Resource Rationality

Thomas Icard

September 12, 2023

Acknowledgements: To be added.

Contents

Chapter 1

Overview and Background

One of the most remarkable human abilities is our capacity to deliberate and reason. We appear to be attuned to the world in a way that transcends reflexive reaction to momentary stimuli. We imagine the merely possible, plan for the future, reflect on the past, wonder what might have been, ponder what is "right" or "good" to do or think, and much more. Rational thought and action seem intimately tied to these paradigmatically rational processes.

At the same time, thinking, reasoning, planning, problem solving, perceiving and learning all cost time, energy, and other precious resources. Rationality is also in part knowing when to bypass more resource-intensive modes of thinking. A second striking feature of human thought and cognition, similar to that of many other organisms, is our ability to cope with challenging and complex problems using highly limited resources. This is true at multiple levels. The mammalian brain—the presumptive substrate for much of individual thought and deliberation—is exquisitely efficient in its allocation and consumption of energy (Niven and Laughlin, 2008). Meanwhile, entire social institutions, such as scientific communities, appear to be structured in part to support effective resource allocation (Kitcher, 1990).

The aim of this Element is to present an approach to rationality that emphasizes the role of resource constraints. A concept of rationality can be understood as a theoretical tool. It should help us understand ourselves, including how we think and how we (think we) ought to think. It can also serve as a means of assessment, providing a normative standard on which to judge behavior in a broad sense. That is, a theory of rationality should help identify and characterize possibilities for improvement. A concept of rationality can serve such purposes more or less well, and a motivating contention here is that sensitivity to resource limitations in particular is paramount. From a scientific perspective, discovering the nature of our own rational faculties demands attention to the limited capacities we bring to problems. From a moral and political perspective, policies and institutions grounded in conceptions of rationality that abstract away from acute resource limits risk exacerbating extant inequities (Mills, 2005; Morton, 2017).

The work presented here is predominantly formal, involving concepts from decision and rational choice theory on the one hand, and the theory of computation on the other. In some ways the framework of *resource rationality* can be seen as a marriage of decision theory and computational theory. The reason for exploring our subject from a mathematical perspective is twofold. The first is simply that we are interested in relatively complex structures and much of mathematics is designed to facilitate the exploration and systematization of closely related structures. The second is that formalization often forces clarity and explicitness about assumptions, sometimes revealing dimensions of a problem that may not be clear at the surface.

While the formal tools we will introduce are quite flexible in their interpretation and in their

accommodation of a wide range of philosophical views, they cannot remain totally neutral. The use of decision theory in particular imposes an emphasis on a kind of "instrumental" rationality. Starting with some primitive conception of what it would mean for things to *go well* for an agent, in a very broad sense, we are interested in what kinds of *behavior*, again in a broad sense, would conduce to things going well. In other words, rather than adjudicating what is inherently good or bad for an agent, the focus will be on tracing out the consequences of assumptions about what is good or bad. Rational pronouncements will thus be of a conditional nature (cf. de Finetti 1974, p. 85): if this is what it means for things to go well, and these are the constraints on the agent's capacities, then this way of proceeding is the best they can do. Such a conclusion will typically follow as a matter of mathematics.

As already mentioned, distinctive of resource rationality is its integration of decision theory and computational theory. The latter is intended to capture crucial details about what agents are really like, with special emphasis on their ability to recognize, store, and process information, as well as the costs involved in doing so. With this emphasis comes greater fidelity to the actual predicaments real agents face. It also thereby narrows the gap between the *normative* and the *descriptive*. Insofar as some of the most interesting rational "oughts" imply "can," we would like a rational framework to hew closely to an agent's assumed capacities. This is just the familiar sense in which the descriptive constrains the normative.

In the other direction, the normative may well be a fruitful guide to the descriptive. Indeed, a prevalent theme in this Element is the use of resource rational analysis to generate promising scientific hypothesis about key aspects of human cognition. Some examples of this will include approaches to cooperation, polarization, oculomotor processing, visual processing in the ventral stream, sequential decision making, reasoning heuristics, and random behavior. The successful application of the methodology depends on an accurate characterization of the problems facing cognition, which typically involve strategic use of limited resources as a core component.

The remainder of this introductory chapter offers further conceptual background and context for the core material to follow in subsequent chapters.

1.1 Conceptions of Rationality

A prominent and traditional conception of rationality in philosophy restricts the concept to organisms possessing a "faculty of reason," that is, a specific suite of internal mental states and processes that canonically includes beliefs and desires (Wedgwood, 2002) and that allow the organism to "recognize, assess, and be moved by reasons" (Scanlon, 1998, p. 23). On this conception, rationality—which can be further divided into *theoretical rationality* and *practical rationality*—is about norms governing these mental states and processes. Theoretical rationality typically concerns what one ought to believe and how one's beliefs ought to change with experience. Practical rationality standardly concerns the resolution of what one ought to do.

Within this traditional ("mentalistic") picture, there is considerable controversy over the exact nature of—not to mention the ultimate source of—rationality, whether theoretical or practical. On one view, rationality is *substantive*: one's attitudes are rational to the extent that they are appropriately responsive to the right *reasons* (Kiesewetter, 2017; Lord, 2018). For instance, a belief ought to be sensitive to the evidential factors that weigh in favor or against it; likewise, a choice of action ought to be properly motivated by practical (notably including moral) reasons, e.g., the fact that one has made a particular commitment (Sen, 1977).

Another view is that there are *structural* requirements of rationality, that an agent's attitudes "cohere" or relate to one another in the right way. For instance, having decided to take a trip to Paris, there is some pressure to figure out how to get there; failing to take the means

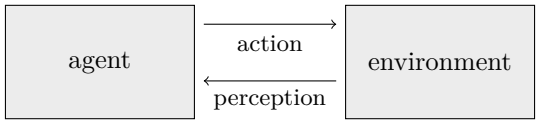


Figure 1.1: A traditional picture of the relation between an agent and their environment, related via perception and action. Similar pictures also appear routinely in the literatures on control theory and reinforcement learning (e.g., Sutton and Barto 1998, Fig. 3.1). In much of engineering the agent is instead called the *controller*, while the environment is the *plant*.

toward an end is a failure of (practical) rationality. In a similar vein, it would be theoretically (or "epistemically") irrational to expect two mutually incompatible events both to occur.

The structural view sits especially well with existing formal treatments of rationality, viz. logic, probability, and decision theory, inasfar as these frameworks concern the synchronic and diachronic relationships among (proposed formalizations of) mental states. For instance, assuming some mental states amount to "degrees of belief," the probability calculus offers a way of characterizing coherence, and potentially also norms for revision (e.g., Rescorla 2021). Such a view is often assumed in cognitive science (e.g., Oaksford and Chater 1999).

Much current debate surrounds both of these distinctions—practical versus theoretical, and substantive versus structural—most saliently about whether one is more fundamental, perhaps even subsuming the other. However, essentially all versions of this traditional conception endorse a picture of the agent relating to the world in a particular way, sometimes called the "sandwich model" of mind (see Hurley 2001). The world (or the "environment") produces "data" for the agent in the form of perceptual input, which transforms their internal ("mental") state; the agent, in turn, chooses actions on the basis of their current mental state, potentially transforming the environment. See Fig. 1.1.

On this picture we can understand different mental states and processes in terms of their "direction of fit," roughly corresponding to the division between theoretical and practical. Some "cognitive" attitudes, like belief, are thought to aim at bringing one's mind in line with the world (i.e., so that the belief is true). Other "conative" attitudes, like desire, aim at bringing the world in line with one's mind (i.e., so that the desire is satisfied). Much of the traditional philosophical discussion concerns when and how norms of rationality produce the right kind of harmony between the world and the agent's mind. The task of one's faculty of reason is precisely to bring about such harmony.

There are variants of the structural view that maintain the basic picture in Fig. 1.1, but eschew mental states like belief and desire as primitives. Such "behavioristic" approaches to rationality replace coherence among mental states with coherence among preferences, in particular preferences as revealed in *choice behavior*. A long tradition in decision theory (Ramsey, 1931; von Neumann and Morgenstern, 1953; Savage, 1954; Jeffrey, 1965) has explored axioms on preference orderings that would be sufficient to guarantee that we can *represent* an agent as maintaining probabilities ("beliefs") and utilities ("desires"), such that their choices maximize the probabilistic expectation of their utilities. By "black-boxing" the agent and the internal causes of their choices, the framework can enjoy a wider range of application, including to non-human animals with potentially divergent mental economies, and it may help to unify choice behavior arising from disparate psychological processes (Vredenburgh, 2020; Thoma, 2021).

A focus on behavior, however, need not proceed in tandem with an axiomatic structural view. On what we might call a *thin* conception, rationality involves nothing more or less than an agent acting (and deliberating) in a way that best conduces to what is good for them.[1] In

[1] Cf. Stich 1990, and further back, American pragmatists like James and Dewey. Note that the thin conception

other words, the thin conception is exhausted by mere instrumental rationality, which Nozick (1993) suggests is, "within the intersection of all theories of rationality (and perhaps nothing else is)" (p. 133). It is generally silent—or to put it more positively, flexible—about what is good for an agent, and about how we parse "conduces to." But here, the tools of decision theory, as a calculus rather than as a set of axioms, can provide some direction.

We can assess an agent using the building blocks of decision theory—states, actions, probabilities, utilities—so long as we can characterize the agent's predicament in terms of these primitives, perhaps alongside others. Unlike the traditional conceptions, but like the behaviorist version of the structural conception, this is by default an *externalist* project.[2] The probabilities and utilities can be, but need not be, the agent's own. Utilities in particular reflect what the theorist takes "going well" to mean for an agent. But they can be grounded in a variety of quite different ways. A few examples include:

1. **Stipulated**: A psychologist studying human reasoning or decision making may simply impose a utility structure on a situation, in order to measure how well experimental participants perform on a well-defined task. Likewise, an engineer might define utilities to characterize a task of interest, using these to induce an artifact into performing the task. As a typical example, where the possible actions are simply guesses about the underlying state, utility might be given by a measure of *accuracy*.

2. **Evolutionary Fitness**: Utility can alternatively be construed in evolutionary terms. As Orr (2007) put the idea, "fitness is the utility that natural selection maximizes" (p. 2998). Fitness for an individual might be understood as something like the number of offspring produced in its lifetime, with the "actions" possible phenotypes (Okasha, 2018). Utilities may play an explanatory role in accounting for features of evolved organisms.

3. **Hedonic**: Early utilitarian philosophers equated utility with certain internal sensations like pleasure or the absence of pain. For example, Edgeworth (1879) famously postulated that, "pleasure is measurable, and all pleasures are commensurable" (p. 396). The problem of how precisely to measure such putative quantities has been a source of much controversy; see Narens and Skyrms (2020) for a thorough recent treatment.

4. **Revealed Preference**: An alternative to hedonic interpretations construes utility in choice-theoretic terms, grounded in in-principle measurable choice behavior. For example, if in some situation an agent would prefer (the result of performing) a to a', then the utility of a ought to exceed that of a'. By extending these preferences to lotteries with known probabilities, von Neumann and Morgenstern (1953) showed how imposition of a few key axioms is sufficient to guarantee a rich utility representation.

5. **Objective Goodness**: Some have contended that we can (and in some contexts should) construe utilities as measures of how *good*, in absolute terms, it would be to take action a in situation s. This might include factors such as what a person prefers, how much pleasure it brings about, and so on. But it also crucially incorporates, for instance, moral considerations. Broome (1991) argues that the von Neumann and Morgenstern approach to utility can be reappropriated for this interpretation.

here is thinner than the view that Elster (1983), citing Rawls (1971), dubs the thin conception. The latter is essentially what we are calling the structural conception.

[2]It is a "moderate" externalism in that it by no means precludes the relevance of internal (e.g., mental) states to understanding behavior (Satz and Ferejohn, 1994). It just does not presuppose any particular cognitive or conative architecture.

Again, only in some of these settings would want to say that the utilities are the agent's own in a meaningful sense. They need not represent utility or desirability in any explicit manner.

A canonical way to analyze what a pattern of behavior "conduces to" is in terms of *expected* utility, that is, how well the agent does on average (or "in expectation"). But even this is not forced upon us. We could instead assess behavior by analyzing a notion of *regret*, or via other modes of combination that, e.g., punish (or reward) risky behavior. Thus, compared to the behavioristic-structural approach, the thin view is more perspectival in its pronouncements: they depend on how we characterize the agent's predicament and how we integrate the theoretical primitives. Yet, what this view lacks in absoluteness it makes up for in breadth of application. Indeed, something like the thin conception of rationality lies at the foundation of a variety of endeavors in science and engineering.

For instance, control theory as a discipline (see, e.g., Todorov 2007) begins with a characterization of a system in terms of a "loss" or "cost" function, which is essentially the inverse of a utility function. Control theory typically deals with noisy environments where uncertainty can be quantified (by the theorist) in order to design an effective controller. This might tell us, e.g., when a system of valves and pumps for managing stormwater flow optimizes environmental and safety concerns (Kerkez et al., 2016). Essentially this same perspective is adopted in much of contemporary artificial intelligence and machine learning (Murphy, 2012), and reinforcement learning in particular (Sutton and Barto, 1998).

Some version of the thin conception arguably grounds some uses of decision theoretic (and game theoretic) ideas in evolutionary biology as well, insofar as we can take utility to measure something like (reproductive) fitness; see Okasha (2018) for discussion. Much of perceptual and motor science is fundamentally couched in the language of decision theory, treating "subpersonal" subsystems of an agent as though they were themselves rational agents, processing data and making decisions; see Ma et al. (2022) for a textbook overview.

Use of the word 'rational' in these contexts may sound like an abuse of language, and there is historical evidence that these various other senses of the word in fact derived from the mentalistic conception sketched above (cf. Broome 2021). In a noteworthy twist, applying this thin behavioristic conception to human agents will promise a distinctive window in our own internal lives, including whatever components might comprise a faculty of reason.

1.2 Bounded Resources

Prominent theories of rationality can be demanding. The expected utility calculus of decision theory seems to require that agents perform extraordinary computations just in order to figure out what to do. It appears to require the agent to be, as Veblen (1898) vividly put it, a "lightning calculator" (p. 389). For instance, when faced with a bet about the millionth digit of π, a person would have to put any amount of money on 5, on pain of irrationality (Good, 1950, p. 49). When playing a game of chess, one of the two players has a winning strategy and so plainly "ought" to play that strategy (Zermelo, 1913).

There is an important sense in which these "oughts" do indeed have rational force. The player has a (presumably decisive) reason to play their winning strategy, the identity of this winning strategy is (structurally) implied what they know, and it would obviously be best for them to play it, given the goal of winning. So it seems rational on any of the conceptions discussed above. The mere fact that discerning this winning strategy is difficult—even that no one currently knows what the winning strategy is (or which player possesses it)—does not obviously call into question its rational status. Sometimes it just takes a lot of time and effort to figure out what one ought, rationally speaking, to do. Drawing the precise connection between rationality and computational limitations can be delicate (Millgram, 1991; Carr, 2022).

Nevertheless, if we want a theory of rationality to relate more closely to the detailed predicaments faced by limited, embodied agents like ourselves, and if we want the theory to shed light on the process of deliberation itself, then some attention to computational resources seems essential. As some have ventured, deliberation for an unbounded agent would be a "waste of time" (Arpaly and Schroeder, 2012, p. 236), perhaps even "unintelligible" (Stalnaker, 1991, p. 429). Deliberation, the suggestion goes, is a *solution* to the problem of bounded resources.

Resource limitations were discussed by many earlier authors, but the concept of *bounded rationality* as a response to standard decision theory in philosophy, psychology, economics, and engineering in particular, is most closely associated with the work of Herbert Simon (1955; 1956). Simon endorses a version of the thin conception of rationality (see, e.g., Simon 1983, pp. 7-8, where he writes that reason is "wholly instrumental"), and this conception lends itself naturally to the incorporation of resource limitations. After all, we want to understand exactly how the agent—as constituted, limitations and all—achieves a desired end.

A second key move in Simon's argument, echoing earlier themes from Dewey (1896), is his critique of the basic picture in Fig. 1.1:

> We must be prepared to accept the possibility that what we call "the environment" may lie, in part, within the skin of the biological organism. That is, some of the constraints that must be taken as givens in an optimization problem may be physiological and psychological limitations of the organism (biologically defined) itself. (Simon, 1955, p. 101)

This move reimagines the distinction between what is "internal" and "external" to the agent. Unlike embodied and enactive approaches that bring the environmental into our understanding of mind (e.g., Hurley 2001), the move is to treat aspects of mind on a par with aspects of the environment. If we want to understand rational agents as solving problems, the problems to be solved must incorporate characteristics of the agents themselves, including memory and processing limitations, and any quirks in how their minds naturally function.

Historically, increased interest in pursuing such an approach during the second half of the twentieth century can be traced to two major developments. The first is the rise of computational approaches to problem solving (starting with seminal advances such as Newell et al. 1959). The very possibility of automating complex tasks put issues of tractability front and center. With the development of artificial intelligence and attempts to automate human-like reasoning and decision making, researchers encountered vexing issues like the so-called *frame problem* (Pylyshyn, 1987). The challenge, which Fodor (1987a) argued "goes as deep as the analysis of rationality" (p. 140), is how to determine which aspects of one's knowledge might be relevant to a given task. Naïvely implementing classical (structural) canons of rationality reveals both how demanding those canons can be, and how remarkably efficient much of human reasoning appears to be in mitigating this problem.

A second contributing factor is the onslaught of empirical work demonstrating the numerous respects in which ordinary human reasoning and decision making seem to part ways with classical canons. Dozens of studies revealed that virtually every pattern in reasoning and decision making that could be violated was, in at least some context (see Slovic et al. 1977; Kahneman et al. 1982 for early summaries). This is obviously a problem if one had hoped to use rational principles to predict individual behavior. But it also raises the possibility that the background rational framework is itself somehow inapt. Appeal to limited computational resources has been one strategy for *vindicating* human rationality in the face of this experimental evidence (Tversky, 1969; Stich, 1990; Gigerenzer et al., 2000; Lieder and Griffiths, 2020). Further support comes from modern AI approaches, which show progress on the frame problem, but also reveal the same characteristic violations of classical canons (Dasgupta et al., 2022).

1.2.1 A Panoply of Research Programs

The term 'bounded rationality' is ambiguous between at least two importantly different readings (cf. Gigerenzer and Selten 2001, pp. 4-6). The first reading construes 'rational' as a gradable adjective, with bounded rationality the study of the many ways in which we fall short of an idealized standard. The perennial finding is that humans are *bounded* in the degree to which they reach this ideal. A second reading construes bounded rationality as a species of rationality, but one that takes resource constraints into account. Bounded rationality, on this view, highlights a distinct normative standard, one that is intended to be more appropriate for assessing human agents in particular. A number of different research programs—with different goals and emphases, and relating differently to these two senses of 'bounded'—have emerged:

> **Heuristics and Biases**: Much of the literature in empirical psychology and behavioral economics takes classical canons of rationality, such as expected utility theory and game theory as well as classical logic and probability, as the standard. The goal is to document and theorize the multitude of ways human judgment and behavior deviate from this standard (e.g., Kahneman et al. 1982). As a typical example, the "cognitive hierarchy theory" (Camerer and Ho, 2015) aims to predict human strategic behavior in terms of a bound on the level of higher-order theory of mind experimental participants can reach.
>
> **Minimal Rationality**: Many normative projects aim at lowering the rational standards so that bounded agents might have some hope of meeting them. In fact, Savage's (1954) original motivation to focus on "small worlds" was to ensure that even mundane decision making would not be "utterly beyond our power" (p. 16). Proposals in economics, like Simon's (1955) original account of "satisficing" or Gilboa and Schmeidler's (2001) "case-based" decision theory, are intended to be tractable by design. Similar projects in philosophy have explored more tractable standards for theoretical/epistemic rationality (Cherniak, 1986; Harman, 1986), for practical decision making (Weirich, 2004; Pollock, 2006), and for rational learning (Skyrms, 1990; Huttegger, 2017). This program might be summarized: "Moderation in all things, including rationality" (Cherniak, 1986, p. 9).
>
> **Ecological Rationality**: Adopting a thoroughly externalist and instrumentalist mode of rational assessment, a major research program has explored "heuristics" that people use to solve difficult problems (Gigerenzer et al., 2000; Gigerenzer and Brighton, 2009), with a focus on features of the environment that render such simple solutions adaptive. One of the key findings is that simple heuristics can be not only simpler, but more effective (e.g., in predictive accuracy) than more complex strategies that attempt to use the calculus of decision theory itself, a phenomenon known as the "less is more" effect.
>
> **Bounded Optimality**: By taking computational and other resource bounds into account, we could try to determine how well an agent could do within those bounds, such that they still have a chance of being "optimally imperfect" (Baumol and Quandt, 1964); cf. Good (1950, 1952). This is a common mode of theorizing in economics (Chapter 3) and a version of it in computer science seeks to identify the optimal program for solving a given problem, net of computational costs (Horvitz, 1987; Russell and Subramanian, 1995). It has even been suggested that, "artificial intelligence can be usefully characterized as the study of bounded optimality" (Russell and Subramanian, 1995, p. 576).

The bounded optimality approach in particular has influenced a closely related research program in cognitive science that has gone under various names, including *computational rationality* (Lewis et al., 2014; Gershman et al., 2015), *algorithmic rationality* (Danks and Eberhardt, 2011; Halpern and Pass, 2015), and *resource rationality* (Lieder and Griffiths, 2020).

Although our understanding of the concept differs in numerous ways from previous treatments, we adopt the term 'resource rationality' for the target of interest in this Element, generalizing beyond its methodological application in cognitive science (the subject of Chapter 5). The framework presented here has much in common with minimal rationality and ecological approaches. The difference is largely one of emphasis, and conceptual and technical primitives. 'Resource rationality', which appeared first in print three decades ago (Cherniak, 1994, p. 103), connotes the idea of being a *resourceful* agent. It also avoids the ambiguity inherent in 'bounded rationality' and highlights the central role of *procedure* in reasoning and decision making.

1.2.2 The Importance of Procedure

The concept of resource rationality presented in this Element supplements the usual components of decision theory—probabilities, utilities, etc.—with assumptions about what the agent is like. That is, while behavioristic conceptions of rationality tend to "black box" inner processes and computations, these are front and center for resource rationality. This gives a *procedural* emphasis (Veblen, 1898; Simon, 1976), highlighting the sequence of steps an agent undergoes in the course of decision making. Roughly speaking, "an action is rational if and only if it is adequately supported by appropriate deliberative procedures" (Gauthier, 1994, p. 700),[3] where "deliberative procedures" is to be understood here very broadly to encompass not only *reasoning* in the ordinary (mentalistic) sense, but also internal mechanisms that generate the behavior of simple systems, including subpersonal subsystems of people.

Concretely, the component we add to traditional decision theory is the concept of a *program*. Interpreted in a very general way, programs receive some type of input (e.g., through perception), process that input in some way, and then generate some output (viz. action). The idea of mental processes as being akin to programs is a familiar one (Turing, 1950; Newell et al., 1959; Putnam, 1967; Pylyshyn, 1984; Rule et al., 2020). We understand programs to include "symbolic" procedures that manipulate explicit internal representations in a way analogous to computer programs, but also "subsymbolic" procedures which generate behavior through cascades of distributed processing units that may or may not manipulate concrete representations (Rumelhart et al., 1986), and indeed many other types of procedures.[4]

Much of the interest in resource rationality stems from the setting of *sequential* decisions, that is, series of decisions made over some period of time. A *strategy* (to be introduced formally in Chapter 2) is essentially a behavioral disposition, that is, a specification of a choice for every possible eventuality. We can understand programs as *implementations* of these abstract strategies. Distinctive of programs is that they may consume resources, and thus "running" a program—that is, implementing a strategy in a certain way—incurs costs.

To give a preliminary sense for how solving a problem under cost constraints might look, we recall a small-scale example from the computing literature:

Example 1. In their presentation of bounded optimality, Russell and Subramanian (1995) study a relatively simple mail-sorting problem; see Fig. 1.2. A machine must process incoming mail by scanning hand-written zip codes and dispatching parcels to appropriate buckets. The goal is to maximize allocation accuracy while minimizing rejections and jams (which occur if the next parcel comes before the last one has been allocated). In this setting there is a mechanism

[3] Note that the view Gauthier is propounding in this paper is more committal about aspects of sequential decision making than we will be in this Element; see §2.1.7, §6.4. Outside action theory, the centrality of process has been emphasized in epistemology too (e.g., Goldman 1986; Ross and Schroeder 2014; Thorstad 2023).

[4] Some authors in the philosophical literature on resource limitations, e.g., Wimsatt (2007), draw a distinction between programs or algorithms on the one hand, and *heuristics* on the other. Following the tradition of Newell et al. (1959) and others, we intend 'program' here to be sufficiently broad, certainly to include anything that has been labeled a heuristic in the literature.

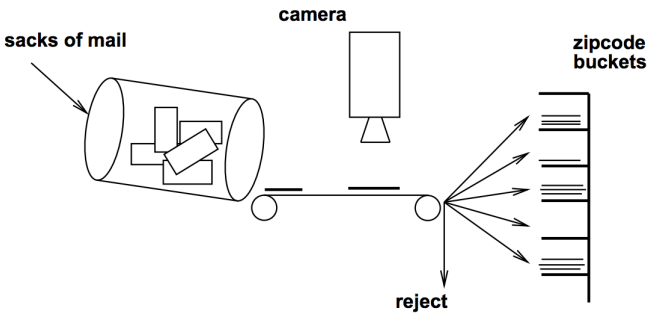


Figure 1.2: The mail-sorting problem analyzed in Russell and Subramanian (1995), reproduced with permission from the *Journal of Artificial Intelligence Research*.

designer, who possesses a set of neural networks for reading zip codes. These networks are of varying qualities (viz. accuracy) and varying execution times, and all of this is known. Having seen the output of several networks, the mechanism will follow the allocation recommended by the most accurate network. The problem is that better networks may take longer to run, risking a jam; so the challenge is to find the resource-optimal *sequence* of networks to run.

Under different cost regimes—fixed deadline, fixed time cost, and stochastic deadline—Russell and Subramanian show how to derive the optimal sequence. Under a fixed time budget, it is optimal simply to run the highest quality network that fits within the budget, since one is sure that it will terminate in time. In other settings, sequencing together several faster networks is optimal, always in order of increasing quality. This happens, for instance, in the stochastic deadline regime when expected arrival rates are high, or when arrival time variance is high.

The mail-sorting task is of course very different from tasks that people ordinarily face. Nonetheless, it provides a very simple illustration of what it means to identify a program—in this case, a sequence of neural networks—that achieves the best performance under given resource constraints. Another noteworthy feature of the example is that the aim of the analysis is to identify not just a good program that can solve the task, but an *optimal* one. Optimality is sometimes viewed as anathema to boundedly rational approaches, especially in much of the literature on minimal rationality, so it is worth clarifying its role in the present Element.

1.3 Constrained Optimization

Fig. 1.1 depicts an agent facing an environment, related to it by perception in one direction and action in the other. The *task*, as construed on the thin conception of rationality, is to take the sequence of actions that best responds to incoming information about environmental circumstances.[5] Suppose an agent's behavioral dispositions are captured by (the strategy implemented by) a program π, and that we have a measure $\mathbf{U}(\pi)$, telling us how effective π is in this environment (to be formalized below in Chapter 2; see Eqs. 2.9 and 2.18). So far this just looks like traditional decision theory. Simon (1955) implores us to acknowledge that part of the problem facing an agent stems from constraints and limitations internal to the agent (that is, within the "agent" box in Fig. 1.1). Let us capture this by imposing a *cost* $C(\pi)$ on the program π. Like the performance measure $\mathbf{U}(\pi)$, the cost $C(\pi)$ may well depend on the

[5]Solving a task well in this sense is sometimes to referred to as *substantive rationality* (Simon, 1976), not to be confused with the substantive conception of rationality that centers attention on reasons (§1.1).

environment, e.g., which perceptual inputs are to be processed. So in the probabilistic setting it too will often be an expectation of a random variable.

A surprisingly robust ("cost-benefit") formulation of the problem facing the agent amounts simply to a linear combination of these two terms:

$$\mathbf{V}(\pi) \quad = \quad \mathbf{U}(\pi) - C(\pi). \tag{1.1}$$

The overall value $\mathbf{V}(\pi)$ of running program π in a given environment is determined by how well the agent performs at the task, externally assessed, minus the internal "computational" costs that π incurs while achieving this performance.

Given a class Π of possible "agent programs" (Lewis et al., 2014; Icard, 2014), or "biologically feasible minds" (Lieder and Griffiths, 2020), we say that $\pi \in \Pi$ is resource rational to the extent that $\mathbf{V}(\pi)$ is high. In practice, there are two different approaches to resource rationality, depending on whether we look at fine-grained costs or simply rule out of Π any programs that would be above a fixed "resource budget":

> **Cost-theoretic approach**: Given some architectural assumptions about a class Π of programs, we can estimate the cost of performing certain operations, or of maintaining and updating relevant internal states. In this way, programs are compared both in how well they perform the task and in how efficiently they mitigate fine-grained costs.
>
> **Panoramic approach**: Without analyzing the costs of specific operations, we can simply restrict Π to a class of programs that we assume are all broadly feasible without attending to any resource differences among programs. In effect, this assumes $\mathbf{V}(\pi) = \mathbf{U}(\pi)$ for $\pi \in \Pi$, that any differences in resource usage are negligible or irrelevant.

We return below (§1.4) to reflect further on what cognitive costs are in the first place, and on when one or the other approach is more appropriate.

In many settings there will be programs that are *optimal*, in the sense that $\mathbf{V}(\pi)$ is not exceeded by any other program in Π. That merely signifies that the agent is performing the task as well as possible, given inherent resource limitations (as in Ex. 1). Numerous authors have rejected appeal to optimality in this setting, on both descriptive and normative grounds.

For descriptive projects aimed at uncovering the mental processes that humans and other agents actually use to solve problems, the objection is that optimality is simply not within the realm of possibility. In his critical review of the rational analysis program in cognitive science (§5.1), Simon (1991) argues that the mind's solutions to hard problems tend to be somewhat arbitrary, and probably do not solve any interesting constrained optimization problem:

> Bounded rationality is what cognitive psychology is all about. And the study of bounded rationality is not the study of optimization in relation to task environments. (Simon, 1991, p. 35)

Authors like Gigerenzer and Selten (2001) and Gigerenzer and Brighton (2009) suggest that simple heuristics (such as "Take-the-Best") are probably not optimal, but they are the strategies that people seem to use much of the time. A further reason for doubting that minds in particular (perhaps unlike some simple computer programs) could be employing resource-optimal strategies is that the problem of *identifying* the resource-optimal strategy is often as difficult as solving the task itself (van Rooij et al., 2019). Even if agents are not individually assumed to discover the resource-optimal strategy themselves, online or through learning, tractability appears to prevent evolutionary emergence of the optimum just the same (Rich et al., 2020).

From a normative point of view, common wisdom is that imperfect agents should typically not attempt to be optimal in any robust sense. This is an important lesson from Gigerenzer

and Brighton's (2009) discussion: not only do people naturally employ simple heuristics, but their performance is decidedly worse when they attempt to use a putatively optimal method. Similar arguments have appeared in epistemology (see, e.g., Staffel 2017).

A related objection to optimization is that it threatens regress. The point of taking costs into consideration is that making good decisions demands intelligent allocation of resources. But if an agent has to figure out the best means of allocating resources before deliberating, then they are faced with a meta-decision-problem that may be at least as demanding. If optimizing something like Eq. (1.1) is the normative target, then we need to worry about how the agent will figure out the optimum of Eq. (1.1), in effect spawning another instance of the same equation "one level up." The lesson that many have drawn is that taking deliberation and resources into account demands that we give up on optimization altogether (Ryle, 1949).

While these objections have considerable force against some "constrained optimization" projects, we can respond to all of them by doubling down on the external and perspectival nature of the thin conception of rationality. There is no assumption that Eq. (1.1) is considered in any explicit way by the agent, and there is no normative requirement to strive to maximize $\mathbf{V}(\pi)$. On the contrary, the framework captures the fact that in some contexts deliberation of any sort is (viewed externally) suboptimal; its costs need only outweigh its benefit. In other words, the framework correctly diagnoses both the external assessment (what would be good to think about) and the internal predicament ("Do not think!"). Cf. Example 31 in §6.5.

The panoramic approach encompasses proposals like satisficing as special cases, for instance, by suitably coarsening $\mathbf{U}(\pi)$, letting all programs that guarantee a fixed "aspiration level" (Simon, 1955) be equivalent in task-performance. As Baumol and Quandt quipped, this makes satisficing a kind of "constrained maximization with only constraints and no maximization" (Baumol and Quandt, 1964, Fn. 2). In a similar vein, Gigerenzer and colleagues' claim that simple heuristics outperform putatively optimal, e.g., Bayesian methods (when the latter are misaligned with the environment), can be seen as a claim that the value of one program is greater than that of another, precisely in the sense of Eq. (1.1). How we characterize the task—namely, $\mathbf{U}(\pi)$—and how we characterize the costs—$C(\pi)$; see the next section, §1.4—together determine how we understand the landscape of resource rational agents.

Of course, this perspectival nature of the framework is a potential hazard when our aim is to *uncover* the mental processes that generate human behavior. As a research strategy in cognitive science, resource rational analysis only has a chance of success to the extent that our characterization of the problem matches some features of the agent's history—whether at the scale of evolution, of development, of learning and online inference, or some combination of these—that could plausibly have led to optimization. Here the concerns about inheritance of intractability (Rich et al., 2020) are quite germane. In suitably complex settings, even we as theorists will not be in a position to solve for the optimal program. A further issue is that we often lack a sufficiently detailed understanding of the task and plausible costs to formulate the problem so that Eq. (1.1) is well-defined in the first place.

Ultimately, the success of the descriptive project of resource rational analysis (Chapter 5) will be judged by its theoretical and empirical fruit. As a counterpoint to the more "heuristic" approach to discovering heuristics advocated by Simon (1955, 1956); Gigerenzer and Selten (2001); Gigerenzer and Brighton (2009) and others, formulating problems in terms of optimization often enjoys, "generalizability, mathematical simplicity, elegance, and heuristic fruitfulness" (Mongin, 2000, p. 105). Somewhat perversely, this is the same argument that has been marshaled in favor of idealized models ignoring cognitive resources altogether (e.g., Friedman and Savage 1948). Indeed, on the view of resource rationality adopted in this Element, there is some merit to the charge that it marks only a moderate departure from traditional accounts of decision theoretic optimization (see, e.g., Simon 1991; Gigerenzer and Selten 2001; Arrow 2004

among others). To some this will be a feature, to others a bug.

Returning to the normative, while adopting an external perspective it would nevertheless be desirable to draw some meaningful connection to the internal context of deliberation. We can understand Eq. (1.1) as assessing "habits of mind" that might be more or less (instrumentally) desirable to cultivate. To that extent, knowing which habits of mind conduce to which desirable ends may be of some limited use for a deliberator. The question of how such habits might be trained or induced, if they can be at all, is a species of a more general family of questions about how to develop skill, virtue, and other qualities whose active pursuit can undermine their realization. As Whitehead colorfully put it,

> Civilization advances by extending the number of important operations which we can perform without thinking about them. Operations of thought are like cavalry charges in a battle—they are strictly limited in number, they require fresh horses, and must only be made at decisive moments. (Whitehead, 1911, p. 46)

This is a familiar theme from discussions of ethics and virtue (e.g., Annas 2011), where the roles of education and reflection are carefully negotiated with the importance of non-deliberative thought and action. But while "intentionally not maximizing" is surely part of a good solution to the deliberative predicament, this is not the same as "intentionally sub-maximizing" (Pettit, 1984), that is, choosing a worse strategy when strictly better ones are conspicuously available.

1.4 Costs and Resources

Distinctive of resource rationality is its incorporation of agent programs, and with this, consideration of resources and their costs. Fundamentally, we are assuming there are ("internal") physical resources that an agent can bring to bear in generating its behavior: energy, memory space, and so on. Such resources do not bear value inherently, but only instrumentally through an appropriate sort of *transaction*. Broadly cognitive resources, similar to dollar bills or oil, are valuable only in the context of a functioning system, and only insofar as they can be "exchanged" for something assumed to have value. For example, probing one's memory for whether the next turn is a right or a left is worth the cognitive resources because it is more important to arrive at the correct destination on time than to direct those same resources toward some other aim. This is not just a notional sense of a foregone opportunity. A physical process has taken place in which concrete resources were "consumed." Resource rationality is marked by the separation and explication of cognitive resources and their functional roles in the production of intelligent behavior and decision making.

Cognitive resources are scarce and scarcity implies cost: by employing resources for one purpose one foregoes the use of these same resources for another purpose. This is why economists often equate cost with *opportunity cost*: "the cost of obtaining anything is what must be surrendered in order to get it" (Robbins, 1934, p. 2).

Applying this to the present setting, we might imagine that the cost of employing a resource for a given purpose has something to do with the value that resource would have if it were not being put to this purpose. This is a *counterfactual* analysis (Nozick, 1977): choosing some option x consumes some resources, and the cost of those resources is the value of the alternative action x' that the agent *would have* taken using the same resources, were x somehow unavailable.

It may seem puzzling why this counterfactual notion of cost is needed at all. It appears redundant: an agent would rationally opt for x just in case the value of x exceeds its cost in resources, but that happens just when x is the best available option, that is, the best use of those resources. In fact, postulating separate costs is unnecessary—indeed, redundant—whenever we possess an explicit, exhaustive list of all the relevant possible uses of the resource.

Such contexts are when the *panoramic* approach is often most apt, as all opportunity costs are already incorporated into the analysis.

Example 2 (Shared Representation)**.** Suppose, for example, that some neural structure is common to multiple different functional pathways. That is, the very same neurons are used as a critical resource for distinct cognitive or perceptual tasks (e.g., recognizing words and recognizing colors). This appears to be a fundamental principle of brain organization not just for reasons of "reuse," but in that sharing neural pathways (and intermediate "representations") facilities faster and more effective learning (Rumelhart et al., 1986; Rogers and McClelland, 2006). At the same time, such sharing raises a problem for multitasking because effective use of the pathway for one purpose (e.g., recognizing a word) comes at the expense of the other (e.g., recognizing a color). The brain thus faces a tradeoff between efficient learning and ease of multitasking. A program—in this case a particular organization of neural pathways, that is, a particular use of a fixed set of resources—might strike a better or worse balance among its various aims. As a preview of resource rational analysis (Chapter 5), Shenhav et al. (2017); Musslick and Cohen (2021) suggest that the brain does in fact optimize this tradeoff. (We will encounter a rather different example of neural wiring optimization in §5.3.2.)

However, many factors that we would like to classify as cognitive resources can be used for an intractably wide array of purposes—they are, as Klein (2018) puts it, causally promiscuous. As a prototypical example, ATP is consumed as part of numerous biochemical (and specifically neural) processes, from maintaining membrane potentials to synthesizing proteins and neurotransmitters like dopamine (Niven and Laughlin, 2008), which are themselves crucial components in a multitude of cognitive functions. A very different example of a highly promiscuous physical resource is *time*, which is "consumed" by virtually any process.

Due to their promiscuity, as well as their fungibility—e.g., any token ATP molecule is as good as any other—it is useful to *quantify* resource usage, as a stand-in for opportunity cost. To understand the role of resources for a particular cognitive problem, it may not be enough to look at all feasible solutions to that very problem (as in the panomaric approach), because those same resources might be involved in solutions to other problems as well. Rather, we want to solve this problem in a way that *minimizes* the use of resources that could be put to other ends. That leads naturally to the cost-theoretic approach.

Costs postulated in cost-theoretic approaches to resource rationality have been multifarious (cf. Ma and Woodford 2020), reflecting different levels of abstraction and dimensions of idealization. For complex behaviors it is often difficult or unhelpful to pinpoint the exact low-level (e.g., biochemical) costs involved, and theorists turn instead to higher-level abstractions like memory space and number of "queries" to memory, attention, "willpower," and many others. Examples like these abstract in space, e.g., considering large neural populations, and also in time, e.g., holistically measuring the cost of a heterogeneous sequence of lower-level operations.

Even for relatively low-level (biochemical) resources, characterizing costs calls for idealization. While it is generally true that having more of a resource is better—greater resource consumption should generally be more costly, since those resources could be used for more ends—it has been recognized at least since the 1730s that the value of resources (like money or time) scales sublinearly with quantity. Because we typically only deal with relatively modest quantities of a resource (ATP molecules, steps of time) it is very often assumed that value forgone, i.e., the cost, is nonetheless proportional to the amount consumed. Suppose executing program π incurs "true" cost $C(\pi)$, but instead we measure a "pseudo-cost" $\widetilde{C}(\pi)$, such as the amount of ATP consumed or the number of time steps taken. The assumption is that at least

the ratios between costs of any two programs are equal (Knight, 1934, Fn. 3):

$$\frac{C(\pi_1)}{C(\pi_2)} \quad = \quad \frac{\widetilde{C}(\pi_1)}{\widetilde{C}(\pi_2)},$$

and thus the ratio $\widetilde{C}(\pi)/C(\pi)$ is the same number—call it β—for any program π. This means that we can rewrite Eq. (1.1) instead in terms of pseudo-costs:

$$\mathbf{V}(\pi) \quad = \quad \mathbf{U}(\pi) - \frac{1}{\beta}\widetilde{C}(\pi). \tag{1.2}$$

The common idealization is that $\widetilde{C}(\pi)$ just is (a factor β of) the true cost $C(\pi)$ of executing π. This idealization is useful to the extent that $\widetilde{C}$ really does track the value of foregone alternatives. In empirical work this "conversion factor" β is often fit from data. Since it is fixed for all possible programs in Π, this severely limits the degree to which costs can be postulated *ad hoc* to rationalize any behavior (cf. Schoemaker 1991; Arrow 2004).

We will encounter many different approaches to cost in this Element. Some examples include, with increasing degrees of abstraction:

1. **Metabolic**: Much study has been devoted to the energetic and metabolic costs involved in maintenance and operation of neural structures (Niven and Laughlin, 2008; Bullmore and Sporns, 2012; Cherniak, 2012; McNamee and Wolpert, 2019; Ma and Woodford, 2020). For instance, the cost of neural connectivity is assumed to scale with "length or distance of inter-neuronal connections and the cross-sectional diameter of axonal projections" (Bullmore and Sporns, 2012, p. 337). Further costs accompany the synthesis of neurotransmitters and "macromolecules" like proteins, maintaining the right concentrations of ions between cell membranes, and much else (Niven and Laughlin, 2008).

2. **Computational**: At a slightly higher level of abstraction, if we liken mental operations to those of a specific computational model, we can use components or operations of the model as putative stand-ins for costly mental operations. This might be the number of steps or amount of memory used in a run of a Turing machine, the number of states in a finite automaton, the width of a feedforward neural network, or the number of steps along a Markov chain in a Monte Carlo inference algorithm.

3. **Informational**: Even more abstractly, information theory captures one sense in which "information" can be understood as a resource. A number of different theorists have suggested using concepts like *(relative) entropy* and *mutual information* to characterize resource costs (Mattsson and Weibull, 2002; Tishby and Polani, 2011; Still and Precup, 2012; Ortega and Braun, 2013; Lai and Gershman, 2021). Often inspired by coding theory or even statistical physics, these approaches are touted as being less committal about the substrate of internal processing (Bialek et al., 2001; Ortega, 2010).

It is worth emphasizing again that the construal of costs is, by default, from an outside theorist's perspective. This by no means precludes the internal representation of costs by the agent; in some cases this may even help explain *how* a given tradeoff is optimized (cf. §6.5). Costs as represented by the agent, however, need not be understood in the same way. They may not be encoded as opportunity costs in any substantial sense, but, e.g., as "intrinsic," in need of no further analysis beyond their inherent disutility (Kool and Botvinick, 2018).

To summarize, peering inside the agent introduces a transactional perspective on decision making: internal currency is being spent in part to bring about external states of affairs. Cognitive processes have costs in the context of a mental economy in which execution of those processes may help conduce to outcomes that are more or less good for the agent.

1.5 Overview of Chapters

The goal of this element is to advance a particular view of rational decision making under resource constraints, and to work toward a reasonably unified framework for theorizing about resource rationality. A number of authors have been skeptical that such a task is achievable. For instance, on the project that Simon catalyzed in the 1950s, Aumann comments, "There is no unified theory of bounded rationality, and probably never will be" (Aumann, 1997, p. 3). As discussed above (§1.2.1), we can agree with this assessment while still pursuing a more unified account of resource rationality in particular, with its characteristic combination of decision theory and computational theory. In some respect, the challenge is to see how far the "thin" externalist, instrumental conception of rationality can take us toward illuminating keystone features of rational thought and action; the contention is that it will take us much further if we incorporate computational resource considerations.

The hope is that what emerges in the pages that follow will help ground a unified set of conceptual and technical tools, building on and integrating a wealth of previous work in philosophy, computer science, cognitive science, statistics, and economics.

Concretely the substantive chapters are set out in the following way:

Chapter 2: We introduce the formal underpinnings of "traditional" decision theory, along the lines of the thin conception sketched above. Special attention is given to sequential decision problems, including ideas and key results from reinforcement learning.

Chapter 3: As a first foray into decision making with bounded resources, we revisit work in economics and game theory from the 1980s and 1990s in which "players" are assumed to be abstract machines, viz. Turing machines or finite automata. Many of the canonical results in game theory, e.g., Nash's theorem on existence of equilibria, fail in this setting.

Chapter 4: As an extended case study, we explore the question of when and how resource bounds might justify behavior that is random or otherwise indeterminate. In addition to the machine model perspective introduced in Chapter 3, we also introduce the relatively abstract model of costs founded on information theory.

Chapter 5: The research program in cognitive science recently labeled *resource rational analysis* is presented and assessed, as a concrete illustration of descriptive projects issuing from the normative project. Several examples are given from the recent empirical and modeling literatures, and methodological issues are discussed.

Chapter 6: In the most speculative part of the Element, based on the more technical work of the previous four chapters, we explore the possibility that many core features of human thought and agency may plausibly be shaped by resource considerations. The form of the conjecture is that the feature in question makes sense in light of (and perhaps *mainly* in light of) the need to make good use of limited resources.

Finally, in the conclusion (Chapter 7), we revisit some of the most central themes that pervade our discussion throughout. As one notable example, a pressing question is how far we want to go in incorporating more "realistic" features of agents like ourselves, particularly given our (perhaps extended but still highly) limited capacities and resources as theorists.

It is worth emphasizing once more that the stance in this Element toward human rationality generally is neither one of vindication nor of criticism. Sensitivity to very real resource constraints does often allow rational sense to be made of otherwise puzzling human behavior. However, the concept of resource rationality developed in the Element is not offered as part of a sweeping hypothesis about human nature, but rather as a theoretical tool, one that may prove more or less useful in understanding and assessing complex agents like ourselves.

Chapter 2

Traditional Decision Theory

The goal of this chapter is to give formal and conceptual background on what we are calling traditional decision theory. This is essentially decision theory where "internal" resource costs are idealized away. Even before introducing agent programs and resource costs, the concepts employed by traditional decision theory—probability, utility, and so on—themselves invite considerably controversy. The presentation in this chapter is motivated by the thin conception of rationality discussed above in §1.1. The philosophical discussion is intertwined with concepts and tools arising in other disciplines, especially control theory and reinforcement learning. Much of this chapter can be used as a reference, skimmed on a first reading.

2.1 Bayesian Decision Theory

Prior to introducing the sequential decision setting, which will be our focus for most of the Element, we begin by dwelling on the conceptual primitives involves in the standard "Bayesian" decision theoretic framework. As a slight elaboration of the "sandwich model" introduced in Fig. 1.1, a depiction of the basic setup—with states, observations, probabilities, actions, and utilities—appears in Fig. 2.1. We proceed to discuss each of these primitives in turn.

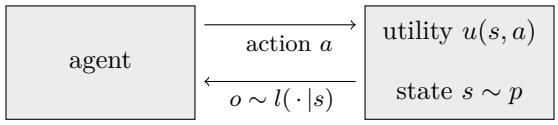


Figure 2.1: Standard decision-theoretic picture. What goes in inside the "agent" box will not be specified in this chapter. A common picture locates representations of probabilities, actions, and utilities here. No such assumption is required on an external ("behavioristic") perspective, though we return to the agent side in later chapters (5 and 6).

2.1.1 States, Observations, and Probabilities

The uses of probability in this Element will be largely elementary. We will typically assume there is a set $S = \{s_1, s_2, \dots\}$ of *states*, viz. relevant ways the world might be at any given time. In theorizing about complex interactive systems it is often useful to assume there is some *probability distribution* associated with the states, measuring how likely the world is to be in a

given state s. This is a function $p : S \to [0, 1]$, assigning a probability $p(s)$ to each state $s \in S$, with the only requirement that $\sum_s p(s) = 1$; the probabilities sum to 1. We let $\Delta(S)$ denote the set of all such probability distributions on S.

In some settings we want to assume that an agent will observe the state s of the world directly. But in others it is useful to capture the idea that an agent only observes some (potentially imperfect) reflection o of the state. Given a set O of possible *observations*, we will often have a *likelihood* function $l(o|s)$, specifying the probability of observing o in state s. Note that, for each s, the function $l(\cdot|s)$ is thus an element of $\Delta(O)$.

Together, p and l determine a *posterior probability distribution* on states. That is, after observing o, the probability that s is what generated o is determined in a Bayesian fashion:

$$p(s|o) \quad = \quad \frac{l(o|s)\, p(s)}{\sum_{s'} l(o|s')\, p(s')}. \tag{2.1}$$

Since the denominator is the same for every state s, it is notationally convenient to write simply that the posterior is *proportional to* the likelihood times the "prior" probability on s:

$$p(s|o) \quad \propto \quad l(o|s)\, p(s).$$

If we imagine a sequence of observations $\vec{o} = o_1, \ldots, o_n$, with each o_i generated independently, then the posterior is proportional to the product:

$$p(s|\vec{o}) \quad \propto \quad \prod_{i \leq n} l(o_i|s)\, p(s),$$

again combining the prior on s with the probability that it would have generated each of these observations. Note that, again assuming the observations are generated independently, we can calculate the posterior on s "all at once," and the order of observations does not matter. By contrast, for an agent that is processing each observation in sequential turn, even if they are probabilistically independent, the order in which they are processed might matter. For bounded agents, we will see that this is often the case (e.g., in §3.5, §5.2).

In a few places (§2.4.2, §4.1.2, §4.4.1, §5.3.4) we will deal with state spaces S that are not discrete, and therefore call upon the resources of measure theory. Assuming a so-called σ-algebra on S—a set $\mathcal{E}$ of "measurable" subsets of S that is closed under complementation and countable union—we define a probability measure to be a function $\mu : \mathcal{E} \to [0, 1]$, such that $\mu(S) = 1$ and $\mu(A \cup B) = \mu(A) + \mu(B)$ whenever $A \cap B = \varnothing$.

The picture we have sketched so far is an external one. We imagine an agent situated in some state in S, and with possible observations O, and we the theorists associate the situation with a prior on S and a likelihood for possible observations. It is against this theoretical backdrop that we will typically be assessing different possible agents, by which we mean, different possible ways of thinking, inferring, deliberating, reacting, and choosing.

Of course, some agents may themselves maintain representations for dealing with their own uncertainty about the world, either implicitly or explicitly. Most obviously, an agent might maintain and update encodings of probabilities $p(s)$ directly (see, e.g., Rescorla 2020). Or one might maintain a more coarsened or qualitative encoding, e.g., tracking merely which possibilities are *more or less likely* than which others (see de Finetti 1937; Ibeling et al. 2023), or coarse-graining the state space by ignoring some distinctions between states (cf. §5.3.3, §6.2). At several points we will consider the possibility that an agent *implicitly* encodes a distribution on states through an ability to *sample* possibilities—that is, to token a representation of a possibility—roughly in proportion to its probability (§4.1.3, §5.3.4, §6.3.1). In all of these cases the agent's internal perspective on the situation may differ substantially from the theorists,

in the probabilistic values and relationships among states, or even more dramatically in the characterization of what the space S of possibilities is in the first place.

An advantage of separating the internal and external perspectives is that we can employ the latter to adjudicate among different instantiations of the former, comparing different representational and inferential strategies to one another, and even to alternative agent architectures that eschew such representations altogether.

2.1.2 Actions

In addition to the set S of states, another primitive is a set A of *actions* that our agent might take. Decision theory as an abstract ("thin") framework presupposes only a minimal sense of agency and action, the latter understood merely as an event that is, in some sense, directly under the agent's control. This may of course include robust intentional action involving rich linguistic or representational capacities such as self-knowledge (Anscombe, 1957) or maintaining an agent "standpoint" (Bratman, 2018). But it also encompasses the movements of simple artifacts such as allocating mail parcels into buckets (Example 1), or neural systems trading off between competing goals like learning and multitasking (Example 2). The framework is intended to be sufficiently broad as to include anything that can sensibly be attributed any kind of *behavior*, even in the absence of anything we would want to call deliberation. The actions A are to be understood as the primitive building blocks of behavior. (More complex *sequential* behaviors will be introduced shortly in §2.2.)

Thus, while resource rationality is characteristically concerned with the thought and deliberation that generate behavior, the decision theoretic framework adopted here is designed for assessing any behavioral dispositions whatsoever, even when deliberation is absent.

A recurring theme—and the topic of Chapter 4—is the idea that some agents may behave in a non-deterministic, or stochastic manner. To capture this, we will sometimes expand the set A of actions to the set $\Delta(A)$, all probability distributions on A, giving the probability with which an agent realizes a given action.

2.1.3 Utility

Assessment of different ways of being in the world depends on some grasp of what it is for a given situation (or sequence of situations) to be better or worse. Specifically, we might be interested in a measure of how "good," or how "desirable" (Jeffrey, 1965), it would be to take action a in situation s. A theoretical tool in this vicinity is the *utility function*, a map $u : S \times A \to \mathbb{R}$. A central question of utility theory is how we should understand what the numbers $u(s, a)$ mean. In §1.1 we listed five possible candidates for a theoretical understanding of utility, all potentially independent of whether the agents in question have their own representations of utility. As with probability, an agent might directly encode a measure of how desirable they take various outcomes to be. Alternatively, they might maintain more abstract representations, e.g., of relative desirability between outcomes. A particularly minimal representation would simply be to encode a "goal" state, without further detail. This is tantamount to saying, e.g., that $u(s, a) = 1$ just in case a achieves the goal in state s, and $u(s, a) = 0$ otherwise (cf. §1.3). And as with probability, our external understanding of utility may depart from the agent's own.[1]

Many of these interpretations face difficult challenges of *comparison*. Most obviously, by taking the values $u(s, a)$ to be (real) numbers, we presuppose that (relative to any state) any

[1] Strikingly, Millgram (2009) argues that it would be positively irrational to maintain overall preferences that could be represented by a utility function in the tradition of von Neumann and Morgenstern (1953).

two actions are comparable. This may be unproblematic when talking about evolutionary fitness or accuracy. But for many decisions—such as Sartre's famous student deciding between caring for an ailing parent and joining the French resistance—the options seem genuinely incommensurable. See Andersson and Herlitz (2022) for a recent compendium on this topic.

Another problem of comparison comes from the multi-agent setting. How are we to trade off between an increase in one's utility against a decrease in another's? Are comparisons between individuals meaningful at all? See Broome (1991) or Narens and Skyrms (2020) for different approaches to this problem motivated by different interpretations of utility.

A seemingly more tractable, but still formidable, problem is how to compare and aggregate utilities for a single individual across time. For instance, how should we compare a good outcome at the present moment with the "same" good outcome at a future moment? We will occasionally assume, following suggestions in psychology and economics, that future utilities are (geometrically) "discounted"—cf. Eq. (2.7) below—though this is obviously not a completely neutral assumption.[2] The questions become even more delicate when agents experience a change of heart in what they value (see §2.1.7 later in this chapter).

For many purposes we will be able to rest content with a stipulated notion of utility, which manages to evade some of these concerns about comparison. However, any full account of (resource) rationality that employs a utility concept will ultimately have to face them. Moreover, crucial for the project of resource rationality, in its cost-theoretic formulation, utilities will need to be commensurate with an appropriate measure of *resource costs* (recall §1.4), in order for characterizations like the fundamental Eqs. (1.1) and (1.2) to make sense.

2.1.4 Expected Utility

What the numbers $u(s, a)$ mean depends not only on their source, but also on how they are going to be used, either by the theorist or directly by the agent. When the state s is known, we assume the best choice a is one that maximizes $u(s, a)$; and in general, the higher $u(s, a)$, the better a is. But what about when s is not known? The most common way of incorporating utility into a theory of decision under uncertainty is to take its *expectation* with respect to the probability on states. The *expected utility* of taking action $a \in A$—written $\mathbb{E}\mathrm{U}(a)$—is defined as the average of $u(s, a)$ where $s \sim p$, that is, where s is drawn with probability $p(s)$:

$$\begin{aligned} \mathbb{E}\mathrm{U}(a) &= \mathbb{E}_{s\sim p} u(s, a) \\ &= \sum_{s\in S} p(s)\, u(s, a). \end{aligned} \tag{2.2}$$

In other words, $\mathbb{E}\mathrm{U}(a)$ is simply a shorthand for this sum over utilities weighted by state probabilities. This quantity offers a plausible sense of "goodness" when the state is unknown but the state probabilities are somehow determined (again, either by the theorist or because they are the "agent's own" subjective probabilities). Thus, as a special case, the *best* an agent can do in a situation of uncertainty is to choose an action that *maximizes* the expected utility.

Remark 1. When measuring actions by their expected utility, we would obtain the same ranking of actions if we took any *positive linear transformation* of the utilities. That is, if we transformed the utility function u by setting $u^*(s, a) = r \cdot u(s, a) + b$ for some $r > 0$ and $b \in \mathbb{R}$, then the ranking of actions by expected utility will be the same whether we take u or u^*. Consequently, in some settings it will sometimes be claimed that utilities are only *meaningful* "up to" positive linear transformation (e.g., Stevens 1946; Narens and Skyrms 2020).

[2] In the context of saving for the future, Ramsey (1928) wrote that any kind of discounting was "ethically indefensible and arises merely from the weakness of the imagination" (p. 543).

Why should we take expected utility as a yardstick for rational choice? One response is simply to insist that this is "rational bedrock" (Briggs, 2019). As Lewis (1974) famously put it, expected utility theory "is the very core of our common-sense theory of persons, dissected out and elegantly systematized" (p. 338). A similar stance is adopted in much of science and engineering. For instance, in a popular textbook on machine learning, where probabilities and utilities (or "loss functions") are often made explicit and typically imposed by the theorist or practitioner, Murphy (2012) characterizes expected utility maximization as "the essence of what we mean by rational behavior" (p. 177).[3]

A substantial tradition in statistics, economics, and philosophy has nevertheless endeavored to justify expected utility maximization without assuming probabilities and utilities as primitives (either of the agent or of the theorist). Rather, the starting point is an agent's *preferences* between lotteries or gambles. As already mentioned, granting probabilities known to the agent, von Neumann and Morgenstern (1953) showed how to derive utilities from a few key axioms. For instance, a continuity axiom states that if X is preferred to Y and Y is preferred to Z, then there is some lottery between X and Z (with probability p of X and $1-p$ of Z) that will make the agent indifferent between that lottery and a sure outcome of Y. The von Neumann and Morgenstern results shows that, a long as an agent's choices obey the axioms, they are behaving *as though* they are maximizing the expectation of some utility function (unique up to positive linear transformation; recall Remark 1).

A more ambitious project is to show that both utilities and (subjective) probabilities can be derived from rational choice behavior. Perhaps the most celebrated attempt is that of Savage (1954), whose representation theorem not only recovers probabilities, but does so uniquely.[4] In addition to the set S of states—the objects of uncertainty—Savage takes as primitive a set of *consequences*—the objects of desirability. Preferences are then over *acts*, viz. functions from states to consequences. Our set A of actions could be understood in these terms: each a defines a function from s to consequences, and $u(s,a)$ measures how good the corresponding consequence is. The force of taking consequences as a primitive, however, is to guarantee a very large and rich space of acts. For instance, Savage assumed that for each consequence there is a "constant act" that has that same consequence no matter what the world is like. Though there are ways of curtailing these richness assumptions (e.g., Gaifman and Liu 2018), the objects of preference will have to be exceedingly rich in order for the theorem to go through.

This is to say nothing of the axioms themselves, whose status as postulates of rational choice have been hotly contested. The Sure Thing Principle is a notable example. Suppose $E \subseteq S$ is an event (that is, a set of states), and suppose an agent prefers one act to another in case E occurs, and also in case E does not occur. Then the agent should have the same preference without knowing anything about E. Putative counterexamples to this and other principles abound (Briggs, 2019; Buchak, 2013, 2022), which in turn have motivated alternative axiomatizations.

2.1.5 Actions and Causes

Although we will not be engaged in debates around axioms for deriving probabilities and utilities, it is worth highlighting an important point of controversy related to the Sure Thing Principle in particular. This axiom is obviously unsound when E's occurrence is in some way

[3]Practice in cognitive science reflects the same stance. For example, in research programs that compare alternative heuristic inference methods, assessment is still given by expected (viz. average) accuracy (Gigerenzer et al., 2000; Gigerenzer and Brighton, 2009). The relevant distributions on problems are not necessarily accessible to the experimental participants, but expected accuracy is the yardstick used to measure degrees of "ecological" rationality. See Okasha (2016) and Thoma (2021) for different sources of doubts about this stance.

[4]Savage's approach was anticipated by earlier work of Ramsey (1931).

dependent on the agent's choice, and this reverberates even into our thin interpretation of decision theory. Consider the following example (from Ahmed 2021, p. 20):

Example 3 (Driving Test)**.** You will have a driver's test next week, and you want to pass. In the meantime you can purchase a lesson, and you know that most people who take the lesson pass, while most people who do not take the lesson fail.

The Sure Thing Principle would apparently have you reason as follows. If you know you will pass, you would rather not pay for the lesson. If you know you will fail, you would still rather not pay for the lesson. And after the fact, no matter how the world turns out, holding either outcome fixed, you would rather not have paid for the lesson. So you should not pay.

Nonetheless, common sense insists that it may well be rational to pay for the lesson, simply because the lesson could help *bring about* the more desirable state of passing.

The example brings to the fore the fact that states and actions are themselves embedded in a larger context that does not always make its way explicitly into the formalism. This wider context can matter. Example 3 illustrates the possibility that the state depends *causally* on which action is chosen. There may also be dependence between the state and the action that is not directly causal. Such examples arise routinely in multi-agent scenarios, as in the following famous illustration from (von Neumann and Morgenstern, 1953, p. 177).

Example 4 (Holmes and Moriarty)**.** Holmes is traveling by train from London to Dover, but he is being pursued by Moriarty, who is on the same train. Moriarty wants to disembark at the same station as Holmes, while Holmes wants the opposite. There is exactly one stop between London and Dover, namely Canterbury. The choice facing both of them is whether to disembark at Canterbury or wait until Dover.

Holmes knows that Moriarty is very clever and, from many past observations, is good at predicting what Holmes will do. Should Holmes get off at Canterbury or not?

Deliberative instability threatens. Merely finding himself inclined toward waiting until Dover seems to provide information about what Moriarty will do, since he knows Moriarty will have anticipated this very inclination. Revising the intention instead to exit at Canterbury does not help, since Moriarty will plausibly have predicted this shift as well. The concern is that Moriarty may always be one mental step ahead.

Indeed, Holmes's attempt (and perhaps also our attempt as theorists) at assessing the probabilities of Moriarty's actions seems to be tied up in some way with Holmes's own deliberation about what he himself will do. How then can Holmes hope to maximize any expected utility?

In this example—typical of scenarios studied in game theory (§3.1; see Ex. 10)—the relevant state is in fact the choice of another agent. And while we may assume that Holmes's choice does not exert a causal influence on Moriarty's choice, it may seem to provide information about it (presumably because of some upstream "common cause," such as an underlying disposition of Holmes about which Moriarty has sufficient information). Many such examples have been discussed in the literature, including Newcomb's problem, the Prisoner's Dilemma (Example 11 below), and many more (see, e.g., Savage 1954; Jeffrey 1965; Skyrms 1980; Joyce 2009). Researchers in reinforcement learning have been concerned with similar issues of so-called confounding (e.g., Bareinboim et al. 2015; Namkoong et al. 2020).

The idea that decisions—and deliberative styles and strategies more generally—can be subject to causal influence by "external" and indeed measurable factors is what makes the psychology of decision making possible at all: to some extent decisions are predictable. Nonetheless, the very context of deliberation seems to involve a presupposition that the deliberating agent may somehow (and at least to some degree) override whatever factors make decisions partly predictable. This tension is nicely summarized by Meek and Glymour:

> One may view decisions, one's own or another's, as the result of a dual system with a default part and an extraordinary part—the default part subject to causes that may also influence the outcome through another mechanism, but the extraordinary part not so influenced and having the power to intervene and displace or modify the productions of the default part. (Meek and Glymour, 1994, p. 1007)

At issue is a fundamental question about how we understand the nature of *choice*. Among the most prominent ways of responding to this set of issues are:

1. Always devise a space S of possibilities so that there will be no (probabilistic or causal) dependence between actions and states (Savage, 1954), for instance by ensuring that S is sufficiently fine-grained (Skyrms, 1980). Ensuring this even in moderately sized problems often requires very large and rich state spaces (cf. Joyce 2009).

2. Assume that the probability distribution is over states and actions together, so that conditional probabilities like $p(s|a)$ are defined. Following Jeffrey (1965), *evidential decision theorists* advocate maximizing the conditional probability of reward, following the adage, "Do what you most want to learn that you will do" (Ahmed, 2021, p. 7).

3. Make the causal structure in a decision problem explicit and construe (at least some) actions as *causal interventions* on that underlying structure (Meek and Glymour, 1994). The theoretical injunction is to clarify precisely how an outcome may depend on an action and how actions and outcomes may both depend on common causes (Hitchcock, 2016).

This in no way exhausts the approaches one can find in the philosophical literature. While much of what follows will remain neutral on how we deal with issues of action-state dependence, it is worth noticing that these issues are not unrelated to resource limitations. For instance, in Example 4, Holmes's predicament appears to stem from a fear of being *outwitted* by Moriarty. Potential richness of the state space, and the possibility that an agent may not be able to introspect on the causes of their decisions are just two more examples. These questions about action-state dependence will thus remain in the background throughout this Element.

2.1.6 Alternatives to Expectation Maximization

Whereas all of the aforementioned approaches to decision theory agree that some kind of expected utility is the relevant quantity for rational assessment, there has also been considerable dissent from this assumption. One type of objection takes issue with one of the basic building blocks. Incommensurability in utility was already mentioned above in §2.1.3; others have argued that ambiguity and imprecision in uncertainty call for alternatives to standard probability representations. Both of these—individually and in concert—have resulted in alternative decision theories. For instance, if we lack a probability function on states, then an agent might pick an action a that *maximizes* the *minimum* utility over all possible states:

$$\max_{a \in A} \min_{s \in S} u(s, a), \tag{2.3}$$

or one that minimizes the maximum *regret* over all states (Savage, 1954, Chapter 9):

$$\min_{a \in A} \max_{s \in S} \max_{a^* \in A} \big[u(s, a^*) - u(s, a)\big], \tag{2.4}$$

These "maximin" and "minimax regret" principles will return below in the discussions of reinforcement learning (§2.3) and game theory (§3.3). Buchak 2022 reviews many other approaches to dealing with the absence of (or generalizations of) probability and utility.

However, even if we admit the same building blocks—a probability p on states (with likelihood l), and a utility u on state-action pairs—concerns have been raised as to whether taking an arithmetical expectation, Eq. (2.2), is the right way of combining these ingredients.

Many of the concerns are motivated by problems of *risk*. For various reasons, losses may loom larger than gains. Suppose, for instance, utility is given by fitness, viz. number of offspring. Should we understand a concept like "adaptiveness" as *expected* fitness, that is, expected number of offspring? A potential danger with taking an arithmetical expectation is that it is not sufficiently sensitive to the most disastrous possible outcome, namely extinction. Consider this simple example from Sober (2001) (see also Gillespie 1977):

Example 5. When a certain bird builds a nest, it has 0.1 chance of that nest evading predators. The bird has two options: (1) lay 10 eggs in a single nest, or (2) lay 5 eggs in each of two different nests. It is easy to calculate that the (arithmetical) expected number of eggs is the same under both options, namely, 1. But they have different probabilities of going extinct. Under option (1) the bird's line dies with probability 0.9, where under option (2) the probability is only 0.81. If we imagine the progeny of this bird (if any survive) carrying out the same strategy over generations, it is easy to show that (2) has a higher chance of long-term survival.

The idea is that adaptation depends not just on the first moment of a distribution (the mean), but also on the second moment (the variance). Following an early suggestion by Bernoulli, a number of authors have advocated taking not the arithmetical expectation, but the *geometric* expectation (Gillespie, 1977; Orr, 2007), which is approximated by $\mu - \mathbb{v}/(2\mu)$, where μ is the mean (e.g., expected utility) and $\mathbb{v}$ is the variance. Penalizing variance in this way provides an alternative mode of combination, one that emphasizes not only the usual (arithmetical) expected utility, but also the degree of dispersion from it across different outcomes. In sequential problems (see §2.2), this is meant to privilege a kind of stability in outcome.

A different way of incorporating risk is not to change the type of expectation, but rather to augment the set of primitives (probability and utility functions) with an additional *risk function*. A particularly well-developed version of this account is Buchak (2013), who allows, e.g., a very desirable but very unlikely outcome to weigh either more or less than it would in the usual expected utility calculation. Debates continue between advocates and critics of expected utility theory, though we will mostly rely on classical expected utility in this Element (the exceptions being in our discussion of learning, §2.3.3, and of game theory, §3.1). Notably, many of the critiques of alternatives—including both of risk-weighted utility theory (Briggs, 2015; Thoma, 2019) and of geometric expectation (Samuelson, 1971)—pertain to the setting of *sequential decisions*, sometimes also known as "dynamic choice."

2.1.7 Dynamic Choice and Preference Change

Before introducing the formal framework for sequential decision making—one that will be applicable to a diverse array of agents, including those that do not engage in any kind of deliberation—it is worth considering how such problems might "look" from the perspective of a deliberating agent. There is an obvious connection between sequential decisions and one-off decisions: in any sequential problem there will be some *first* choice point where a decision has to be made. Like in the multi-agent setting (cf. Example 4 above, and §3.1 below), the quality of a single decision can depend on choices made by other agents; in this case it will be choices made by the same agent, only at a different time.

Given this *dynamic* aspect of sequential choice, how should an agent make that first choice? The notion of an optimal strategy in a sequential decision—to be formally defined below in Eq. (2.9)—implicitly assumes that the best action now is the one that achieves the optimal

balance between immediate utility and (possibly discounted) future rewards, *assuming that you will continue to pursue this optimal balance at future choice points*, given your current view of future choice points. This is quite explicit in the so-called Q-value introduced later in Eq. (2.11). Such a "naïve" (Steele, 2010) assumption may be problematic for at least two reasons.

The first is simply that it may be overly optimistic about the agent's future computational abilities, the very topic of this Element. For instance, a person could plan the shortest and most efficient path from the train station to City Hall. But when it comes time to execute that plan, if it involves a complex pattern of twists and turns, the agent's attempt to follow it may instead leave them lost. In other words, even independent of the difficulties in formulating good plans or strategies, carrying out an optimal strategy may itself be too costly. This will be a main theme of the next several chapters, so we leave it to the side here.

A second reason to question future optimality is that one's impression of what is better or worse may change. This intuitive example comes from Thoma (2018):

Example 6. During a coffee break from work you would like to watch a television episode. It would not be good overall, you currently judge, to watch more than one episode. However, you know that once you have watched the first, you will prefer to watch a second than to return to work. Should you watch the first episode or not?

One response to this type of challenge—an approach known as *sophisticated choice* (e.g., Steele 2010)—is to say that the person has two options: (1) watch the first episode, but simultaneously implement a (potentially costly) measure to guarantee that there is no possibility of watching the second; or if that is not feasible, (2) do not watch the first episode. The obvious challenge for sophisticated choice is the realization that there is a sequence of actions that, at *any* given point, they would prefer to the sequence they in fact pursue: simply watch the first episode and then stop before the second, without any costly intervention guaranteeing the latter. (Notably, this structure is shared with iterated prisoner's dilemmas; see Example 11).

An alternative response that attempts to make good on this intuition—known as *resolute choice* (e.g., McClennen 1990; Gauthier 1994)—is to permit watching the first episode, so long as the person pre-commits to a plan of stopping after the first episode. In other words, rather than rely on some external commitment device, simply exercise your own will. This of course assumes that the individual has the cognitive resources necessary to promote "reason above passion" (Hume, 1739, II.iii.3). But an even more basic challenge for the proponent of resolute choice is to clarify why, at the point of decision, it is rationally required—or indeed even permitted—to follow through with the plan in the first place. By assumption, after the first episode the person judges it *overall better for them* to watch a second episode than to return to work. What rational force does a plan devised by a previous version of oneself have over one's current choices, and why is insistence that one follow the plan anything other than irrational "rule workship" (Smart, 1956; Bratman, 2018)?

A different, though related set of puzzles arises from the possibility that some decisions may change what one values in a more fundamental, long-term manner. For instance, the decision to give birth can alter what one comes to find better or worse, including in ways that cannot be adequately anticipated beforehand (Ullmann-Margalit, 2006). The challenge here is thus also epistemic: not only does one need to account for future changes in utility assessment, there is a problem of imagining what those future utilities will be in the first place (Paul, 2014).

These various challenges related to dynamic choice may motivate incorporating additional components into the decision theoretic formalism introduced so far, viz. time-relative utilities, "planning states" as a basic component of an agent, and so on. It is important to emphasize, however, that this does not by itself answer the fundamental philosophical questions. We still face further issues such as how—at the time of decision—one ought to aggregate past and future

utilities (cf. Pettigrew 2019), or how—again at the time of decision—a plan constrains what one ought to do at that time (cf. Bratman 2018 and §6.4).

At issue are substantive questions about what it means for things to "go well" for an agent across an extended period of time, and specifically how this relates to the agent's own subjective states of *valuing* at each particular point in time. Though we will not attempt to answer these substantive questions here, in Chapter 6 we will address some such considerations that pertain to—or perhaps even emerge from—resources consideration in particular (§6.4).

Putting aside these subtleties about dynamic choice, and how it might look to an agent sophisticated enough to think about it, we now move on to the standard framework we will employ throughout this Element for assessing sequential behaviors.

2.2 Sequential Decisions

In the formulation of one-off decision problems, there is one choice to be made, namely, a selection from A. Then, depending on the state s, some utility $u(s, a)$ is obtained. Yet much of what is challenging about decision making stems from the need to make a whole series of decisions over time. Given, as before, a set S of possible states and a set A of possible actions, we assume an interaction between the agent and its environment produces a *history* h:

$$h \quad = \quad s_0, a_0, \ldots, s_{t-1}, a_{t-1}, s_t$$

of length $2t+1$, with each of the t actions a_i producing a new state s_{i+1}. We of course allow that some "actions" may amount to the agent doing nothing, in which case the state may change without any intervention on the part of the agent. Let $\mathcal{H}_t$ be the set of all histories of length $2t + 1$, and let $\mathcal{H} = \bigcup_t \mathcal{H}_t$ be the set of all histories.

To capture the environmental dynamics, we assume there is some *transition function* $q : \mathcal{H} \times A \to \Delta(S)$, with $q_{h,a}(s)$ the probability of entering state s after history h when the agent performs action a. The environment begins in state s with probability $q_0(s)$. This leads to:

Definition 1. A *sequential decision problem* is a 4-tuple (S, A, q, u):

- S is a set of states;
- A is a set of actions;
- $q_{h,a}(s)$ gives the probability of reaching s in one step following history h when a is taken; meanwhile $q_0(s)$ is the probability of initially being in s;
- $u(h, a)$ is the utility, or reward, obtained from taking action a following history h.

In this simple setting there is no uncertainty about what state the agent is in or about what history has taken place. (Uncertainy about the state will be introduced in §2.4, while uncertainty about the history will appear when we put limits on available memory; see §3.4.) In particular, the utilities of all actions are transparent. The only uncertainty is about what state will result from a given action. This already introduces a distinctive strategic component: choosing worse (viz. lower utility) options now may lead to the possibility of better actions later on, and perhaps greater cumulative utility. In the sequential setting we assess not just individual actions, or even sequences of actions, but instead entire *strategies*. We might think of a strategy as a behavioral disposition, or as a "policy."

Definition 2. A *strategy* is a function $\sigma : \mathcal{H} \to \Delta(A)$, specifying a distribution on actions for every possible history. We write $\sigma_h(a)$ for the probability of action a at history h.

As an agent implements a strategy in a sequential problem, there are two potential sources of nondeterminism: the transition function and any randomness in action selection. Putting these together, for a given number of steps T, they induce a probability distribution $\mathbf{P}_T^\sigma$ on histories in $\mathcal{H}_T$. As a piece of notation, for $h \in \mathcal{H}_T$ and $t < T$, let h_t be the initial segment of h that stops at s_t; thus, $h_t \in \mathcal{H}_t$. Then the distribution $\mathbf{P}_T^\sigma$ is defined as follows.

$$\mathbf{P}_T^\sigma(h) \quad = \quad q_0(s_0) \cdot \prod_{t=0}^{T-1} \sigma_{h_t}(a_t) \cdot q_{h_t,a_t}(s_{t+1}). \tag{2.5}$$

How are we to assess a strategy? If only a finite sequence (i.e., a history) is generated (with probability 1), then we could simply add up the utilities obtained along that history:

$$U(h) \quad = \quad \sum_{t=0}^{T-1} u(h_t, a_t). \tag{2.6}$$

This amounts to assuming that every "time slice" of the agent is weighed equally. Such an assumption might be especially sensible in the context of externally and holistically evaluating "how well things go" for an agent over an entire (finite) lifespan.[5] since there may be no distinguished time of evaluation. It may be worth noting, empirical studies suggest that, when reflecting abstractly on how a good life looks, people actually prefer that things become better over time rather than worse. Thus, realistically, the order of outcomes may matter.

For a variety of reasons, decision theorists have found it useful to introduce *temporal discounting* into the framework, particularly when we think of there being a "now"—a so-called anchor point—the very first decision in a series of future decisions. Rather than emphasizing improvement over time, timepoints in the future actually receive *less* weight relative to the anchor. The assumption is that there is a *discount* $\gamma > 0$, such that:

$$U(h) \quad = \quad \sum_{t=0}^{T-1} \gamma^t u(h_t, a_t) \tag{2.7}$$

Of course, Eq. (2.6) is the special case where $\gamma = 1$. But if $\gamma < 1$, then future utilities count less toward the sum. One interpretation of γ is that it specified the probability of continuing on to another round of decision making.[6] Another is as a "measure of impatience" (Aumann, 1981), consistent with the empirical finding that people care less about future rewards than about immediate rewards. Though this "exponential discounting" expression does not perfectly match empirical behavior (Frederick et al., 2002), it is a common and convenient assumption.

Whether we adopt Eq. (2.6), (2.7), or some other means of assessing histories, we can use Eqs. (2.5) and (2.6) to calculate the T-step *expected* utility for strategy σ:

$$\mathbf{U}_T(\sigma) \quad = \quad \mathbb{E}_{h \sim \mathbf{P}_T^\sigma} U(h). \tag{2.8}$$

This is just like the expected utility measure in the non-sequential case (Eq. (2.2)), but the expectation now is with respect to a distribution defined on entire histories.

In some settings we may not want to impose any specific finite bound on the length of time. As long as $\gamma < 1$, the T-step expected utilities will converge as T increases, so we can assess a strategy σ in general by taking the limit:

$$\mathbf{U}(\sigma) \quad = \quad \lim_{T \to \infty} \mathbf{U}_T(\sigma). \tag{2.9}$$

[5] In support of this suggestion, Rawls writes, "Rationality requires an impartial concern for all parts of our life. The mere difference of location in time, of something's being earlier or later, is not a rational ground for having more or less regard for it" (Rawls, 1971, p. 293). Cf. also Fn. 2.

[6] However, under this interpretation we would need to add a term, $1 - \gamma$, for the stopping probability.

Naturally, Eq. (2.9) will always be well-defined if the process terminates with probability 1, even if $\gamma = 1$ (no discounting). We thus take Eq. (2.9) as the official assessment measure for strategies in sequential decision problems.

Remark 2 (Strategies versus Plans)**.** While strategies specify an action for every possible history, there is no assumption that an agent actually maintain or encode a strategy in any explicit way. A strategy is rather an abstract characterization of an agent's behavioral dispositions. No matter which history h comes to pass, the agent will do something (or nothing) at that point, and $\sigma_h(a)$ specifies the probability of the agent actualizing a.

Of course, many agents maintain a mental economy that does involve explicit intentions to perform specifies actions at future times. This mental state of *planning* is usually understood to be highly partial (Bratman, 1987, Chapter 3), far less detailed than what is required to specify a strategy in the sense of Def. 2 (cf. Zhi-Xuan et al. 2020 for relevant recent computational work). Partiality may be advantageous if deliberation can be postponed to a later point, and it may be resource-efficient to the extent that there is no need to plan for contingencies that (the agent believes) are not going to happen. Plans in this sense fit into the framework in a different way, namely as part of an agent architecture, with operations over plans represented in the theory by programs (recall §1.2.2, and see §6.4 below for more).

Returning to the technical apparatus of sequential decision making, a particularly important setting for sequential problems is when utilities and the next state do not depend on the entire history, but only on the action and the *current state*. This "Markovian" assumption (Bellman, 1957) leads to a simplification of Def. 1:

Definition 3 (MDP)**.** A *Markov decision process* is a sequential decision problem (S, A, q, u) in which $q_{h,a} = q_{h',a}$ and $u(h, a) = u(h', a)$, whenever h, h' agree on their last state. When dealing with MDPs we simply write $q_{s,a}$ instead of $q_{h,a}$, and likewise $u(s, a)$ instead of $u(h, a)$.

The Markov assumption may appear to be a limitation. After all, what is good for an agent at a particular time may depend on what happened at an earlier time. For instance, one might be concerned with whether a current choice is consistent with a pattern of past choices.[7] Yet there is an important sense in which the Markov assumption by itself involves no loss in generality. The reason is that we can always convert a sequential decision problem into an MDP simply by letting the states of the MDP encompass all possible histories in the sequential decision problem (cf. Sutton and Barto 1998, Section 3.5).

Correspondingly, we can simplify the notion of a strategy, so that it too depends only on the current state rather than the entire history:

Definition 4 (Stationary Strategy)**.** A strategy σ is *stationary* if it depends only on the current state. Assuming stationarity, a strategy is a function $\sigma : S \to A$, and we simply write σ_s.

We state without proof the following landmark result:

Theorem 1 (Blackwell 1970)**.** If an MDP admits any optimal strategy at all, then there is an optimal strategy that is stationary.

While the conversion of any sequential decision problem into an MDP—and with it, the restriction to stationary strategies—is thus possible in principle, the conversion does result in a much larger state space, and therefore potentially more complex strategies. How we represent the decision problem may impact how we capture resource costs:

[7] Perhaps one's life going well inherently involves a certain consistency in choice across time, for example.

Remark 3 (Preview of Strategic Cost)**.** We will encounter two ways of imposing resource costs on a strategy, both of which involve presumptive limitations on *memory*.

In the general setting of sequential problems, we could study the degree to which a strategy depends on history: do we need to remember what happened many steps back, or do we only need to look at the current state? We study this question in the next chapter using automata theory (see Def. 11). Since stationary strategies have minimal cost in this setting, Theorem 1 shows that MDPs always have optimal strategies with minimal cost.

However, there is a second approach to strategy cost that penalizes not dependence on history, but sensitivity to a diversity of states. If there are many relevantly different states in an MDP, it may require substantial resources to respond differentially to all of them. One prominent approach to formalizing this idea uses information theory (see Def. 14, 15).

Evidently, when we convert a sequential decision problem into an MDP, we may trade in the first type of complexity for the second: though our strategy will no longer be history-dependent, it must be defined relative to a significantly larger state space.

A theoretical advantage of working with MDPs and assuming stationarity is that we can formulate an elegant recursive specification of the assessment function in Eq. (2.9) above:

Remark 4 (The Bellman Equation)**.** Given a strategy σ, we might be interested in the value of being in a state s. This leads to the recursive Bellman equation (after Bellman 1957):

$$\mathbf{U}^{\sigma}(s) \quad = \quad \mathbb{E}_{a\sim\sigma_s}\Big[u(s,a) + \gamma\, \mathbb{E}_{s'\sim q_{s,a}}\mathbf{U}^{\sigma}(s')\Big], \tag{2.10}$$

such that we recover Eq. (2.9) as $\mathbf{U}(\sigma) = \mathbb{E}_{s\sim q_0}\mathbf{U}^{\sigma}(s)$, so long as $\mathbf{U}(\sigma)$ is well-defined.

Another useful fact for MDPs with stationary strategies is that, under common assumptions, they together define a "stationary distribution" (not to be confused with a stationary strategy):

Remark 5 (Stationary Distribution)**.** As a technical point—which can be skipped here but will be invoked in later discussions (§4.3.4, §5.3.3)—the strategy and transition functions together define a so-called stationary distribution $P_{\sigma}(s)$ specifying the "long-term" probability of visiting state s. Suppose the states in S are listed as $s_1, s_2, \ldots$. Then σ and q give a transition matrix $\mathbf{T}$, such that $\mathbf{T}[i,j] = \sum_a \sigma_{s_i}(a)\, q_{s_i,a}(s_j)$. This is the average probability (given strategy σ) of moving from state s_i to state s_j in one step.

A stationary distribution is a "fixed point" vector $\mathbf{s}$ satisfying $\mathbf{sT} = \mathbf{s}$. Under relatively minimal conditions, such a distribution exists uniquely. For example, as long as every state is reachable from every other state with positive probability ("irreducibility"), and for every state s there is no number k such that the system will return to s in k steps with probability 1 ("aperiodicity"), this is guaranteed. (See, e.g., Chapter 3, Theorem 1.3 of Karlin and Taylor 1975.) We assume both of these are satisfied, which justifies reference to *the* stationary distribution, whereby $P_{\sigma}(s_i) = \mathbf{s}[i]$ for the stationary distribution vector $\mathbf{s}$.

If we know all of the relevant parameters in an MDP, solving for the optimal policy can actually be reasonably tractable (Papadimitriou and Tsitsiklis, 1987), assuming the spaces of states S and actions A are not too large. However, we the theorists—not to mention the agents themselves—often do not know all the details of the environment. In this case, we need to study not just strategies, but learning methods for inducing strategies.

2.3 Reinforcement Learning

Imagine an agent introduced to a new environment in which the relevant features (e.g., the transition probabilities q or utilities u) are not known or previously experienced. How might

the agent learn to flourish in this new environment? And how might we—at the theoretical level—assess different approaches for learning? What makes a learning procedure better or worse? In discussing these questions we will assume the environment is Markovian (Def. 3). While the task is therefore to learn an optimal stationary strategy (Def. 4), an agent will typically not employ a stationary strategy during the course of learning. The whole point of learning is that the strategy may change with experience—it will thus be history dependent.

There are two families of approaches to reinforcement learning, *model-free* and *model-based*, and we discuss each in turn.

2.3.1 Model Free Methods: Q-Learning

Model-free approaches are quite simple, and date back to the "law of effect" introduced in psychology by Thorndike (1911). All the agent needs to do is learn to take the right actions in the right states so that expected cumulative utility (Eq. (2.9)) is maximized. In other words, it suffices to encode a "cached" value for each state-action pair. This idea of cached values leads to a very simple model-free method known as *Q-learning*. Intuitively, the value of taking a in s equals the immediate reward $u(s,a)$ plus the expected future reward from that point on. Fixing a strategy σ, we define the *Q-value* to be:

$$Q^\sigma(s,a) \quad = \quad u(s,a) + \gamma\, \mathbb{E}_{s' \sim q_{s,a}} \mathbf{U}^\sigma(s'), \tag{2.11}$$

where $\mathbf{U}^\sigma(s')$ is given by the Bellman Equation (2.10). Note that the *best possible* Q-values—those resulting from an optimal strategy σ^*—will then satisfy another recursion, with $Q^* = Q^{\sigma^*}$:

$$Q^*(s,a) \quad = \quad u(s,a) + \gamma \big[\mathbb{E}_{s' \sim q_{s,a}} \max_{a' \in A} Q^*(s',a') \big]. \tag{2.12}$$

That is, taking a in s is as good as its immediate reward and the best one could do from that point forward. This insight is the basis for the Q-learning algorithm, which iteratively updates estimates $\widetilde{Q}_t$ of the optimal function Q^*, essentially following Eq. (2.12). Suppose at step $t+1$ that the agent is in state s, takes action a, and moves on to a next state s' (drawn from $q_{s,a}$). Then we update the previous estimate $\widetilde{Q}_t(s,a)$ by:

$$\widetilde{Q}_{t+1}(s,a) \quad \leftarrow \quad (1-\eta)\, \widetilde{Q}_t(s,a) + \eta \big[u(s,a) + \gamma \max_{a' \in A} \widetilde{Q}_t(s',a') \big], \tag{2.13}$$

for some "learning rate" $\eta \in (0,1)$ that determines how quickly we move away from the previous estimate. Under minimal conditions, as long as each action is taken at each state infinitely often, it can be shown that $\widetilde{Q}_t$ converges to the optimal Q^* as t increases (e.g., Jaakkola et al. 1994). So any Q-learning agent will eventually be acting optimally, simply by choosing the action that has the highest Q-value for any given state.

Remark 6 (Deep Reinforcement Learning)**.** Recent successes in learning to play complex games—including Go, chess, and a range of Atari video games—involve an approximate Q-learning algorithm known as *deep Q-learning* (Mnih et al., 2015; Silver et al., 2018). Rather than encoding an entire table of values $Q(s,a)$ for all s and a, these values can instead be estimated using a pair of (convolutional) neural networks.

A remarkable feature of Q-learning is that it works no matter how the actions are chosen during learning (it is "off policy" in the reinforcement learning jargon). But an agent in a real environment might want to learn quickly while also performing as well as possible during the learning process. So we would like to say something about how the agent should behave during learning. This leads to the well-known tension between *exploration* and *exploitation*:

2.3.2 Exploration/Exploitation Tradeoff

After observing the current state s at stage t in the learning process, the agent needs to choose a next action. How should this action be chosen? One obvious possibility is to pick a that maximizes the current estimate $\widetilde{Q}_t(s, a)$. Given past experience, a may have led to the best outcomes. This is pure "exploitation" of known rewards.

However, there may well be actions that would be significantly better than a, but that the agent has not yet tried. Pure "exploration" would involve picking the least known action, for instance, the one that has been attempted in this state the least number of times, or perhaps by choosing an action totally at random.

The exploration/exploitation tradeoff arises from the need to balance these two considerations: ensuring that the agent does well enough while still learning, and ensuring that learning happens efficiently. This fundamental tension, which has been called a "basic feature of the human condition" (Huttegger, 2017, p. 36), has prompted a variety of approaches:

Example 7 (Random Exploration: Softmax)**.** Suppose the current estimate of a is $V(a)$. For instance, we might have $V(a) = \widetilde{Q}_t(s, a)$ for a Q-learning agent. Then the *softmax* (or *Boltzmann*) decision rule (Bishop, 1995; Sutton and Barto, 1998) selects action a with probability:

$$\frac{e^{\beta V(a)}}{\sum_{a' \in A} e^{\beta V(a')}}, \tag{2.14}$$

where β is a so-called *inverse temperature* parameter (see §4.3.3 for the source of the nomenclature). As β goes to 0, the rule comes ever closer to complete randomization (high exploration); as β increases, the rule comes ever closer to maximizing estimated value (high exploitation). We will revisit this rule below in Chapter 4; see especially §4.1.2.

Example 8 (Uncertainty-Directed Exploration: Upper Confidence Bound)**.** An alternative to random exploration is to alter the update rule in (2.13) by adding a premium for actions that have been less explored. Recall that $\widetilde{Q}_{t+1}(s, a)$ is a weighted combination of $\widetilde{Q}_t(s, a)$, the previous estimate, and

$$u(s, a) + \gamma \max_{a' \in A} \widetilde{Q}_t(s', a'). \tag{2.15}$$

If we simply add to (2.15) an "uncertainty premium" ρ as another summand, then we arrive at the family of *upper confidence bound* (UCB) methods. The uncertainty premium ρ is chosen proportional to $1/\sqrt{n}$, where n is the number of times a has been chosen in state s so far. As ρ ensures sufficient exploration, the agent can simply maximize estimated Q-values at each step.

Both softmax exploration and UCB-type methods have been implicated in studies of human choice behavior (Schulz and Gershman, 2019). Common to all approaches to the dilemma is an assumption that an agent should begin with a strong bias toward exploration, and then as time goes on transition more and more toward exploiting. For softmax exploration this amounts to a kind of "annealing" by increasing β; for UCB it amounts to decreasing the uncertainty premium ρ over time. Construing an entire human lifespan as one long sequential decision problem, Gopnik (2020) suggests that *childhood* can be understood in part as nature's solution to a large-scale exploration/exploitation dilemma, insofar as children appear to be in a privileged and extended position to learn and explore in novel environments.

The exploration/exploitation tradeoff is of course deeply related to issues of resource limitations, particularly when we think of exploration as somehow "internal" to the agent's mind (Gopnik, 2020; Aronowitz, 2021), a theme explored further in §6.5. Yet it arises for virtually any learning agent, independent of computational resources. A difficult methodological question is how to assess a learning process, even before worrying about resource costs.

2.3.3 Assessing Learning Methods

If we think of the learning problem as simply one large (meta-)decision problem, and we proceed as we did in §2.1, then we would have to put a prior distribution on a suitable space of MDPs. Construing entire strategies as basic actions, Eq. (2.9) specifies a utility for a pair of an MDP and a choice of strategy. So a Bayesian solution would be to maximize expected utility, where the expectation is now with respect to the distribution on MDPs. It is possible to formulate the problem this way (Bellman, 1957)—cf. Remark 7 below—and intricate methods have been developed for this formulation (Ghavamzadeh et al., 2015). One of the putative advantages of the Bayesian approach is that it solves the exploration/exploitation problem in a principled way: the tradeoff is implicit in whatever sequence of actions maximizes expected utility.[8]

Nevertheless, for a host of reasons, most work in reinforcement learning has not adopted such a formulation. Instead, emphasis has been on two criteria that do not require putting a prior distribution on the space of MDPs:

> **Sample complexity:** In the worst case (among MDPs), how many time steps does it take before the learning agent is within some sufficiently small range of the optimal strategy, with high probability (see, e.g., Strehl et al. 2006)?
>
> **Worst-case regret:** In the worst case (among MDPs), how much worse does the agent perform than the optimal agent? This is a version of minimax regret, Eq. (2.4) above.

To unpack the second criterion, suppose we divide the learning phase into I "episodes" each lasting some finite number of steps. In each episode the agent employs a strategy σ^i based on their experience (e.g., estimated Q-values) from the previous episodes. Then the worst-case regret for a given MDP measures how much worse the agent is performing across all episodes, compared with the optimal strategy σ^* for that MDP:

$$\sum_{i=1}^{I} \big[\mathbf{U}(\sigma^*) - \mathbf{U}(\sigma^i)\big]. \tag{2.16}$$

Because the MDP is not known ahead of time, no learning algorithm can avoid some degree of regret. In fact, fundamental lower bounds on worst-case regret can be derived (Azar et al., 2017; Jin et al., 2018), which scale (sublinearly) with the size of the MDP (S and A) and the number of episodes and steps per episode. In other words, these results show that for *any* algorithm, as the size of the MDP and the number of episodes increases, there will be MDPs that prevent the algorithm from reaching the optimal policy too quickly.

Jin et al. (2018) show that worst-case regret and sample complexity are closely related (see also Dann et al. 2017; Osband and Van Roy 2017), and moreover that specific versions of UCB (Example 8) in fact approach the theoretical optimal performance. Softmax exploration, while conceptually simple, is less well understood theoretically, and in any case appears to be less sample-efficient than more "directed" (deterministic) approaches (Cesa-Bianchi et al., 2017). Of course, it evidently requires even less memory than UCB (since it does not have to store any data about how often actions have been chosen), which raises the possibility that it is nonetheless a *resource* rational approach, balancing statistical and computational efficiency. We consider this question in a broader context (beyond learning) in Chapter 4; see Cor. 8.

[8] It will follow from Theorem 5 below (in §4.2) that such a solution will not involve any randomization at all, thereby ruling out any type of softmax exploration.

2.3.4 Model-Based Methods

Agents that learn in a model-free way, e.g., through Q-learning, are inducing conditioned responses to stimuli, akin to stimulus-response approaches in psychology (e.g., Thorndike 1911): when in state s, perform action a. But what happens if, say, the transition probabilities remain as they were, but the utility structure changes dramatically? In other words, what is good for the agent changes. For this new environment, the learned strategy may be next to useless, and the agent will have to start learning again from scratch.

An alternative approach is to learn not just the estimated value of taking a in s, but actual features of the environment itself, most saliently the transition probabilities (q) and the utility structure (u). To the extent that an agent has internalized these components, if one of them changes, all they need to do is learn about those changes, and then they can engage in some type of *replanning*, taking the changes into account. The idea that humans and other animals learn by building "internal models" was a key motivating idea in the development of cognitive science (Craik, 1943), a theme to which we will return in §6.1.

A model-based approach thus involves explicitly learning about features of the environment.[9] Given some history $h = s_0, a_0, \ldots, s_T$ of interaction, we might suppose the agent maintains a probability $\mathbf{P}(M|h)$ that this history was generated by MDP M. On a Bayesian formulation, this could be derived from an assumed prior distribution $\mathbf{P}(M)$ on MDPs, together with the obvious likelihood $\mathbf{P}(h|M) = q_0(s_0) \cdot \prod_{t=0}^{T-1} q_{s_t,a_t}(s_{t+1})$. Further experience thus allows the agent to refine their uncertainty about the underlying MDP.

Once M has been learned, the agent can solve for the optimal strategy σ^* and behave according to σ^*. But in the course of learning, just as in the model-free setting, the agent faces a tradeoff between exploration and exploitation. Model-based UCB-type algorithms (Example 8) have been studied in this setting as well, and shown to achieve optimal regret bounds (Azar et al., 2017). There are also prominent approaches that rely on random behavior:

Example 9 (Posterior Sampling)**.** One method, knowing as *posterior sampling* or *Thompson sampling* (after Thompson 1933), works as follows. At the beginning of each episode, having observed history h, the agent randomly draws an MDP M with probability $P(M|h)$, solves for the optimal strategy σ^* in M, and then uses σ^* for the remainder of that episode, generating a longer history for the next draw at the start of the next episode.

Theoretical work has shown that this approach also attains optimal regret bounds in a certain sense (Osband and Van Roy, 2017).[10]

When should an agent employ a model-free mode of learning, and when should they invoke an internal model? While a model may be computationally costly to learn, maintain, and use, it facilitates quick and flexible adaptation (cf. Godfrey-Smith 1998). By contrast, model-free approaches demand fewer computational resources, though they may adapt less efficiently.

This is a paradigm example of a tradeoff between efficiency and computational costs. There is evidence that brains in fact employ both types of learning in concert—both a "habitual" model-free system, and "goal-directed" model-based system—and they may come into conflict (Daw et al., 2005). When conflict occurs, and a decision must be made, the system must

[9] A precise distinction between model-free and model-based learning offered in the literature (Strehl et al., 2006; Jin et al., 2018) invokes computational resources, specifically memory: a method is model-free just in case its space requirements scale sublinearly with the space needed to encode the entire MDP. So if the "internal model" were sufficiently compact, it would not actually be counted as model-based learning. Note that this definition is best read as offering necessary conditions for being a model, but not sufficient conditions.

[10] Osband and Van Roy (2017) use a notion of *Bayesian regret* that is just like Eq. (2.16), except that the worst case MDP is replaced by taking an expectation of (2.16) with respect to a prior distribution on MDPs. In anticipation of Chapter 4, this does not show randomization is *necessary* to minimize (Bayesian) regret.

somehow adjudicate between the two. Daw et al. (2005) give evidence that this is done in a statistically efficient way, while Gershman et al. (2014) suggest that there is even more direct interaction between the systems, e.g., whereby the internal model supplies "simulated data" to help the habitual module learn more efficiently. This body of work offers another illustration of resource rational analysis, the topic of Chapter 5.

2.4 Sequential Decisions with Partial Observability

As a final topic in traditional decision theory, we now consider a generalization of Markov decisions processes that allows imperfect observations of the underlying state (Åström, 1965; Kaelbling et al., 1998). We augment the MDP formalism (Def. 3) with a set O of observations, and a likelihood function $l(o|s)$ for the probability of observing $o \in O$ when in state s.

Definition 5 (POMPDP)**.** A *partially observable Markov decision process* is specified by a 6-tuple, (S, A, O, q, l, u):

- S is a set of possible *states*;
- A is a set of possible *actions*;
- O is a set of possible *observations*;
- $q : S \times A \to \Delta(S)$ is a *transition function*, whereby $q_{s,a}(s')$ is the probability of reaching s' when a is taken in s; meanwhile, q_0 is a distribution on initial states;
- $l : S \to \Delta(O)$ is a *likelihood* function, with $l(o|s)$ the probability of observing o in s;
- $u : S \times A \to \mathbb{R}$ is the *utility* function.

Notably, POMDPs have been used in cognitive science to capture intuitive "theory of mind," including with infants (Jara-Ettinger et al., 2016).

In the partially observable setting, since the agent does not perceive states but only noisy indications o of them, let us call a sequence $\omega = o_0, a_0, \ldots, o_{t-1}, a_{t-1}, o_t$ an *observation history*. Let $\mathcal{O}_t$ be the set of all observation histories of length $2t+1$, i.e., those with t actions and $t+1$ observations, and let $\mathcal{O} = \bigcup_t \mathcal{O}_t$ be the set of all observation histories.

Definition 6 (Strategy in a POMDP)**.** A strategy is a function $\sigma : \mathcal{O} \to \Delta(A)$. We write $\sigma_\omega(a)$ for the probability of action a given observation history ω.

Note that, even though the underlying dynamics (represented by q) are assumed here to be Markovian, depending only on the current state, we cannot necessarily restrict attention to stationary policies that depend only on the last observation. Since the agent does not directly observe the state, the whole observation sequence might be relevant.

A strategy σ, with a likelihood function l and a transition function q, still gives a distribution $\mathbf{P}^\sigma_T$ on histories parallel to (2.5), though histories are now slightly more complex objects. In the partially observable setting a history will be a sequence $h = s_0, o_0, a_0, \ldots, s_{t-1}, o_{t-1}, a_{t-1}, s_t, o_t$. Let us write $\omega(h)$ for the subsequence obtained by removing all state occurrences. Thus, $\omega(h) = o_0, a_0, \ldots, o_{t-1}, a_{t-1}, o_t$ is an observation history as defined above.

Given σ, we define a distribution on T-step histories as follows:

$$\mathbf{P}^\sigma_T(h) \quad = \quad q_0(s_0) \cdot l(o_0|s_0) \cdot \prod_{t=0}^{T-1} \Big(\sigma_{\omega(h)\restriction t}(a_t) \cdot q_{s_t,a_t}(s_{t+1}) \cdot l(o_{t+1}|s_{t+1}) \Big) \tag{2.17}$$

which in turn gives a T-step expected utility analogously to Eq. (2.8),

$$\begin{aligned} U(h) &= \sum_{t=0}^{T-1} \gamma^t u(s_t, a_t) \\ \mathbf{U}_T(\sigma) &= \mathbb{E}_{h \sim \mathbf{P}_T^\sigma} U(h). \end{aligned}$$

and finally an overall expected utility just as in (2.9):

$$\mathbf{U}(\sigma) = \lim_{T \to \infty} \mathbf{U}_T(\sigma). \tag{2.18}$$

An optimal policy is a function σ for which $\mathbf{U}(\sigma)$ is maximal.

Remark 7. On a Bayesian formulation (recall §2.3.3), the problem of learning a fixed MDP can be characterized as a POMDP. Suppose, for instance, that the agent knows the states S, the actions A, and the utility function u, but is uncertainty about the transition function.

Then a POMDP representation of the learning problem takes the set of states to be all pairs $\langle q, s \rangle$ where q is a transition function and $s \in S$, the actions A are the same, and observations are the states $O = S$, with likelihood function $l(s|\langle q, s \rangle) = 1$. The new utilities are $u^*(\langle q, s \rangle, a) = u(s, a)$. The transition function q^* in the POMDP is given by $q^*_{\langle q,s \rangle, a}(s') = q_{s,a}(s')$. In other words, the only uncertainty comes from the initial distribution q^*_0, defined so that $q^*_0(\langle q, s \rangle) = q_0(s)\, p(q)$ for some prior p on transition functions.

Solving this POMDP thus produces the optimal learning strategy relative to the prior p on transition functions, again balancing exploration and exploitation in a principled way.

2.4.1 Bayesian Filtering

Given partial observability, it seems that an agent would need—in some way or another—to track (their uncertainty about) the underlying sequence of states that likely generated the observation history $\omega = o_0, a_0, \ldots, o_t, a_t, o_{t+1}$ encountered so far. Doing so in a Bayesian manner—known as *filtering*—leads to the following posterior probabilities:

Proposition 1. The posterior probability $P(s|\omega)$ of state s, given observation sequence ω, is defined by recursion on the length of ω:

$$\begin{aligned} P(s|o_0) &\propto l(o_0|s)\, q_0(s) \\ P(s|o_0, a_0, \ldots, o_t, a_t, o_{t+1}) &\propto l(o_{t+1}|s) \sum_{s'} P(s'|o_0, a_0, \ldots, o_t)\, q_{s', a_t}(s). \end{aligned}$$

The first probability, $P(s|o_0)$, is just the standard posterior after the first observation (cf. Eq. (2.1)). For the recursion we multiply the likelihood of the last observation given s, by the probability of having arrived at s after taking the last action a_t. An optimal solution to a POMDP would generally seem to demand sensitivity to these posterior probabilities specified in Prop. 1. In fact, there is a precise sense in which an agent tracking beliefs in a POMDP can be understood as inhabiting an ordinary MDP, in which the states are *belief states*.

2.4.2 Belief MDPs

If an agent were tracking the posterior distribution on states, following the expressions above in Prop. 1, then we could think of them as traveling between different belief states, depending

on what actions are taken and what observations are made. As Prop. 1 shows, knowing the previous belief state "screens off" the new probability from the rest of the history:

$$\begin{aligned} \text{belief } b,\ \text{action } a &\mapsto \text{new belief } b' \text{ such that } b'(s) = \sum_{s'} b(s')\, q_{s',a}(s) \\ \text{belief } b,\ \text{observation } o &\mapsto \text{new belief } b' \text{ such that } b'(s) \propto b(s)\, l(o|s) \end{aligned}$$

Given a belief state b, let us write $b[a]$ or $b[o]$ for the respective resulting belief state.

This itself defines a kind of dynamics, whereby an agent decides on an action a given their belief state b, which then establishes a new belief state $b[a]$ that can change again with a further observation; and we finally end up where we can repeat the same process but relative to the new belief state $b[a][o]$. What this informal observation shows is that POMDPs can actually be construed as (continuous-state) MDPs, in which the states are the *possible beliefs* of the agent. A variation on the following appears in Åström (1965):

Proposition 2. For every POMDP $M = (S, A, O, q, l, u)$ there corresponds an MDP (as in Definition 3), $M^* = (B, A, q^*, u^*)$, such that:

- The set of states is $B = \Delta(S)$, that is, all probability distributions on S;
- A is the same in M^* as it is in M;
- The transition function $q^* : B \times A \to B$ is defined so that $q^*_{b,a}(b')$ is the probability of there being some observation that leads to belief state b'; that is,[11]

$$q^*_{b,a}(b') \quad = \textstyle\sum_{s,o} \big(b[a](s) \cdot l(o|s) \cdot \mathbb{1}_{b'=b[a][o]}\big).$$

- Finally, the utility $u : B \times A \to \mathbb{R}$ is defined so that $u^*(b, a) = \sum_{s \in S} b(s)\, u(s, a)$; this is just the *expected* utility of a under b (Eq. (2.2)).

Any optimal strategy σ^* for M^* canonically induces an optimal strategy σ for M, whereby $\sigma_\omega = \sigma^*_b$ with b the belief state defined by $b(s) = P(s|\omega)$ as given in Prop. 1.

In contrast to MDPs, it is known that the worst-case computational complexity of solving POMDPs is relatively high (Papadimitriou and Tsitsiklis, 1987), typically intractable in theory and in practice. Unsurprisingly, dealing with continuous-state MDPs is also generally intractable. While the unrestricted framework is attractive in virtue of its significant generality, particular applications—especially when we as theorists actually want to *solve* for the best strategies—may call for simplifications or restrictions.

2.5 Summary

Prior to introduction of considerations of resources or resource costs, we have identified some of the key tools and themes in traditional approaches to decision theory. Despite points of philosophical controversy and contention at nearly every turn, we can identify a relatively canonical set of structures implied by a typical (especially sequential) decision problem. As a summary of core concepts and notation, we have encountered the following components:

[11] Here $\mathbb{1}_\varphi$ is the indicator function, equal to 1 if φ is true, 0 otherwise.

- S: set of states
- A: set of actions
- O: set of observations
- $q, p, P, \mathbf{P}$: a probability distribution
- $\Delta(X)$: set of distributions on X
- $\mathbb{E}_{x\sim p} f(x)$: p-expected value of f
- $u(s, a)$: utility of a at s
- σ: strategy
- $\mathbf{U}(\sigma)$: expected utility of a strategy
- $Q^\sigma(s, a)$: Q-value of a at s
- π: a program
- Π: a set of programs
- $C(\pi)$: cost of a program
- $\mathbf{V}(\pi)$: overall value of a program

These components highlight numerous sources of costs and complexity. We will encounter a range of approaches to characterizing such costs, from those that depend concretely on details of agent architecture (e.g., how representations are encoded), to those that purport to place limits on any physically embodied agent whatsoever.

Chapter 3

Machines Playing Games

As discussed at end of the previous chapter there are numerous sources of complexity in traditional models of decision making. The present chapter presents a line of work originating in Economics in the 1980s that construes *players in games* as resource constrained machines. Rather than looking at the topic from the perspective of a single decision maker, this body of research takes the multi-agent perspective of game theory. A prominent theme here is that familiar results and patterns in standard approaches to game theory look quite different when the individuals are even minimally constrained in what they can compute.

Compared to some other approaches that hew closely to human psychological processes (see especially Chapter 5), this line of work is relatively coarse-grained, studying agents that are human-like only in being subject to very general computational constraints. This alone is sufficient to reveal some subtle—and indeed seemingly human-like—behaviors.

Since the previous chapter dealt only with single-agent decision problems, our first task is to introduce game theory against this background. As a segue to later chapters, this chapter is also tasked with introducing two fundamental concepts from the theory of computation, namely (probabilistic) Turing machines and (probabilistic) finite state automata.

3.1 Games and Decisions

Game theory is the study of rational decision making among groups of agents who may have different desires or preferences. At first blush it may not be clear why multi-agent contexts introduce anything fundamentally new beyond the Bayesian decision theoretic setting covered in the previous chapter.[1] The trouble, of course, is that Holmes's attempt to arrive such an assessment is frustrated by a potential symmetry in the strategic situation, which in turn threatens deliberative instability. After all, Moriarty faces essentially the same decision problem, with the goal of assessing what Holmes is most likely to do. Each assumes the other is shrewd and neither wants to disappoint.

3.1.1 Game Theoretic Equilibrium

The motivating idea behind the theory of games is that these multi-agent scenarios cannot be easily reduced to separate individual-level decision problems—there is something holistic about strategic situations among intelligent agents. Recall that in a single-agent, one-shot

[1] Indeed, we will briefly discuss conservative multiagent extensions to single-agent decision theory in §6.6.

decision problem, we assume a set A of actions, a set S of possible states, and a utility function $u : S \times A \to \mathbb{R}$. In standard game theoretic scenarios we assume that the primary source of uncertainty is what action another agent (or agents) will take. Thus, let $N = \{0, \ldots, n-1\}$ denote a set of n "players," such that each player $i < n$ has a set A_i of possible actions, with $\mathbf{A} = (A_0, \ldots, A_{n-1})$. The utility now depends on what all of the players do; so player i will be associated with a utility function $u_i : \mathbf{A} \to \mathbb{R}$. It will be convenient to assume in most of this chapter that there are only two agents. So a game will be simply a pair $\mathcal{G} = (\mathbf{A}, \mathbf{u})$, where $\mathbf{A} = (A_0, A_1)$ and $\mathbf{u} = (u_1, u_2)$. Here is a very simple rendering of Example 4:

Example 10. Suppose Holmes (player 0) and Moriarty (player 1) both have actions $A_0 = A_1 = \{c, d\}$ (Canterbury and Dover). Their respective utilities can be depicted in a table, with the first number the payoff for 0, the second the payoff for player 1. Actions along the leftmost column are those of 0, while actions on the top row are those of 1:

	c	d
c	$-10, 10$	$5, 0$
d	$10, -5$	$-10, 10$

For instance, if Moriarty disembarks early at Canterbury but Holmes continues to Dover (bottom left), this is especially bad for Moriarty, as he misses Holmes and is stranded in Canterbury.

If Holmes did have a prior probability over Moriarty's actions, then it would make sense for him to maximize expected utility with respect to that prior. For instance, if he judged Moriarty equally likely to choose Canterbury or Dover, then Holmes would stay on until Dover. However, as discussed in §2.1.5, playing against a clever agent introduces the possibility of dependence between states and actions, in which case maximizing expected utility may no longer make sense. For instance, it might seem reasonable to expect Moriarty to avoid Canterbury, since he would not want to be stranded there. The problem is that Moriarty, having anticipated this very expectation, will choose Canterbury precisely to outwit Holmes. And so on.

Just as in Chapter 2, a strategy for player i is a distribution $\sigma_i \in \Delta(A_i)$ over actions A_i. Given a pair $\sigma = (\sigma_0, \sigma_1)$ of strategies for the two players in game $\mathcal{G}$, the expected utility for each player i is a sum of products:

$$\mathbf{U}_i^{\sigma} = \sum_{a_0 \in A_0} \sum_{a_1 \in A_1} \sigma_0(a_0) \times \sigma_1(a_1) \times u_i(a_0, a_1). \tag{3.1}$$

The operative notion in game theory is that of a Nash equilibrium.

Definition 7 (Nash Equilibrium)**.** A pair $\sigma = (\sigma_0, \sigma_1)$ of strategies forms a *Nash Equilibrium* if, for each i and each σ' that differs only on σ_i, we have $\mathbf{U}_i^{\sigma} \geq \mathbf{U}_i^{\sigma'}$.

Nash (1950) famously proved that every finite game has a Nash equilibrium, though in general this result requires "mixed"—that is, randomized—strategies. For instance, it is easy to check that in Example 10, the only Nash equilibrium is fully randomized: Holmes chooses Canterbury with probability 1/3 and Moriarty chooses Canterbury with probability 3/7.

This result is in stark contrast to the setting of individual Bayesian decision making, where it can be shown that there is always a deterministic strategy that maximizes utility. (A strong version is proven in Theorem 5 of Chapter 4.) The question naturally arises of how the pronouncements of decision theory and game theory should be reconciled. That is, how does maximizing expected utility relate to playing a Nash equilibrium strategy?

Some have argued that decision theory should be understood as fundamental, and that equilibrium notions make sense when (and only when) one's subjective probabilities match

those of the equilibrium solution (see, e.g., Kadane and Larkey 1982). In their introduction of modern game theory, von Neumann and Morgenstern (1953) called this the "Robinson Crusoe" model of strategic decision making, since it treats other agent decisions as just another aspect of the world about which one might be uncertain. For instance, if Holmes judges that Moriarty will choose Canterbury with probability 3/7 then it would be acceptable to play his mixed strategy of 1/3 and 2/3. At the same time, given these beliefs, *any* strategy for Holmes would maximize his expected utility (namely $-10/7$), including, say, deterministically opting for Dover. And if he judged Moriarty only minimally more likely to wait until Dover, the only way to maximize expected utility would be to (deterministically) disembark at Canterbury.

Recall, however, that the distinctive challenge of the strategic multi-agent context is precisely how to assess such probabilities. Some have suggested that, at least in certain circumstances with as-good-as-ideal players, equilibrium considerations might furnish a stable prior probability for each of the players (Harsanyi, 1982). On this interpretation the probabilities in a mixed strategy do not describe randomization over actions (Zollman, 2023), but rather the subjective degrees of belief for the other players in the game (Aumann, 1987). In other words, in Example 10 taking Moriarty to choose c with probability 3/7 and d with probability 4/7 identifies the beliefs that Holmes *ought*, by appeal to deliberative symmetry, to have.

There are yet further proposals for ameliorating the potential tension between expected utility maximization and game theoretic equilibrium. For example, in the spirit of resource rationality—specifically an instance of the "minimal rationality" approach described in §1.2.1—Skyrms (1990) shows how very simple learning agents aiming to maximize expected utility may reach game theoretic equilibria by virtue of achieving a kind of "dynamic deliberative" equilibrium. Other work abandons the idea of equilibrium play as a matter of rationality at all, and instead treats game theory as an analytic tool for describing or predicting group-level behavior, for instance in penalty kicks at soccer games (Chiappori et al., 2002) or in the evolved distribution of traits in a large population (Maynard Smith and Price, 1973).

3.1.2 Prisoner's Dilemma

Perhaps the most famous game is one for which decision theory (viz. expected utility maximization) and game theory (Nash equilibrium play) in fact agree:

Example 11 (Prisoner's Dilemma)**.** Two players, $\{0, 1\}$, each have possible actions $\{c, d\}$ (mnemonic for *c*ooperate and *d*efect), with utilities given by the table:

	c	d
c	$3, 3$	$0, 4$
d	$4, 0$	$1, 1$

On one story, after Hume (1739), two farmers stand to benefit from the other helping to harvest their crops. Farmer 0 has a choice of whether to help farmer 1, and 1 has a choice of whether to help 0. Let us suppose (contrary to Hume's discussion) that both choices will be made without knowledge of the other's choice. If both help each other, then they do reasonably well (utility 3). But if, say, 0 helps 1 but not vice versa, then 1 does even better (utility 4).

For both players choice d "dominates" choice c, in the sense that it gives a superior outcome no matter what the other does. The only Nash equilibrium in this game is thus the pair of pure strategies according to which both players defect. This is also the only strategy that maximizes expected utility, relative to any prior probability on the other's actions.

Games like the prisoner's dilemma have raised questions about the rational pronouncements of standard decision and game theory; see Peterson (2015) for a compendium. Many have the

intuition that, e.g., Hume's farmers ought to be able to cooperate. After all, they both end up better off if they, collectively, do not play the strategies that rationality evidently demands of them. Moreover, in experimental studies people cooperate about half of the time in (one-shot) decision problems that resemble prisoner's dilemma situations (Camerer, 2003, p. 46).

Some have suggested that we might draw parallels between prisoners' dilemmas and phenomena of temptation and potential preference change discussed in §2.1.7, such that the two predicaments may enjoy a common solution. For instance, Gauthier (1994) argues that, just as in a case of temptation, maintaining a policy of cooperation might conduce better to one's life going well overall (cf. §6.4). Other authors like Gold and Sugden (2007) maintain that we can, and often do, elevate to a "group" or "team" perspective and figure out what course of action is best for the group. Whether such attempts to rationalize cooperation amount to changing the decision situation fundamentally (see Binmore 2015) remains a matter of controversy.

In any case, it is well understood that *sequential* versions of the prisoner's dilemma can induce cooperative behavior in rational players, consistently with classical decision and game theory (see Example 13 below). Before approaching the sequential game setting we first introduce some foundational concepts from computability theory.

3.2 Turing Machines and Implementable Strategies

The fundamental idea behind resource rationality is that agents implement *programs* that govern their behavior. A foundational approach to understanding which behaviors can, in principle, be implemented by some program takes Turing's (1936) celebrated analysis of computability as a starting point. The very idea that the mind can be likened to some kind of computing device takes inspiration from this analysis (McCulloch and Pitts, 1943; Putnam, 1967). The use of Turing machine programs *per se* in studying game theoretic strategies was explored thoroughly starting in the 1980s (e.g., Neyman 1985; Nachbar and Zame 1996, among many others).

We leave the definition of Turing Machines here semi-formal, since the details can be found in any textbook on theory of computation (e.g., Hopcroft and Ullman 1979). We assume:

Definition 8. A Turing machine is given by:

1. An infinite read/write tape with inputs ("observations") written on the tape at the beginning of computation and outputs ("actions") written on the tape at the time of halting;
2. A finite set of *states* that the machine can be in;
3. Transition rules that can flip fair coins to determine what to do next; specifically, these rules have the form:
$$\langle q, o, b; a, d, q' \rangle,$$
and are read: when in state q, reading symbol o and random bit b, rewrite o as a, go direction d on the tape (left or right), and enter state q'.

From Turing's (1936) analysis and much subsequent work, it is generally assumed that deterministic Turing machines—those whose actions do not depend on the random input—define exactly the *computable* functions. This class is widely assumed to delimit the functions ("input-output" relations) that could be computed by (reasonable idealizations) of the human mind. As von Neumann famously put it:

> You insist that there is something a [Turing] machine cannot do. If you will tell me precisely what it is that a machine cannot do, then I can always make a machine that will do just that. (quoted in Jaynes 2003, p. 7)

A similar analysis can be given for the more general class of *probabilistic* Turing machines (PTMs). By allowing output to depend on random bits, PTMs define functions from observations to *distributions* on actions, or in other words, *strategies* in the sense of Def. 2. A strategy that can be implemented by a Turing machine we call a *computable strategy*.

One-shot games, as defined above in §3.1, do not involve any observations. So a strategy will simply be a machine for effecting a distribution on actions (a "mixed" strategy). A basic question is whether equilibrium play can be achieved by a pair of PTMs. That is, for cases like Example 10 that require mixed strategies, when are these strategies computable by Turing machines? Turing (1936) defined the notion of a *computable real number*: roughly, one for which there is a Turing machine that can compute a rational approximation to the number for any desired degree of approximation. As long as the utilities in the game are all computable numbers, it can be shown that equilibrium strategies will be Turing computable. The reason is that equilibria can be construed as solutions to polynomial equations, always giving algebraic (and thus computable) numbers (Lipton and Markakis, 2004). Specifically, a pair of (mixed) strategies (σ_0, σ_1) is associated with probabilities $x_1, \ldots, x_n$ and $y_1, \ldots, y_m$ over A_0 and A_1, respectively; and an equilibrium will be a solution to the following system:

$$\begin{aligned} x_1, \ldots, x_n, y_1, \ldots, y_m &\geq 0 & \mathbf{U}_0^{(\sigma_0,\sigma_1)} &\geq \mathbf{U}_0^{(\sigma_0^*,\sigma_1)} \\ \sum_{i \leq n} x_i = \sum_{j \leq m} y_j &= 1 & \mathbf{U}_1^{(\sigma_0,\sigma_1)} &\geq \mathbf{U}_1^{(\sigma_0,\sigma_1^*)} \end{aligned}$$

where $\mathbf{U}_0, \mathbf{U}_1$ are the polynomials defined above in (3.1), and σ_0^* and σ_1^* range over the (finitely many) *pure* (non-mixed) strategies for players 0 and 1, respectively.[2] As any solutions will therefore give computable real numbers, there will be a probabilistic Turing machine that exhibits this behavior; see, e.g., Icard (2020) for a proof. We thus have:

Theorem 2. Equilibrium play in a one-shot game can be accomplished by Turing machines.

Recall, however, that the interest in agent programs is the fact that different programs may incur different costs. Suppose a pair of (Turing machine) programs $\pi = (\pi_0, \pi_1)$ generates a pair of strategies $\sigma = (\sigma_0, \sigma_1)$ with expected utilities $\mathbf{U}_0(\sigma), \mathbf{U}_1(\sigma)$, and that using π_i incurs cost $C(\pi_i)$. Then we arrive at a version of (1.1) for each player $i \in \{0, 1\}$:

$$\mathbf{V}_i(\pi) \quad = \quad \mathbf{U}_i(\sigma) - C(\pi_i).$$

In this new *machine game*, an equilibrium would be a pair $\pi = (\pi_0, \pi_1)$ such that neither has any incentive to adopt a different program given the other player's choice of program. In a machine game, equilibria are no longer guaranteed to exist in general:

Example 12. A simple observation due to Halpern and Pass (2015) is that in games with only mixed strategy equilibria—games like that in Example 10—charging for randomness results in a situation with no equilibria. For instance, if deterministic programs incur no cost while any probabilistic program incurs positive cost, then for any mixed strategy that player $1 - i$ might adopt, player i would prefer some pure strategy to any mixed strategy. But by assumption there are no pure strategy equilibria in the game.

There is a sense in which this situation is representative, in that equilibria are guaranteed to exist so long as there is no cost for randomization (Halpern and Pass, 2015, Thm. 5.4). Thus, costs for other resources such as time and space do not affect Theorem 2. It is perhaps interesting to consider in what conceivable settings access to random bits, but not other resources like time or memory, would come with opportunity cost.

[2] It is easy to check that restriction to pure strategies suffices here; cf. 4.4 below.

Remark 8. To the extent that a decision problem—whether multi-agent or single-agent—requires updating subjective probabilities with new observations, the question arises of whether *probabilistic conditioning* (recall Eq. (2.1)) is a computable operation. In general, it is not. Even if an agent's prior probabilities are expressed by a Turing machine, it does not follow in general that conditional probabilities will be; see Ackerman et al. (2019) for the continuous case, and Icard (2020) for the discrete case. As both of these papers show, there are nonetheless substantial classes of instances where computability is closed under conditioning.

Remark 9. It is worth mentioning here the topic of (asymptotic, worst-case) complexity theory. Taking the Turing machine as a model of computation, this approach concerns the amount of space (viz. number of tape cells) or time (viz. number of steps) required for computational problems, particularly as a function of the "size" of the problem instance. If the required space or time (for the worst-case problem instance) increases too quickly, it is said to be intractable. Finding a Nash equilibrium in a two-person is believed to be intractable in this sense, which some have argued calls its normative and predictive significance into question (Daskalakis et al., 2009). Computing conditional probabilities is also believed to be intractable, and, as mentioned in §2.4, computing a solution to a POMDP is known to be intractable.

This type of complexity analysis can be a useful tool for *a priori* falsification of cognitive models: if a mind is hypothesized to compute some function whose resource requirements grow too quickly even for moderately sized instances, then that hypothesize should either be rejected or else refined, e.g., by identifying natural "subproblems" that remain tractable (van Rooij et al., 2019). We will not be pursuing such an approach in this Element, and we will not be using the so-called complexity classes—Polynomial time, NP-time, Polynomial space, etc.—to classify problems. The study of resource rationality, as proposed in this Element, typically benefits from more fine-grained approaches to resource costs.

Questions of computability and resource limitations take on special interest in repeated, or sequential, games, to which we now turn.

3.3 Repeated Games

Given a game $\mathcal{G}$, we can imagine playing it multiple times. This essentially becomes a sequential decision problem, just as in §2.2. As in the more general setting, a history is a finite sequence of elements from $\mathbf{A}$—here including at each point an action for all of the players—and a strategy for player i is a function $\sigma_i : \mathcal{H} \to \Delta(A_i)$, where $\mathcal{H}$ is the set of all histories. Recall the progression of definitions in §2.2, starting with the T-step utility for history h:

$$U_i(h) \quad = \quad \sum_{t=0}^{T-1} \gamma^t u_i(h_t, a_t)$$

For a pair of strategies $\sigma = (\sigma_0, \sigma_1)$ we define a T-step distribution on histories, $\mathbf{P}^{\sigma}_{T}$, in a similar way, replacing the prior q_0 and the transition distribution $q_{h,a}$ on states with player $i-1$'s strategy, which analogously produces a distribution on actions (player $i-1$'s actions are like states for agent i) for each history. This allows defining the T-step expected utility for i:

$$(\mathbf{U}^{\sigma}_{T})_i \quad = \quad \mathbb{E}_{h \sim \mathbf{P}^{\sigma}_{T}} U_i(h)$$

and finally the overall utility for agent i, when the strategies $\sigma = (\sigma_0, \sigma_1)$ are being played:

$$\mathbf{U}_i(\sigma) \quad = \quad \lim_{T \to \infty} (\mathbf{U}^{\sigma}_{T})_i. \tag{3.2}$$

Holding fixed 1's strategy σ_1, we can ask which strategy σ_0 maximizes $\mathbf{U}_0(\sigma_0, \sigma_1)$. This strategy is called 0's *best response* to σ_1, and it brings us back to the ordinary setting of sequential decision problems when the "transition function" (i.e., agent 1's strategy) is known.

Note that, from each agent's perspective a sequential game is a typically non-Markovian decision problem (Def. 3), since what the other player will do next may itself depend on earlier interaction in the history of play. Thus, strategies cannot necessarily be assumed stationary (Def. 4). One could again take an individualistic ("Robinson Crusoe") Bayesian approach by putting a prior probability on the possible strategies that the other agent could possibly be pursuing. The most common game theoretic approach, however, instead extends the notion of Nash equilibrium (Eq. (3.1)) from the one-shot context.

Definition 9. In a sequential game, we say that $\sigma = (\sigma_0, \sigma_1)$ is a *Nash equilibrium* if, for each $i \in \{0, 1\}$, we have $\mathbf{U}_i(\sigma) \geq \mathbf{U}_i(\sigma')$ for any σ' that differs only on σ_i. In other words, each agent's strategy is a best response to the other agent's strategy.

3.3.1 The Folk Theorem

One of the main results on repeated games is the Folk Theorem (e.g., Aumann 1981; Fudenberg and Maskin 1986), which tells us what combinations of rewards are attainable from equilibrium play in infinitely repeated games. Somewhat dual to the maximin concept (see Eq. (2.3)), let us say that the *minimax actions* for players 0 and 1, respectively, are defined:

$$a_0^* = \arg\min_{a_0 \in A_0} \max_{a_1 \in A_1} u_1(a_0, a_1) \qquad\qquad a_1^* = \arg\min_{a_1 \in A_1} \max_{a_0 \in A_0} u_0(a_1, a_0).$$

In other words, a_0^* is the action by 0 that pushes 1's utility as low as possible, provided 1 is playing the best response to the action. Define the *minimax values*, respectively, as:

$$m_0 = \max_{a_0 \in A_0} u_0(a_0, a_1^*) \qquad\qquad m_1 = \max_{a_1 \in A_1} u_1(a_0^*, a_1).$$

Then, e.g., m_1 is the best utility 1 can achieve when 0 is playing their minimax action. The Folk Theorem then says that, in a repeated (discounted) game, if there is a pair of actions that guarantees each some utility greater than their minimax values, then that pair of actions will form an equilibrium. More formally:

Theorem 3 (Folk Theorem)**.** Suppose there is a pair of actions $a = (a_0, a_1)$ such that for each $i \in \{0, 1\}$, $u_i(a) > m_i$ for the minimax value m_i. Then there is an equilibrium pair of strategies σ for the repeated game, such that $\mathbf{U}_i(\sigma) = \sum_{k=1}^{\infty} \gamma^{k-1} u_i(a)$ for both $i \in \{0, 1\}$.

Proof sketch. If agent 0 plays the strategy that begins with a_0 but switches to the minimax action a_0^* forever as soon as the agent 1 deviates from playing a_1, then agent 1 cannot benefit by deviating, as long as the discount is suitably high. A symmetric argument applies to agent 1 playing such a strategy. Thus, they play $a = (a_0, a_1)$ for the duration of the game, each $i \in \{0, 1\}$ achieving cumulative reward $\sum_{k=1}^{\infty} \gamma^{k-1} u_i(a)$. □

A notable instance of the Folk Theorem is the repeated prisoner's dilemma:

Example 13 (Iterated Prisoner's Dilemma)**.** The minimax values for the prisoner's dilemma (Example 11) are $(1, 1)$. However, when both cooperate (both choose c) they achieve values $(3, 3)$. The Folk Theorem thus tells us that there are strategies allowing both to cooperate on every play of the game. In particular, both agents can play a strategy of cooperation up until the other fails to cooperate, at which point they defect forever. This pair of strategies (known as "trigger" strategies—see Example 15 below) forms a Nash equilibrium.

3.3.2 Turing Machines Playing Repeated Games

Repeated games—and indeed sequential decision problems more generally—present a distinctive challenge for resource limited agents. One's strategy needs to be sensitive to an increasing stream of observations. In the game theoretic context the observations are of other players' actions. Intuitively, strategies may be more or less costly to implement, depending on their sensitivity to the other player's actions.

Recall that the study of resource rationality can be divided into cost-theoretic and panoramic approaches (§1.3). Taking a panoramic approach, we can start by asking which strategies can be implemented by a Turing machine at all. We know from general computability theory that many functions from an infinite set (like the set of all histories) to any other non-empty set (like the set of actions) are not computable by any Turing machine. The question is, how much of the theory of repeated games survives a restriction to computable strategies?

In one study of iterated prisoners dilemmas (and many games like it), Nachbar and Zame (1996) show that there is a pair of equilibrium (deterministic) strategies (σ_0, σ_1) such that σ_1 can be implemented by a Turing machine, while no best response to σ_1 is computable. The strategy σ_1 amounts to a complicated pattern of delayed "punishments" for defecting behavior, such that responding optimally would be tantamount to solving Turing's (1936) famous halting problem. Furthermore, there are classes of computable strategies for player 1 that each admits some computable best response for player 0, but the function that computes a best response is itself uncomputable. In other words, for each strategy in the class there is a Turing machine that would be a best response to it (and thus they would be in equilibrium), but there is no Turing machine that takes a game as input and outputs the requisite strategy for that game.

Despite these limitative results, some of the standard results still hold, including the Folk Theorem 3. This is because the simple pair of strategies described in the proof of Theorem 3 can easily be carried out by Turing machines. In fact, those strategies can be carried out by a special class of Turing machines known an *finite state automata*, which can be construed as Turing machines with a fixed finite bound on available memory (see, e.g., Minsky 1967). Much of the literature in economics has been devoted specifically to automata playing games, and this will take us back to the cost-theoretic approach to resource rationality.

3.4 Finite Automata

Following a suggestion by Aumann (1987), a sizable body of work has investigated the *memory costs* involved in carrying out sequential strategies (recall Remark 3). Reviews of this research tradition can be found in Kalai (1990); Rubinstein (1998). Most of the literature focuses on deterministic automata, but we turn to probabilistic automata below in §3.4.2.

3.4.1 Deterministic Automata

A very general type of deterministic automaton is sometimes called a *transducer* since it transforms sequences of inputs to sequences of outputs:

Definition 10. A (deterministic) *automaton* is given by a quadruple $\mathcal{A} = (Q, q_0, O, A, \delta, \tau)$:

- Q is a non-empty set of states, with distinguished *start state* q_0;
- O is the input alphabet, or *space of possible observations*;
- A is the output alphabet, or *space of possible actions*;

- $\delta : Q \to A$ is the *output function*;
- $\tau : Q \times O \to Q$ is the *transition function*;

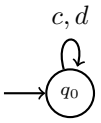
If Q is finite, we call $\mathcal{A}$ a *finite-state automaton*.

In the context of a two-player game $\mathcal{G} = ((A_0, A_1), u)$, it is natural to take $A = A_i$ to be player i's actions, and $O = A_{1-i}$ to be the actions of the other player. Then we have:

Fact 1. Every deterministic strategy σ_i can be represented by an automaton.

To see this, simply let $Q = \mathcal{H}$ be the set of states, with ϵ (the empty history) the initial state q_0. We can define $\delta(h) = \sigma_i(h)$ and let $\tau(h, a)$ be simply h concatenated with a. Of course, this is most interesting when the resulting automaton is finite. Not every strategy can be represented by a finite-state automaton, but many can. Here are some elementary examples:

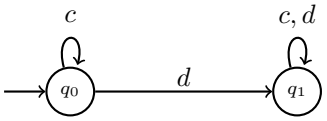
Example 14 ("Always Cooperate")**.** A simple strategy to cooperate unconditionally gives a one-state automaton, beginning in state q_0 and remaining there no matter whether the other player chooses c or d (notated on the transition arrow from q_0 back to itself):



$\delta(q_0) = c$

By declaring $\delta(q_0) = d$ instead, we obtain an "Always Defecting" agent.

Example 15 ("Trigger")**.** Another simple strategy is to cooperate until the other agent defects, and to defect from then on:



$\delta(q_0) = c$

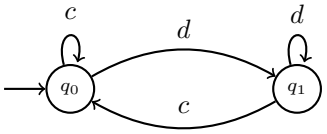
$\delta(q_1) = d$

This is the strategy we used to illustrate one special case of the Folk Theorem. In a bit more detail, suppose our discount factor γ is $\frac{3}{4}$. If both players opt for Trigger, then, according to Eq. (3.2), they both have an expected payoff of $\sum_{k=1}^{\infty}(\frac{3}{4})^{k-1} \times 3 = 4 \times 3 = 12$. Given this strategy of the other player, no deviation can result in a better expected payoff. For example, if at any stage n of the iterated game, agent 1 plays d, they will gain 1 utile by doing so (achieving payoff 4 rather than 3). But since player 0 will henceforth play d, the best they can do from then on is to play d as well, which will result in a (discounted) payoff of 1 at all subsequent rounds. The point is that for any $n > 0$:

$$\sum_{k=n}^{\infty} \left(\frac{3}{4}\right)^{k-1} 3 \quad > \quad \left(\frac{3}{4}\right)^{n-1} 4 + \sum_{k=n+1}^{\infty} \left(\frac{3}{4}\right)^{k-1}.$$

A similar calculation shows that two players adopting the following strategy will also be in equilibrium, so this provides yet another illustration of the Folk Theorem.

Example 16 ("Tit-for-Tat")**.** Another strategy is to play whatever the other just played:



$\delta(q_0) = c$

$\delta(q_1) = d$

In fact, if 0 opts for Trigger and 1 opts for Tit-for-Tat, they will again be in equilibrium. It is perhaps noteworthy that in empirical studies of iterated prisoners dilemma, all of the strategies mentioned above are attested (dal Bó and Fréchette, 2018). Furthermore, there appears to be a preference for simpler, or less costly, strategies across human experiments.

A natural cost to associate with an automaton $\mathcal{A}$ is the *number of states* in Q. We can think of these as "states of mind" whereby an agent can only be in one of finitely many. Correspondingly, let us say that the inherent (memory) complexity of a strategy is the size of the smallest automaton implementing the strategy:

Definition 11 (Complexity of a Strategy)**.** The complexity of a strategy σ, written $C(\sigma)$, is the smallest number k, such that there is an automaton with k states that implements σ.

An alternative way of understanding this same complexity measure in terms of the size of an "information system" needed to encode the strategy (Kalai, 1990).

Definition 12 (Complexity of a Strategy, Version 2)**.** Given a strategy σ and a history h, write $(\sigma \restriction h)$ to be the function defined by $(\sigma \restriction h)(h') = \sigma(h \cdot h')$. Then complexity can be taken as the cardinality of the set $\{(\sigma_i \restriction h) : h \in \mathcal{H}\}$.

The next result follows from the Myhill-Nerode Theorem (e.g., Hopcroft and Ullman 1979):

Proposition 3. The complexity measures in Defs. 11 and 12 agree: the cardinality of the information system coincides with the number of states needed to implement the strategy.

As another special case of (1.1)—and specifically Eq. (1.2)—we can define the *cost-adjusted* utility of strategy σ_i to be (where, again, $\mathbf{U}_i(\sigma)$ is defined as in Eq. (3.2)):

$$\mathbf{V}_i(\sigma) \quad = \quad \mathbf{U}_i(\sigma) - \frac{1}{\beta} C(\sigma_i). \tag{3.3}$$

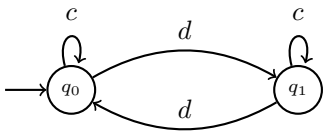
for some $\beta > 0$, measuring how much additional states cost. All of the automata above are minimal for their respective strategies, possessing costs of β and $2/\beta$.

Suppose we take, for instance, $\beta = 1$. Will the pair (Trigger, Trigger) still be in equilibrium? According to Eq. (3.3) both agents can expect a utility of $12 - 2 = 10$. If one agent is playing Trigger—or indeed any strategy that always cooperates whenever the other agent is cooperating—then the other can benefit by switching to Always Cooperate: this has the same long run payoff, but the cost is only 1 rather than 2. However, when one agent is playing Always Cooperate the other agent would clearly benefit by playing Always Defect. These observations reveal that there is *no* pair of machine strategies in equilibrium according to which both agents always cooperate. So in fact:

Proposition 4. For sequential decision problems with strategic cost measured as in Def. 11—that is, by the size of the smallest implementing automaton—the Folk Theorem 3 fails.

Nevertheless, equilibria in these games are always guaranteed to exist. Here is one interesting example of equilibrium strategies presented in Rubinstein (1998):

Example 17 ("Initial Threat")**.** Suppose both agents play the following strategy:



$\delta(q_0) = d$

$\delta(q_1) = c$

In other words, both agents start off defecting, which moves both to state q_1, where they will cooperate as long as the other is cooperating. With $\delta = \frac{3}{4}$ and $\beta = 1$ both expect utility of $1 + \sum_{k=1}^{\infty}(\frac{3}{4})^k 3 - 2 = 8$. Neither pure strategy is a best response to Initial Threat: Always Cooperate leaves the other defecting forever and gives a utility of -2, while Always Defect has the other defecting every other step, which again results in long run utility of less than 8. Moreover, once the second agent is in state q_1 a similar analysis to that above shows that the first gains no advantage by deviating, as the subsequent punishment outweighs the momentary gain. Both agents playing Initial Threat forms an equilibrium.

Somewhat surprisingly, in (pure strategy) equilibria of these games, the strategies opted by the two players always have the same number of states, and even much of the same structure (Rubinstein, 1998). There are numerous other results about machine games. For instance, Neyman (1985) shows that if the players have an upper bound on the size of their automaton, then cooperation can result even in the prisoners dilemma repeated only finitely many times (see also Halpern and Pass 2015). Meanwhile, Gilboa and Samet (1989) studied repeated games in which one player can choose an arbitrarily complex Turing machine while the other is restricted to a finite state automaton, showing that under certain assumptions the limited player has a strategic advantage (the "tyranny of the weak" effect). For further discussion and many more results, see the surveys in Kalai (1990); Rubinstein (1998).

3.4.2 Probabilistic Automata

The literature on games played by automata has tended to focus on deterministic automata. But as in our discussion of Turing machines above in §3.2, it makes sense to consider stochastic machines in addition. These will play an important role in the next chapter as well.

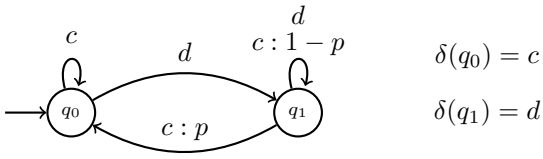
Definition 13. A *probabilistic automaton* is just like in Def. 10, except that the transition function now brings a state/observation pair to a distribution on next states:

$$\tau : Q \times O \to \Delta(Q).$$

We allow generalizing from a start state q_0 to an initial distribution $p_0 \in \Delta(Q)$ over states.

A probabilistic automaton thereby implements a strategy in the sense of Chapter 2.

Example 18 ("Stochastic Tit-for-Tat")**.** The following specifies a mixed strategy for the iterated prisoners dilemma, whereby the agent (deterministically while in q_0) begins cooperating, defecting (moving to q_1) whenever the other defects, and then moving back to q_0 only with some probability p whenever the other cooperates.



The usual Tit-for-Tat strategy (Ex. 16) is the special case where $p = 1$, while Trigger (Ex. 15) is the special case where $p = 0$.

This automaton is a kind of interpolation between Tit-for-Tat and Trigger, in the sense that one could, at each stage, flip a p-biased coin to determine whether to play Tit-for-Tat or Trigger for the next move. By a similar calculation to those above in §3.4.1, any two players

adopting any version of Stochastic Tit-for-Tat (that is, any choices for the parameter p) will be in equilibrium. However, again for similar reasons, they will no longer be in equilibrium once we impose a cost on each additional state: in this case both would want to opt for the simpler Always Cooperate strategy, which will again thwart equilibrium.

Importantly, once we limit the size of automata, a stochastic strategy—that is, play by a probabilistic automaton—is not the same thing as a probabilistic mixture of automata. Some of the literature (e.g., Neyman 1985) has considered settings where a player can adopt a mixed strategy by randomly choosing a (deterministic) automaton from among those below a given complexity, and then using that selected automaton for the duration of the repeated game. It is straightforward to check that there is no probabilistic mixture of two-state automata that will enjoy the same behavior as Stochastic Tit-for-Tat (Ex. 18). See §4.4.2 for more on this.

3.5 Automata and Polarization

One of the most striking multi-agent phenomena is belief *polarization*, whereby two individuals with only slightly differing views on a subject come to have dramatically different opinions after being presented with the same evidence. This phenomenon has been identified in numerous domains, including gun control and capital punishment (Taber and Lodge, 2006), climate change (Cook and Lewandowsky, 2016), and many other consequential topics.

Under very minimal conditions, two "ideal Bayesian" agents updating on the same evidence will come to agree, even given very different prior judgments (Blackwell and Dubins, 1962), a result that even holds when priors are assumed computable (Zaffora Blando, 2022). Polarization is evidently due to some departure from this idealized picture. Many different explanations have been given, but one theme that emerges in many studies is the idea that people are not perfect in the way they process and encode evidence (see Taber and Lodge 2006). For instance, while it seems sensible to update one's opinion about the reliability of a source based on how plausible one finds their testimony, polarization can result if such dependencies are routinely forgotten (Pilgrim et al., 2022). Similarly, polarization may persist in cases where people cannot remember all the evidence supporting their belief, even though they originally incorporated that (potentially ambiguous) evidence in a rational manner (Dorst, 2023).

Remarkably, the very general assumption that people have limited memory is enough to produce polarization, even among agents who are in every other way idealized. Thus, whereas the full story about how and why people polarize is likely to be multifaceted, the mere assumption of limited memory already affords a window in the phenomenon.

Suppose a decision maker faces a choice between two options a and b, and there are two relevant states, h and l, with probabilities $p(h)$ and $p(l)$, and the following payoff matrix:

	h	l
a	"high"	"low"
b	0	0

In other words, b is the "safe" action guaranteeing status quo, while a is the best action to take just in case state h obtains. For instance, a might be the action of *asserting that state h obtains*, while b could be publicly withholding judgment about whether h or l obtains, a choice that especially makes sense when the agent thinks l is the more likely state.

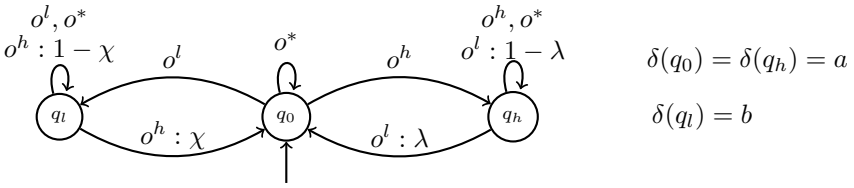
Suppose moreover that there is a finite set O of possible observations, with likelihoods $p(o|h)$ and $p(o|l)$ for $o \in O$. Then given some finite sequence $\vec{o} = (o_1, \ldots, o_k)$ of independent observations, the optimal action is straightforward: the decision maker should choose a just if

$$\frac{p(h|\vec{o})}{p(l|\vec{o})} > \frac{\text{"low"}}{\text{"high"}}$$

where $p(h|\vec{o}) = p(h) \times \prod_{i \leq k} p(o_i|h)$ is determined by Bayes rule, and likewise for $p(l|\vec{o})$.

But now suppose that our decision maker has only limited memory and thus cannot process data in an arbitrarily fine-grained manner. A very general way of imposing a constraint of finite memory on the agent is to assume that their decision making strategy must be implementable by a (possibly probabilistic) finite state automaton (Def. 13), with a hard upper bound on the number of states. In other words, the agent can employ any decision making strategy whatsoever, as long as it can be done with a limited number of functionally different states. This is another instance of the panoramic approach to resource rationality (§1.3).

In such a setting Wilson (2014)[3] has provided a general analysis of what the *optimal* finite state machine for this problem would be. Wilson assumes there is some termination probability η, such that after each successive observation, with probability η the sequence ends and an action is chosen. For fixed η and p, and values for "high" and "low," the aim is to determine for each k the automaton of size $\leq k$ that has the highest expected reward. It is shown not only that an optimum exists, but that it always takes on a particular form. As an illustrative example, for small η, the optimal 3-state automata all take the following form:[4]



Here o^h is the observation with highest likelihood ratio in favor of h—i.e., the o maximizing $\frac{p(o|h)}{p(o|l)}$—and likewise for o^l, while o^* stands for every other observation in O. In other words, the optimal decision making agent *ignores all but the most extreme signals*. In general, having reached one of the extreme states (q_l or q_h), the probabilities χ and λ of leaving those states upon presentation of contrary evidence will be quite small (and moreover, $\chi < \lambda$).

This analysis affords a number of different perspectives on rational polarization. The first point is that, even for two agents who have the same priors and likelihoods and the same (memory-optimal) decision making policy, if they make their observations in a different order, they are likely to occupy diametrically opposed positions. Even in this simple example, seeing o^l and then o^h is very different from seeing o^h before o^l; the first is almost sure to bring the agent to q_l, the second to q_h. Once there, arriving back at the "more neutral" q_0 is unlikely. This is of course in stark contrast to "ideal" Bayesian inference (recall §2.1.1; cf. Example 28).

For $k > 3$ the optimal automata look similar to that above, but with more intermediate states between the two extremes:

Some of the intermediate transitions will be deterministic (e.g., upon seeing o^h go right with probability 1), while others will be stochastic. Again, all non-extreme signals are ignored altogether. Thus, we see at least three more potential sources of polarization, aside from the order dependence already mentioned: (1) given the same priors and the same memory bound,

[3] See also Halpern et al. (2014) for further development of these ideas. The foregoing results were anticipated by a much earlier line of work from Hellman and Cover (1970, 1971), which will return below in §4.4.

[4] Assume moreover that $p(h) = p(l) = 0.5$, while "high" = 1 and "low" = -1. The probability parameters λ and χ are then a function only of η and the likelihoods $p(o|h)$ and $p(o|l)$.

polarization could result from random transitions; (2) two agents starting in two different intermediate states may respond differently to the same signals; (3) slightly different priors or likelihoods (or different upper bounds on size) would produce slightly different optimal automata, which again could result in two agents moving to opposite extremes with the same data points, in contrast to the classical merging-of-opinions results.

The full explanation for polarization phenomena in human agents is surely complex and multifaceted—see Bramson et al. (2017) for a systematic review and discussion. It is remarkable, and potentially illuminating, that optimization under limited memory is already sufficient to recapitulate the basic pattern. Wilson's results on optimal automata—as well as older results from Hellman and Cover (1970, 1971)—lead us directly to the theme of the next chapter, concerning the question of when randomization may be beneficial for a resource-limited agent.

3.6 Conclusion

The goal of this chapter has been to introduce some relatively simple models of "agent programs" modeled on Turing machines and finite state automata. Relative to standard game theory, this resource rational perspective already highlights some important departures, including two of the most canonical results in game theory, Nash's Theorem and the Folk Theorem.

With the next few chapters we will be slowly moving away from the still highly idealized picture of agent programs studied in this literature, toward analyzing increasingly detailed features of human agents, which in turn calls for greater psychological fidelity in what we take to be an agent program.

Chapter 4

Resource Rational Randomness

A remarkable aspect of resource rationality is the emergence of *randomness* in characterizations of rational thought and behavior. One might have expected randomness and randomization to be somehow in opposition to rational, purposive behavior. After all, the very etymology of 'random' brings in ideas of being "without aim or purpose, or principle" (Oxford English Dictionary). In fact, there is a very general reason that non-trivial randomization is prohibited by standard Bayesian decision theory, reviewed below in §4.2. Yet, as we have already seen in the previous chapter (e.g., §3.5), random behavior may nonetheless emerge as optimal once we take resource limitations into account. Evidently, introducing computational cost into the picture can be sufficient to defeat the decision theoretic argument. The task of the present chapter is to interrogate this phenomenon in more depth.

Aside from being of interest in its own right, randomization provides a useful case study for illustrating two alternative approaches to thinking about costs and resources. The first considers random behavior to be a kind of *default*, whereby computational resources are required in order for the agent to override that default. Even if the ultimate goal is to choose the "right" action deterministically, resource limitations make it so that one ought only go so far in that direction. The second approach construes randomness itself as a *resource*, which might trade off with other resources such as memory. That is, if an agent could somehow harness a suitably random source, this might be helpful for mitigating other resource limitations.

These two ways of thinking about rational randomness—*randomness as default* and *as resource*—draw upon different technical tools. The second is based on the machine models introduced in the previous chapter (Turing machines, automata, etc.). The first invokes ideas from *information theory*. Thus, a further task of the present chapter is to introduce these ideas from information theory. Additionally, randomization provides a natural segue to the next two chapters, which concern the use of resource rationality to understand key aspects of cognition.

4.1 Randomness in Humans

There is considerable controversy over what processes count as *random*, and indeed whether any processes in the world are "genuinely" random. For present purposes, it is enough to consider processes that are *well modeled* as being probabilistic or random. This may of course be because the process is in some sense indeterminate, but it could also be because—from an appropriate perspective—the process is *unpredictable* (Eagle, 2005). In this case, the probabilities summarize variations in behavior that cannot be (practically) predicted. Once we admit these more perspectival varieties of randomness, random processes evidently abound in nature, and

in the human sciences in particular; see, e.g., Glimcher (2005) and Icard (2021) for reviews.

4.1.1 Neural Noise

Some of the most striking evidence for random processes in humans comes from neuroscience. As Niven and Laughlin (2008) summarize in their review of energy limitations in perception, "intrinsic noise occurs at all stages of sensory processing, including the transduction of the sensory stimulus into an electrical signal, the transmission of electrical signals within neurons and synaptic transmission of signals between neurons" (p. 1796). Since information only ever decreases along a noisy channel—a fact known as the data-processing inequality (Shannon, 1948)—each additional source of corruption only seems to make the problem of extracting meaningful sensory signals from the environment more difficult.

Below sensory processing, even at the cellular level there appears to be inherent noise in neural firing rates. It is well known, for example, that neurons in the visual cortex respond preferentially to different properties of visual stimuli. A neuron's average firing rate over a given time span might be a reliable indicator of, say, the orientation of a line in some part of the visual field (Example 29). When researchers have attempted to characterize the firing pattern in greater detail, it has been repeatedly found to be well modeled as otherwise random. The "most random" probability distribution for a given average rate is the *Poisson distribution*:[1]

$$p_\lambda(k) \quad = \quad \frac{\lambda^k e^{-\lambda}}{k!} \tag{4.1}$$

For a given firing rate λ within a fixed time interval, $p_\lambda(k)$ is the probability of observing exactly k spikes in that interval. While neural firing is not exactly Poisson distributed—e.g., due to refractory periods in which firing is momentarily suppressed—numerous studies have employed Eq. (4.1) as an accurate approximation (see Faisal et al. 2008 for a review and discussion). Furthermore, there is reason to think randomness, or at least unpredictability, is relatively fundamental in the nervous system (Mainen and Sejnowski, 1995; Glimcher, 2005).

This short summary reveals two potential links to resource limitations, with randomness appearing either as a feature or as a bug. The ubiquity of noise in neural processing could be seen as a feature, were it determined that some extractable source of randomness could be beneficial. At the same time, it is a potential bug if we focus instead on the resources necessary to preserve and transmit as much of the incoming stimulus as possible.

4.1.2 Softmax Decisions

Apparent randomness has been observed not just in the brain but also in observable high-level behavior. From the earliest days of empirical psychology, researchers like Fechner and Thurstone assumed that behavior could only be summarized probabilistically, and subsequent research explicitly modeled choice behavior as stochastic (see Luce and Suppes 1965 for an early review). One particularly striking phenomenon is known as *probability matching* (e.g., Estes 1959). Suppose that exactly one of two levers, A and B, could deliver a reward. Imagine that 70% of the time lever A delivers a reward, while 30% of the time it is B that delivers the reward. If presented with these two levers it seems that the optimal strategy would always be to pull lever A. Empirically, however, people often *match* the success probabilities, pulling A around 70% of the time and B around 30% of the time (see Vulcan 2000 for a review).

[1] Specifically, it is the discrete distribution with maximum entropy among a reasonable family of distributions that give the same average rate (Harremoës, 2001). For more on entropy see 4.3.1 below.

Such matching phenomena also extend to more "notional" probability distributions such as those derived by Bayes' rule from assumed priors and likelihoods (Peterson and Beach, 1967). For instance, when asked to estimate how much a movie will ultimately gross after observing that it has made $40 million so far, experimental participants give responses that in aggregate match the posterior derived from the actual empirical distribution (Griffiths and Tenenbaum, 2006), rather than giving the response that would maximize posterior probability. (Note the similarity to *posterior sampling* discussed in Chapter 2, Example 9, in the context of learning.)

This and a plethora of other behavioral patterns have been well modeled in terms of a decision strategy first formulated by Luce (1963). Given a set A of actions, suppose we have a measure $V(a)$ of how "valuable" action $a \in A$ is in the present context. The strategy suggested by Luce is to select a with probability given as follows:

$$\frac{e^{\beta V(a)}}{\sum_{a' \in A} e^{\beta V(a')}}, \tag{4.2}$$

for some parameter β. This is in fact the same softmax function that we considered in Example 7 during our discussion of reinforcement learning (§2.3). In that setting $V(a)$ was an estimate of the expected utility of a during the course of learning. Aside from taking $V(a)$ as expected utility, Eq. (4.2) will also encompass examples like the following:

Example 19. As a special case, consider a setting with possible states S and prior $p(s)$, together with possible observations O and likelihood $l(o|s)$, so that the posterior $p(s|o)$ is given in terms of Bayes' rule, $p(s|o) \propto p(s)l(o|s)$. Assume that $S = A$ and the utility of guessing state s given observation o is $V(s) = \log p(s|o)$. In such a setting the task is to guess the right state, and we assess how valuable a guess is by (the natural logarithm of) its posterior probability. Then, assuming also that $\beta = 1$, the softmax expression in (4.2) becomes $p(s|o)$. In other words, we recover posterior matching exactly in this setting.

A conservative way to think of the softmax function in (4.2) is in terms of a theorist's uncertainty about the agent's own latent utility function. Under suitable assumptions about the distribution of utility functions in a population—see Theorem 4 below—the softmax rule emerges as the correct way of analyzing choice probabilities in the population (McFadden, 1973). This so-called random utility model is, by design, fully consistent with classical decision theory. Each individual agent is assumed to maximize expected utility deterministically. Apparent randomness arises from a noisy sample of a heterogenous population.

An earlier, more radical interpretation, following Thurstone (1927) and Luce and Suppes (1965), attributes *stochastic behavior* to the individual agent. We might assume the agent *samples* a utility function u from a distribution on such functions, and then maximizes (expected) utility according to u. Suppose there is a random variable U_a for each action a, giving the distribution on utility values for a, and that these are all independent for different actions a. Suppose, moreover, that the probability density function for possible values x of U_a is:

$$f(\mathsf{x}) \quad = \quad \begin{cases} e^{\beta V(a) + e^{\mathsf{x}\beta V(a)}} & \mathsf{x} \leq 0 \\ 0 & \mathsf{x} > 0 \end{cases}$$

giving a so-called Gumbel distribution on each U_a. Then we can show the following:

Theorem 4. The probability that U_a is maximal is given by Eq. (4.2), the softmax rule.

A proof of Theorem 4 is provided in Appendix 4.A.1. Thus, a specific assumption about the distribution on utility functions provides one way of deriving the softmax rule. This interpretation again does not concern resource limitations per se.

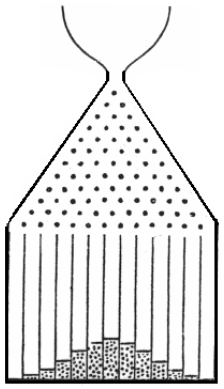


Figure 4.1: Illustration of physical artifact capable of drawing "samples" from a normal (Gaussian) distribution. Adapted from Galton (1889).

4.1.3 Sampling Propensities

While these random utility interpretations (stochastic or deterministic) may be appropriate for many settings, they do not obviously apply to examples like posterior matching (Example 19), where we assume the utility structure is relatively clear (typically: guess the state with highest probability). A third interpretation, suggested first in the literature by Vul (2010), is that agents stochastically sample, not a utility function, but a state s in proportion to its (posterior) probability. The idea is that people have an inbuilt ability to generate possibilities in a random way, and to use these "samples" for decision making.

In particular, if an agent drew some number k of samples $s_1, \ldots, s_k$ from $p(s)$, then they might choose an action a that maximizes the *estimated* expected utility in the following sense:

$$\text{Choose } a \text{ that maximizes} \quad \widetilde{V}(a) \quad = \quad \frac{1}{k}\sum_{i=1}^{k} u(s_i, a). \tag{4.3}$$

When $k = 1$, $S = A$, and $u(s, s') = \mathbb{1}_{s=s'}$, we recover precisely posterior matching. But this formulation also captures a wider array of choice problems in which state estimation is an important component. For example, in the course of deciding which of two highways to take a person might (unconsciously) conjure up several possible states of the world—imagining what the traffic situation is likely to be, how pleasant each one might be, etc.—and compute an estimate $\widetilde{V}(a)$ of how good each option a appears across these sampled states.

Just as (4.2) converges to maximizing behavior as β increases, so (4.3) converges to expected utility maximization as k increases. Smaller k will in general produce more random decisions. There is evidence that experimental participants do in fact produce responses probabilistically, in a way that seems consistent with a method like (4.3) with relatively small k, i.e., few samples; see, e.g., Vul and Pashler (2008); Collins and Frank (2012); Vul et al. (2014), among others.

As these authors have reported, behavior tends to become less random given either higher stakes or more time to reach a decision (e.g., Vul et al. 2014). Whether we interpret this as modulating β in (4.2) or k in (4.3), it seems clear that stochastic choice may be closely related to issues of resource limitations. It has even been suggested, motivated by computational considerations, that a primary format for representing uncertainty may be via generative procedures that primitively draw samples from implicit probability distributions (Vul et al., 2014; Icard, 2016; Sanborn and Chater, 2016), much the same way a "Galton box" (Fig. 4.1) has

a propensity to sample (approximately) from a normal distribution without, e.g., explicitly encoding a mean or variance.

An important inspiration for this suggestion about sampling propensity comes from the observation that computational work in statistics and machine learning often employs "Monte Carlo" sampling algorithms to perform complex probabilistic inference tasks (e.g., Murphy 2012; cf. the discussion below in §5.3.4). In fact, the original revolution in Monte Carlo methods in science and engineering came from the need to simplify difficult calculations (Metropolis et al., 1953). The intuition is that, while exact calculations can be computationally prohibitive, it can be relatively easy to generate random samples from a related distribution and use these samples to approximate the quantity of interest. The question thus arises of how and why, exactly, would randomness confer resource-related benefits?

4.2 Counteracting Convexity

It is worth pausing to see exactly why randomness is prohibited by the Bayesian decision theoretic framework reviewed in Chapter 2. It is an instance of a more general pattern.

Given some deterministic programs π_i, let us write $\sum_i p_i \pi_i$ for the randomized program that pursues π_i with probability p_i. A value function $\mathbf{V}$ on programs is said to be *convex* if

$$\mathbf{V}\big(\sum_i p_i \pi_i\big) \quad \leq \quad \sum_i p_i \mathbf{V}(\pi_i). \tag{4.4}$$

This means that the value of a randomized program can be no greater than that of the *most* valuable program in the mixture. Unless all of the programs are equally valuable, an agent always stands to gain by deterministically opting for the most valuable one.

An example of a convex value function $\mathbf{V}$ is expected utility—as in Eqs. (2.2) and (2.9)—in that case, the inequality in (4.4) becomes an equality. In a one-shot decision problem this is because $\mathbb{E}_s \mathbb{E}_a u(s,a) = \mathbb{E}_a \mathbb{E}_s u(s,a)$, simply by the commutativity of addition. The argument is less immediate in the sequential decision setting, and it is closely related to a famous result in game theory known as Kuhn's Theorem (after Kuhn 1950).

Recall from §2.4 the notion of a strategy in a (partially observable) sequential decision problem, viz. a function $\sigma : \mathcal{O} \to \Delta(A)$ from observation histories to distributions on actions. In the present context—following the game theory literature—let us call this a *behavioral strategy*.

An alternative setting would be to consider the set Σ of deterministic strategies, that is, functions $\varsigma : \mathcal{O} \to A$, and the set $\Delta(\Sigma)$ of distributions over these deterministic strategies. Let us call the elements $\mu \in \Delta(\Sigma)$—again following game theory nomenclature—*mixed strategies*.

Suppose that O and A are both finite, i.e., there are only finitely many possible observations and only finite many actions available to an agent.[2] Then:

Theorem 5. For every behavioral strategy producing a distribution on histories (in the sense of Eq. (2.17)), there is a mixed strategy that produces the same distribution.

Proof sketch. Fix T, the number of steps, and consider any behavioral strategy σ. We define a mixed strategy μ as follows. The probability of adopting deterministic strategy ς is

$$\mu(\varsigma) \quad = \quad \prod_{\omega \in \mathcal{O}_t : t < T} \sigma_\omega(\varsigma(\omega)).$$

That is, the probability of a deterministic strategy ς is the product of the probabilities that σ would make the same choice as ς at each observation history. We show in Appendix §4.A.2 that this defines the same distribution on T-step histories as σ. □

[2]The finiteness restriction is necessary; see Megiddo (1994) for a counterexample to Thm. 5 if A is infinite.

Assuming the strategy σ is a computable function, it follows immediately by the argument above that μ will also be computable. Hence we have the following corollary.

Corollary 1. Any finitary behavior achievable by a probabilistic Turing machine can be achieved by a probabilistic mixture over deterministic Turing machines.

Theorem 5 and Corollary 1 essentially allow us to reduce the sequential setting to the one-shot setting: any (behavioral) strategy can be replaced by a deterministic strategy, with all randomization occurring upfront. Although game theory still countenances mixed strategies when the state is chosen by another intelligent agent (§3.1), once we fix the transition function and the likelihood (q and l from Def. 5), a similar argument applies. Recall Eq. (2.9):

$$\begin{aligned}
\mathbf{U}(\sum_i p_i \pi_i) &= \lim_{T\to\infty} \mathbb{E}_{h\sim \mathbf{P}_T^{\sum_i p_i \pi_i}} U(h) \\
&= \lim_{T\to\infty} \sum_i p_i \, \mathbb{E}_{h\sim \mathbf{P}_T^{\pi_i}} U(h) \\
&= \sum_i p_i \lim_{T\to\infty} \mathbb{E}_{h\sim \mathbf{P}_T^{\pi_i}} U(h) \\
&= \sum_i p_i \, \mathbf{U}(\pi_i).
\end{aligned}$$

The first step is allowed because all randomization occurs upfront, before the first observation. Thus, once again, randomization is never better, and typically leads to a worse result.

How could resources come in to defeat this argument? In broad strokes, convexity could fail simply due to the introduction of a cost term. Recall the fundamental Eq. (1.1):

$$\mathbf{V}(\pi) = \mathbf{U}(\pi) - C(\pi)$$

Suppose π^* is equivalent (in the sense of Theorem 5) to a mixture $\sum_i p_i \pi_i$, and hence that $\mathbf{U}(\sum_i p_i \pi_i) = \mathbf{U}(\pi^*)$. The mere possibility that the cost $C(\pi^*)$ could be less than the average cost $\sum_i p_i C(\pi_i)$ of the deterministic programs in the mixture (or any other behaviorally equivalent mixture) is enough to invalidate Eq. (4.4). In such a case, assuming that the cost of the mixture is just the average costs of the programs in it, we would have

$$\begin{aligned}
\mathbf{V}(\sum_i p_i \pi_i) &= \mathbf{U}(\sum_i p_i \pi_i) - \sum_i p_i C(\pi_i) \\
&= \mathbf{U}(\pi^*) - \sum_i p_i C(\pi_i) \\
&< \mathbf{U}(\pi^*) - C(\pi^*) \\
&= \mathbf{V}(\pi^*).
\end{aligned}$$

Of course, the mathematical possibility of this does not yet show that it ever happens in plausible or relevant circumstances. The remainder of the chapter will consider different assumptions about cost that lead to such a reversal, and even to the optimality of strategies like softmax.

4.3 Randomness as Default

One common perspective is that randomness is not somehow "added" in the decision making process. Exactly the opposite, randomness is a kind of default behavior ("without aim or

purpose") and deliberation is understood as sharpening one's action distribution in the direction of, e.g., deterministically maximizing expected utility.

As already noted above in §4.1.1, much of neural processing seems to be subject to random noise at various levels. While biological brains do appear to have mechanisms for mitigating the noise, randomness may well still percolate up to the level of observable behavior.

Let us suppose that our agent begins in what Skyrms (1990) calls a *state of indecision*, corresponding to a (typically randomized) *default strategy* δ (cf. also Huttegger 2017 on "choice propensities"). Perhaps the ideal strategy is one that opts deterministically for the action a^* of highest expected utility. For various reasons it may be costly to reach that alternative strategy starting from δ. For instance, δ may be a strongly ingrained habit, difficult to overcome. Or perhaps it is computationally demanding to figure out what a^* is in the first place. Whatever the reason, resource costs may in effect anchor the agent closer to the default from which they began. Our question in this section will be how to assess—in a suitably general way—the costs of moving from the default δ to some other strategy π.

The focus here will be on one specific proposal that has been offered independently by a number of different researchers. That is to use concepts of information theory. After introducing the basic constructions, we will see three potential justifications for essentially the same proposal: axiomatic (§4.3.2), thermodynamic (§4.3.3), and coding theoretic (§4.3.4).

4.3.1 Information Theoretic Cost

Information theory, as developed by Shannon (1948), concerns optimization of communication codes, and is thus essentially about dealing with limited resources. It is perhaps unsurprising, therefore, that the theory would appear in analyses of resource limited decision making.

Suppose we have a random variable X taking on some possible values x, with distribution $p(\mathsf{X} = x)$, or simply $p(x)$. We will observe successive values of X and our task is to communicate those values to some third party. If before sending any messages we have the opportunity to decide on a *code* for each value x—a mapping from x to a binary string of 1's and 0's—what would be the best code to optimize for average length of the signal sent? Assuming perfect communication (a "noiseless channel"), a moment's thought suggests that the highly probable values will appear more often and so should be given the shortest codes. Since there are only so many short codes, some of the less probably values will need to be assigned longer codes.

One can show that the number of bits used to code x in an optimal code for distribution p is roughly $-\log_2 p(x)$. Consider the following examples:

Example 20. For instance, if X has 8 possible values, all equally likely with $q(\mathsf{X} = x) = 1/8$, then an optimal code for q would use $-\log \frac{1}{8} = \log 8 - \log 1 = 3$ bits for each value of X. That is, each value of X is assigned to exactly one of the binary strings of length 3.

On the other hand, the distribution p in this table can have different length codes:

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
$p(x)$	1/2	1/8	1/8	1/16	1/16	1/16	1/32	1/32
code	1	001	000	0111	0101	0100	01101	01100

On the second line is an example of an optimal code (the so-called Huffman code). Observe that the code lengths for p are 1, 3, 4, and 5, in different proportions.

Computing the average number of bits that need to be transmitted in such an optimal code gives us $\sum_x p(x)(-\log p(x))$, or:

$$H(p) \quad = \quad -\sum_x p(x) \log p(x) \tag{4.5}$$

This quantity $H(p)$ is known as the *entropy* of X. Aside from the coding intuition, entropy is often understood as a measure of "disorder" or noise. If we construe $-\log_2 p(x)$ not as the length of a code, but as the *information* conveyed by x, then $H(p)$ measures how much information is conveyed on average. Continuing Example 20:

Example 21. Obviously the average for the distribution q in Example 20 is 3 bits, corresponding to an entropy of 3 for that distribution. Intuitively, such a distribution is unpredictable.

For the second distribution, p, the outcome is more predictable: half the time we will expect to see x_1—thus, on the coding interpretation we would expect to see a message of length 1 half the time—and three fourths of the time we expect to see one of only three values, x_1, x_2, or x_3. Correspondingly, the entropy in this case is lower, at about 2.3.

Next, imagine that we had a code that was optimized for a distribution q, but that the true distribution is actually p. How "inefficient" would it be to use this code for q? In particular, how many extra bits would we need to code samples drawn from p when using this code optimized for q? This is given by the following:

$$\begin{aligned} D_{KL}(p\,||\,q) &= \text{“average length of codes optimized for } q\text{”} - \\ &\quad \text{“average length of codes optimized for } p\text{”} \\ &= -\sum_x p(x)\log q(x) - \big(-\sum_x p(x)\log p(x)\big) \\ &= \sum_x p(x)\log\frac{p(x)}{q(x)}, \end{aligned} \tag{4.6}$$

known as the *Kullback-Leibler divergence* (or *KL-divergence*), also known as *relative entropy*. A special case is when q is the uniform distribution: $q(x) = 1/K$, where K is the number of possible values of X. Because $H(q) = \log K$, we have $D_{KL}(p\,||\,q) = \log K - H(p)$.

Example 22. Continuing again with Example 20, the divergence of p from q is $D_{KL}(p\,||\,q) = \log 8 - H(p) \approx 0.7$. In other words, adopting a scheme that assigns every value of X a code of length 3, our message would on average be 0.7 bits too long.

Similarly, using a code optimized for p when the true distribution is q (uniform) results in KL-divergence of about 0.6, as we are using inefficient codes of length 4 and 5 half the time.

The bold idea that we will be exploring here is that KL-divergence can be seen not just as a measure of coding inefficiency, but also as a measure of the costs involved in the transformation of one action distribution into another.

Definition 14 (Information theoretic cost of a strategy)**.** Hold fixed a default action distribution $\delta \in \Delta(A)$. Define the cost $C(\pi)$ of implementing alternative program π as:

$$C(\pi) \;=\; D_{KL}(\pi\,||\,\delta).$$

Then, the present version of cost-adjusted value (1.1) becomes, for some $\beta > 0$:

$$\mathbf{V}(\pi) \;=\; \mathbf{U}(\pi) - \frac{1}{\beta}\,D_{KL}(\pi\,||\,\delta), \tag{4.7}$$

where $\mathbf{U}(\pi) = \sum_a \pi(a)\,\mathbb{E}\mathrm{U}(a)$ in this case (recall Eq. (2.2)).

Theorem 6. With cost measured by (a multiple of) KL-divergence (Def. 14), the strategy $\pi^* \in \Delta(A)$ that maximizes value $\mathbf{V}(\pi^*)$ according to Eq. (4.7) is

$$\pi^*(a) \;\propto\; \delta(a)\,e^{\beta\cdot\mathbb{E}\mathrm{U}(a)} \tag{4.8}$$

The proof of this result is given in a technical appendix to this chapter, §4.A.3.

This connects directly to the discussion of probability matching and the softmax rule (Example 7, §4.1.2) via the following easy consequence of Theorem 6.

Corollary 2. When the default distribution δ is uniform—that is, $\delta(a) = 1/|A|$ for each a—then the optimal strategy simplifies to:

$$\pi^*(a) \quad \propto \quad e^{\beta \cdot \mathbb{E}\mathrm{U}(a)},$$

that is, precisely the softmax distribution (4.2). This is equivalent to assuming that we are *maximizing (a multiple of) entropy* $H(\pi)$, rather than minimizing KL-divergence to δ.

The cost function in Def. 14 has been proposed and discussed across many different literatures, most explicitly by Mattsson and Weibull (2002) and later by Ortega and Braun (2013), though versions have also appeared in neuroscience and in reinforcement learning (see §4.3.4 below), and even in recent theories of linguistic pragmatics (Example 32 in §6.6). The question is, how can we justify Def. 14 as a sensible analysis of resource costs?

4.3.2 The Cost of Control

Mattsson and Weibull (2002) suggest that $D_{KL}(\pi)$ can be justified axiomatically, so long as the axioms appear intuitive. The starting point is a control cost function $\mathfrak{c}$, measuring the cost $\mathfrak{c}(\pi, \delta)$ of moving from default distribution δ to some new distribution π. It is assumed to be continuous in π and δ (as vectors of probabilities), and we presume that $\mathfrak{c}(\pi, \delta) = 0$ if and only if $\pi = \delta$. Mattsson and Weibull then consider three further axioms:

> **Axiom 1**. Cost is invariant under relabeling. That is, if π^* and δ^* are permutations of π and δ, respectively, then $\mathfrak{c}(\pi, \delta) = \mathfrak{c}(\pi^*, \delta^*)$. In other words, all "atomic" events are treated the same, and only the probability values matter.
>
> **Axiom 2**. It is costlier to rule out a larger set of actions than to rule out a proper subset. Specifically, if $A_1 \subset A_2$, then the uniform distribution on A_1 is no more costly than the uniform distribution on A_2.
>
> **Axiom 3**. If A is divided into two subsets, B and C, then the cost $\mathfrak{c}(\pi, \delta)$ of moving from δ to π (both distributions on A) is the same as the *sum* of costs of (i) moving from the probabilities $\delta(B)$ and $\delta(C)$ to $\pi(B)$ and $\pi(C)$, and (ii) the weighted (by $\pi(B)$ and $\pi(C)$, respectively) costs of moving from $\delta(\cdot\,|B)$ to $\pi(\cdot\,|B)$ and $\delta(\cdot\,|C)$ to $\pi(\cdot\,|C)$.

The key result follows from a generalization of the derivation of entropy due to Shannon (1948).

Theorem 7 (Hobson 1969)**.** Any cost function that satisfies Axioms 1-3 is proportional to KL divergence. That is, if Axioms 1-3 hold of $\mathfrak{c}$, then

$$\mathfrak{c}(\pi, \delta) \quad \propto \quad D_{KL}(\pi \,||\, \delta);$$

that is, we recover exactly Definition 14.

As an alternative interpretation, also from the the economics literature, Matějka and McKay (2015) suggest we could take $\mathfrak{c}$ as a measure of "rational inattention," with entropy reduction measuring something like how many questions an agent needs to ask (possibly to themselves) in order to narrow down their decision. On either interpretation, the ultimate success of this analysis depends on how plausible we find Axioms 1-3.

Figure 4.2: Compressing X (with volume V_1) to Y (with volume V_2). Adapted from Feynman (1998), Fig. 5.2. It is assumed that there is no difference between the total energy in X and in Y. Because we had to put in some energy to compress, this "missing" energy that got absorbed into the heat bath is what we want to measure. The point is that this is a change in entropy.

4.3.3 Information and Energy

An alternative justification, offered by Ortega and Braun (2013) (see also Ortega 2010 and Genewein et al. 2015), starts from the commonplace that computation requires physical energy. In fact, it has been suggested that we should understand a computer just to be "an engine that dissipates energy in order to perform mathematical work" (Bennett, 1982, p. 906). Each time a choice gets made—each time a computation step occurs—there is some thermodynamic work that must be done in order to "record" or "register" that choice. The core idea of this approach is to postulate a general correspondence between units of information (e.g., bits) and units of thermodynamic work, so that the amount of information processed in moving from distribution δ to distribution π—as measured by KL-divergence—can be related to the *amount of (e.g., thermodynamic) work* involved in such processing.

Consider a simple thought experiment about ideal gases from Feynman (1998). Imagine a box X of gas with volume V_X containing N atoms, and suppose we wanted to compress the gas to a smaller box Y with volume V_Y. Assuming we can do this in a way that keeps the temperature T constant, the physical work involved in this process is given by the negative of the so-called *(Helmholtz) free energy difference* ΔF between the state before and after the compression. This is the well known equation:

$$\Delta F \quad = \quad NkT \ln \frac{V_X}{V_Y} \tag{4.9}$$

where k is the Boltzmann constant. As Feynman explains, the thought experiment still makes sense even if we only have $N = 1$ atom in the box. We can think of the compression as reducing our uncertainty about the location of the atom; see Fig. 4.2.

The free energy difference ΔF thus quantifies how much work must go in to achieve this reduction. For example, if compressing cuts the volume of X in half, we obtain $kT \ln(2)$, which von Neumann (1966) famously estimated as the "thermodynamic minimum of energy per elementary act of information" (p. 66).

Ortega and Braun (2013) suggest extending the thought experiment by partitioning X into some cells a. Suppose our primary interest is locating in which cell the particle resides. We then have an obvious distribution $\delta(a) = V_X(a)/V_X$, where $V_X(a)$ is the volume of a within X, quantifying our uncertainty about the location of the particle before compression. Likewise, for the probability after compression, let $\pi(a) = V_Y(a)/V_Y$, where $V_Y(a)$ is the volume of a in Y. Assuming that the atom could equally likely be in any position within box X before compression, and anywhere in Y afterward, the probabilities δ and $\pi = \delta(\cdot \,|\, Y)$ intuitively quantify our uncertainty about which cell the atom occupies. See Fig. 4.3.

Figure 4.3: In this elaboration of Fig. 4.2, adapted from Ortega and Braun (2013), there are four actions corresponding to the four colors, with area representing probability mass. Here we consider both the separate amounts of work required to compress each compartment, and the overall reduction in entropy with respect to the coarsened (i.e., compartmentalized) space.

The free energy difference can thus be written in the following way (where now $N = 1$):

$$\begin{aligned}
\Delta F &= kT \ln \frac{V_X}{V_Y} \\
&= kT \sum_a \pi(a) \ln \Big(\frac{V_X}{V_Y} \frac{V_X(a)}{V_X(a)} \frac{V_Y(a)}{V_Y(a)} \Big) \\
&= \sum_a \pi(a)\, kT \ln \frac{V_X(a)}{V_Y(a)} + kT \sum_a \pi(a) \ln \Big(\frac{V_X}{V_X(a)} \frac{V_Y(a)}{V_Y} \Big) \\
&= \mathbb{E}_{a\sim\pi} \Delta F[a] + kT \sum_a \pi(a) \ln \frac{\pi(a)}{\delta(a)} \\
&= \mathbb{E}_{a\sim\pi} \Delta F[a] + kT\, D_{KL}(\pi \,||\, \delta).
\end{aligned}$$

In other words, the overall thermodynamic cost of this operation can be broken down into a sum of two terms. The first term is the expected free energy difference *per compartment*, written here as $\Delta F[a]$. This can be seen as the work involved in compressing just that single cell a. However, this does not exhaust the work performed. The second term captures the work involved in *reducing our uncertainty* about which compartment houses the particle. This is measured by the KL-divergence from the old distribution δ to the new distribution π, with a conversion factor of kT. Note that this gloss depends on π being sharper than δ.

At this point Ortega and Braun (2013) suggest a substantive reinterpretation by switching from the location of a particle in a box to an action performed by an agent. Suppose instead that a ranges over actions and that δ and π are action distributions. The suggestion is to replace the per-compartment free energy $\Delta F[a]$ with the *disutility* $-\mathrm{U}(a)$ of opting for action a. In the (very) special case that the action happens to be compressing a cell within an ideal box, this is not a reinterpretation at all. But we might imagine more generally that $\mathrm{U}(a)$ gives some independent measure of how good (or "not bad") action a is; e.g., $\mathrm{U}(a) = \mathbb{E}_s\, u(s, a)$ could be defined as expected utility over some assumed distribution on relevant states.

On this reinterpretation, minimizing the free energy difference is tantamount to maximizing the expected utility of using π minus the cost of overriding the default δ, in line with Eq. (4.7):

$$-\Delta F \quad = \quad \mathbf{U}(\pi) - kT\, D_{KL}(\pi \,||\, \delta),$$

where $kT = 1/\beta$. This explains why β is often called the inverse temperature. As stated in Theorem 6, the maximum is given precisely by the softmax decision rule.

The specific conversion factor kT is motivated by the theory of ideal gases. In any concrete physical system (such as a brain or a computer) the costs of information processing will be greater. The bold hypothesis (cf. §1.4) is that there will be *some* conversion factor β that

accurately characterizes this cost, and in particular that the (opportunity) cost of moving from δ to π will be proportional to the relative entropy, $D_{KL}(\pi \,||\, \delta)$. To the extent that this hypothesis holds, it gives a resource rational justification for random behavior, and specifically for (posterior) probability matching.

4.3.4 Information Coding

The discussion of information theory so far has focused on one-shot decision problems and probability distributions over single actions. Interestingly, moving to the setting of (fully observable) sequential decision problems (§2.2) supplies us with a canonical default distribution. A natural choice is to let the default probability of a be something like the long-run frequency of action a. Then we could measure the complexity of a strategy—relative to the probabilities characterizing an environment—as the average cost of *diverging* from that default distribution to a situation-specific distribution.

More precisely, recall (Remark 5) the *stationary distribution* $P_\sigma(\mathsf{S} = s)$ for a given MDP strategy σ, capturing the long-term frequency with which we expect to visit state s. We can define an analogous random variable A for the stationary distribution on actions:

$$P_\sigma(\mathsf{A} = a) \quad = \quad \sum_{s \in S} P_\sigma(\mathsf{S} = s)\, \sigma_s(a). \tag{4.10}$$

We thus have two random variables S and A with distributions $P_\sigma(\mathsf{S})$ and $P_\sigma(\mathsf{A})$. Note that randomness in action selection is actually inherited from (apparent) randomness in the environment. Even if the agent were to behave deterministically in any given state, the stationary distribution $P_\sigma(\mathsf{A})$ will still generally be non-deterministic.

Taking $P_\sigma(\mathsf{A})$ as the default, we can then measure cost by averaging over KL-divergences:

Definition 15 (Information theoretic cost of a sequential strategy)**.** The cost $C(\sigma)$ of implementing sequential strategy σ is,

$$C(\sigma) \quad = \quad \mathbb{E}_{s \sim P_\sigma} D_{KL}\big(\sigma_s \,||\, P_\sigma(\mathsf{A})\big), \tag{4.11}$$

that is, the KL-divergence from the default distribution $P_\sigma(\mathsf{A})$ to the situation-specific distribution σ_s, *averaged over states*.

Assuming we can justify KL-divergence as the right way of measuring the cost of moving from a default to a sharpened action distribution (e.g., via §4.3.2 or §4.3.3), this seems like the right extension to the sequential setting; the question is, how much on average does it cost to move from the stationary distribution on actions to a situation-specific disposition?

Very similar to the derivation of Theorem 6, we have the sequential generalization of it:

Theorem 8. Where $\mathbf{U}(\sigma)$ is defined as in Eq. (2.9) and $C(\sigma)$ as in Eq. (4.11), the value

$$\mathbf{V}(\sigma) \quad = \quad \mathbf{U}(\sigma) - \frac{1}{\beta}\, C(\sigma)$$

is maximized by σ^*, defined so that for each $s \in S$ and $a \in A$:

$$\sigma^*_s(a) \quad \propto \quad P_{\sigma^*}(\mathsf{A} = a)\; e^{\beta Q^{\sigma^*}(s,a)}, \tag{4.12}$$

where Q^{σ^*} is the Q-function introduced in Eq. (2.11).

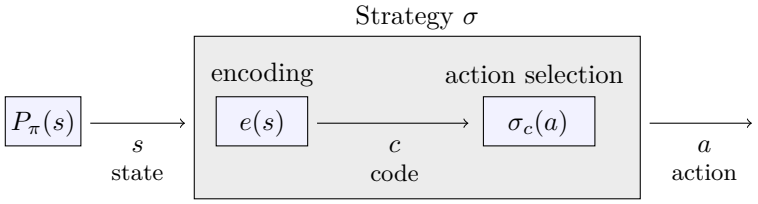


Figure 4.4: Inside a strategy σ for the Markov decision process. The agent is assumed to encode the state s via an encoding $c = e(s)$, and then draw an action from the (typically lossy) action distribution σ_c. Higher mutual information between state and action distributions presumably implies more complex encodings. Adapted from Lai and Gershman (2021), Fig. 1A.

Observe that Eq. (4.12) is just like Eq. (4.8), the original default-relative softmax expression, except that Q replaces expected utility of a, and the default distribution is $P_{\sigma^*}(\mathsf{A})$.

There is another way of interpreting Definition 14, revealed in the following derivation:

$$\begin{aligned} C(\sigma) &= \mathbb{E}_{s\sim P_\sigma} D_{KL}\big(\sigma_s \,||\, P_\sigma(\mathsf{A})\big) \\ &= \sum_{s,a} P_\sigma(\mathsf{S}=s)\,\sigma_s(a) \log \frac{\sigma_s(a)}{P_\sigma(\mathsf{A}=a)} \\ &= \sum_{s,a} P_\sigma(\mathsf{S}=s)\,\sigma_s(a) \log \frac{P_\sigma(\mathsf{S}=s)\,\sigma_s(a)}{P_\sigma(\mathsf{S}=s)P_\sigma(\mathsf{A}=a)} \\ &= \sum_{s,a} P_\sigma(\mathsf{S}=s,\mathsf{A}=a) \log \frac{P_\sigma(\mathsf{S}=s,\mathsf{A}=a)}{P_\sigma(\mathsf{S}=s)P_\sigma(\mathsf{A}=a)}, \end{aligned} \tag{4.13}$$

the equality $P_\sigma(\mathsf{S}=s,\mathsf{A}=a) = P_\sigma(\mathsf{S}=s)\,\sigma_s(a)$ following from Eq. (4.10). This last expression (4.13) is known as the *mutual information* between S and A, often written $I(\mathsf{S};\mathsf{A})$. What this simple derivation shows is thus that Definition 14 is equivalent to saying we measure cost by the mutual information between the (stationary) state and action distributions.

The concept of mutual information—also called *predictive information*—has a long and significant history in neuroscience, dating back to the "efficient coding hypothesis" of Barlow (1961). As a measure of the information carried by one random variable about another, mutual information has been proposed as a way of measuring how well a population of neurons encodes some class of stimuli (e.g., Brunel and Nadal 1998). Presumably, a better encoding will demand greater resources and thus incur greater cost. This is only exacerbated by the various sources of noise in the nervous system (§4.1.1), which seems to demand considerable *redundancy* in neural coding in order to overcome the noise (Niven and Laughlin, 2008).

Motivated by considerations like these, mutual information has been independently proposed as a cost term in reinforcement learning and theoretical neuroscience (e.g., Tishby and Polani 2011; Still and Precup 2012). The idea is intuitive: the more an agent's action distribution carries information about the state—that is, the more it would be possible to predict the state from the action—the more the agent must be differentially responding to different states. Responding differentially in turn requires that the agent somehow *encode* states appropriately. More sensitive responses require better encodings, and better encodings demand resources. See Fig. 4.4 for one way of illustrating the core idea, adapted from Lai and Gershman (2021).

We will return to these ideas below in §5.3.3 when discussing resource rational analysis in cognitive science. For now, it merits emphasis that all of these approaches to information theoretic cost are relatively abstract. Under auxiliary assumptions, it can be proven that

mutual information does place a lower bound on the amount of memory needed to implement a given strategy (see, e.g., Shalizi and Crutchfield 2001, Thm. 5). At the same time, the perspective is often promoted as a "model-neutral" approach that makes minimal assumptions about computational architecture (see, e.g., Bialek et al. 2001; Ortega 2010).

4.4 Randomness as Resource

A different starting point on the question of when randomness could be (resource) rational takes it to be, not a default, but a *resource*. Perhaps the default behavior of an agent is actually deterministic: despite the presence of neural noise, when put in a particular circumstance, without deliberation the agent would always carry out the same (perhaps suboptimal) strategy. On this view, a source of random noise—whether originating in the nervous system or external to the agent—might actually help the agent improve upon suboptimal deterministic strategies. Here we explore two ways that might happen.

4.4.1 Stochastic Resonance

Imagine a system with a binary "output" variable Y that depends on another binary "input" variable X. The system operates with a *threshold* $\theta > 0$, such that Y takes on value 1 just in case its input value surpasses θ. If $\theta < 1$, then such a system would behave deterministically, with Y always being present (i.e., having value 1) whenever X is present. If $P(\mathsf{X} = 1) = 0.5$, say, then they will both be present together half the time.

But suppose $\theta > 1$. Then Y will not be sensitive to the presence of X at all. The idea of *stochastic resonance* is that it could be helpful in such a situation to add some random noise to the signal X. Imagine a third variable Z whose distribution is Gaussian with mean 0 and variance $\mathbb{v}$, and let:

$$\mathsf{Y} \quad = \quad \begin{cases} 1 & \text{if } \mathsf{X} + \mathsf{Z} > \theta \\ 0 & \text{otherwise.} \end{cases}$$

If $\mathbb{v} = 0$, then we are back in the previous case. But it now seems intuitive that having positive variance could actually be useful here. Let the objective be to maximize the *mutual information* between X and Y (Eq. (4.13) above), and assume $\theta = 1.25$. With $\mathbb{v} = 0$, the mutual information is 0. However, if the variance is instead, say, $1/4$, then $I(\mathsf{X};\mathsf{Y}) > 1/10$. See Fig. 4.5. Intuitively, even though Y sometimes "misfires" when X is absent, it exhibits a strong enough correlation with X so that some signal passes through the system. Of course, if $\mathbb{v}$ is too high, X and Y will again become decorrelated. There is in fact an optimal noise level (Bulsara and Zador, 1996).

Some evidence of stochastic resonance has been found in the biological world. In one famous study, Russell et al. (1999) showed that injecting the optimal amount of electric noise into the surrounding water of a paddle fish allowed the fish to detect planktonic prey more effectively. Whether any organisms use "intrinsic" sources of randomness—e.g., intrinsic neural noise—for such purposes remains a matter of debate; see McDonnell and Abbott (2009) for a review.

It is important to emphasize that stochastic resonance depends on the threshold θ being a fixed feature of the organism. If the threshold can be set by the agent, then noise can again confer no benefit. Thus, whether stochastic resonance counts as a resource rational justification for randomness depends in part on whether we consider the thresholding mechanism to be an inherent part of the agent in question (recall Fig. 1.1). Some have suggested that dynamically setting the threshold could itself be a costly process (cf. McDonnell and Abbott 2009), but substantiating this suggestion requires a further argument, perhaps drawing on ideas about resource requirements such as memory. This leads to our last rationale for randomness.

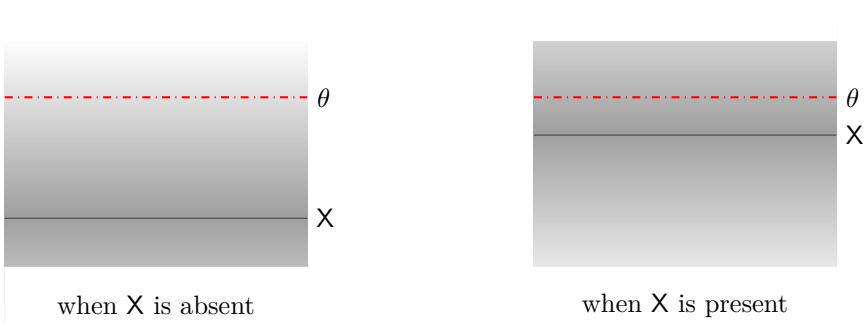


Figure 4.5: An example of stochastic resonance. Gray shading denotes Gaussian noise ($\mathsf{X}+\mathsf{Z}$), with darker signifying greater probability. In order for Y to register the presence of X, activity needs to surpass the dotted red threshold (θ).

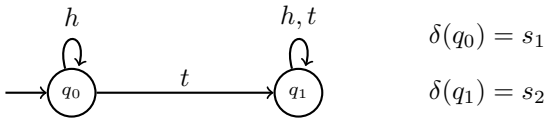
4.4.2 Randomization and Memory

We have already seen in Chapter 3 that agents with a finite memory appear to benefit from randomization. Recall that a general way of enforcing finite memory is to assume the agent can only be in one of finitely many behaviorally distinct states; in other words, the agent's deliberative strategy must be implementable by a finite automaton (Defs. 10 and 13). On this approach memory is not formalized explicitly or assumed to have any particular format. Instead, restrictions on memory are captured by restrictions on the size of the automaton.

The work on polarization by Wilson (2014) discussed in §3.5 was partly anticipated in earlier work by Hellman and Cover (1970, 1971). They looked at the problem of *hypothesis testing* by finite automata. Given a prior over hypotheses and a likelihood function for observations given hypotheses, the best strategy for an unbounded agent is simply to guess the hypothesis with highest posterior probability. However, this strategy generally requires unbounded memory. The question is, what happens when memory is bounded? The following simple illustration is adapted from Icard (2021), inspired by Hellman and Cover (1970, 1971):

Example 23 (Hypothesis testing with a finite memory)**.** Imagine there are just two states, $S = \{s_1, s_2\}$ ("coin 1" or "coin 2"), and there are two possible observations $O = \{h, t\}$ ("heads" and "tails"). The two coins are equally likely a priori—$p(s_1) = p(s_2) = 1/2$—and the likelihoods are $l(h|s_1) = 0.99$ and $l(h|s_2) = 0.9$. Thus, heads is more likely than tails for both coins, but it is vastly more likely in s_1. Suppose that we will encounter T observations ("flips of the coin"), after which we have to guess which coin generated the observations. That is, $A = S$. What is the best we can do if our strategy must be implemented by a finite state machine? Specifically, suppose our finite state machine can have at most *two* states.

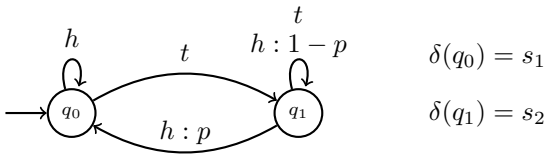
If we are forced to use a deterministic strategy, the best we can do is the following:



We begin by guessing coin 1 (in q_0), but as soon as the first tails is observed we guess coin 2, and that is our guess forever after. Note, incidentally, that this is just the "trigger" strategy discussed in Example 15 of the previous chapter. If $T = 100$, this strategy will be correct about 68% of the time. If $T = 200$ this drops to around 57%, and by the time we reach $T \geq 500$, the

agent's performance is no better than chance: there is almost always at least one flip of tails, no matter whether it is coin 1 or coin 2.

If we are allowed to employ stochastic transitions, however, we can do much better.



This strategy is just like the deterministic strategy above, except that in state q_1 there is some probability p of returning back to q_0 (guessing s_1). If we see enough heads in a row, even if p is very small, we will eventually make that transition. Notice that this is just "stochastic tit-for-tat" discussed in Example 18. The analysis from Hellman and Cover (1970, 1971) shows $p \approx 0.03$ leads to the best performance in this specific case. With more than a few dozen observations, the strategy will be correct about 77% of the time, no matter how large T is.

The intuition behind Example 23 is clear enough: random transitions provide the memory-bounded agent with a means of escaping hypotheses that may have initially seemed plausible, but become decreasingly plausible given more data.

An even simpler illustration comes from a well known example in game theory, where strategies are restricted to those implementable by an automaton with just a single state:

Example 24 (Absent Minded Driver)**.** This is a sequential decision problem introduced by Piccione and Rubinstein (1997). Suppose an agent is out late at a concert and needs to drive home afterward. Taking the first turn on the route would lead to a swamp, which is highly undesirable. The second turn, however, leads home. If the agent does not turn at all, then they arrive at a hotel and have to pay for a room, which is again slightly undesirable.

Because they are very tired, they will not remember whether they are at the first or second stop. A depiction of the scenario appears in Fig. 4.6. We now impose the requirement that the strategy be implementable by a *one-state* machine. What is the best deterministic strategy? Simply never turn, giving an expected utility of 1. By the argument from convexity (§4.2), no probabilistic mixture of deterministic one-state strategies does better.

But what if we can employ a behavioral strategy and randomize our decision at each step, e.g., turning with probability p at both choice points? Call this strategy σ. Then:

$$\begin{aligned} \text{Expected utility of strategy } \sigma &= p(1-p)4 + (1-p)(1-p) \\ &= 2p - 3p^2 + 1. \end{aligned}$$

This takes on a maximum when $p = 1/3$, giving expected utility $4/3$. The best strategy is behavioral, and no mixed strategy can achieve the same.

Both Example 23 and Example 24 demonstrate an intriguing failure of Theorem 5, the result that every behavioral strategy is equivalent to some mixed strategy. While this is true when we ignore the resource costs required to implement the strategy—even if we impose the requirement that the strategy be implementable by some computer program; recall Corollary 1—it fails if we put an upper bound on available memory. Both examples here reveal cases in which the performance of a probabilistic automaton of a given size cannot be matched by a probabilistic mixture of deterministic automaton of the same size.

Consequently, if it is a fixed feature of an agent that their memory capacity is limited, that by itself raises the possibility that randomization could confer a definitive benefit. This

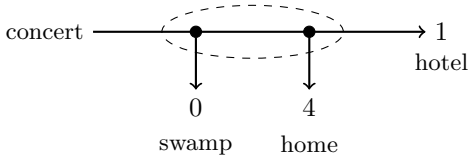


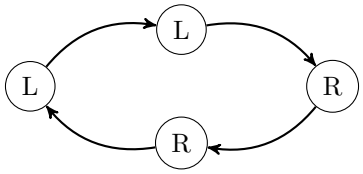
Figure 4.6: Schematic for the absent minded driver (Example 24). The dotted oval around the two choice points signifies that the agent cannot distinguish between the two. So any admissible strategy will have to treat these choice points identically.

is especially plausible, as Example 23 makes salient, when the amount of data confronting the agent vastly outstrips available memory and processing capacity (cf. Christiansen and Chater 2016 on the specific challenges of real-time language processing). It is argued in Icard (2021) that this situation is common in humans and other agents of sufficient complexity, which might in turn (resource) rationalize widespread randomization. Here is one last illustration.

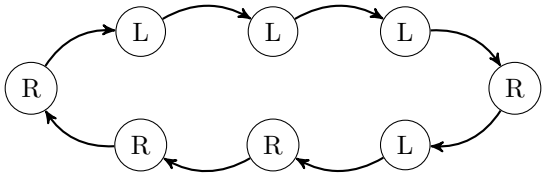
Example 25 (Random Pigeons)**.** A study by Machado (1993) had pigeons producing sequences of pecks, on a left key (L) or a right key (R). The goal is for the sequence to be *normal* at level k, meaning that the proportion of all k-length subsequences should be uniform. This can be construed as a kind of search problem. Imagine there is some "prize" k-length sequence, such that producing that sequence gives positive probability of a large prize. Assuming all k-length sequences are equally likely, the goal is to produce each subsequence the same number of times so as to maximize the overall chance of a prize. In these experiments, reward in the form of food is delivered throughout the trial, with probability of reward proportional to how balanced the produced sequence is among subsequences of length k.

There exists an optimal deterministic strategy for this task, provided by so called de Bruijn sequences. In the case of $k = 1$, the optimal strategy is simple alternation, captured by this deterministic automaton (with no observations, and the action, L or R, appearing in the state):

guaranteeing the same number of L as R throughout. Pigeons routinely learn this strategy. For $k = 2$, the optimal strategy is slightly less obvious:



This balances all subsequences of length 2. With slightly more difficulty, pigeons are able to learn this strategy (Machado, 1993, Exp. 1). The answer for $k = 3$, however, is not at all obvious. It turns out that repeating the following sequence is the optimal strategy:



No pigeons in this study demonstrated such behavior. More interesting, however, is the finding that almost all of the birds in this experiment behaved in a way that was eventually indistinguishable from a random Bernoulli process. The produced sequences passed a barrage of statistical tests, suggesting that they had learned to achieve variability by acting more or less randomly (Machado, 1993, Exp. 2). Evidently, when the memory requirements become too demanding, the pigeons in this study latch on to the optimal alternative, which in this case is simply to act randomly, all but guaranteeing approximate balance for long enough sequences.

4.5 Conclusion

There have been a number of arguments for the rationality of randomization across various literatures. The need for randomized experiments to control for unknown confounders in causal inference is one famous example (going back at least to R. Fisher). Game theorists have argued that randomized strategies might confer protection against being outwitted in strategic scenarios (§3.1). Meanwhile, researchers in reinforcement learning have suggested that randomizing one's decisions—especially in accord with the softmax distribution in (4.2)—can help balance *exploitation* of known rewards with *exploration* for even higher sources of reward (recall §2.3.2). In a very general sense, each of these arguments relates to resource limitations. For instance, they all stem from a concern that formulating (and computing with) a suitable *prior distribution* on relevant states of the world may be difficult.

Such appeals to resources tend to be informal. The approaches discussed in the present chapter offer concrete rationales for random behavior by reference to a formalization of resource costs. Aside from shedding light on the intriguing question of when randomness in behavior might in fact be resource rational, an important function of this chapter has been to introduce the information theoretic approach to understanding resources more generally, as a potential alternative to the mechanistic approach grounded in explicit computational models. Both show how the convexity argument in §4.2 can fail, though in rather different ways. We will see examples of both in the next chapter on resource rational analysis in cognitive science.

4.A Technical Appendices

These appendices include technical material from the present chapter. In §4.A.1 we derive the softmax function from the random utility model introduced in §4.1.2. Next, in §4.A.2 we prove Theorem 5, showing that every behavioral strategy is equivalent to a mixed strategy. Finally, §4.A.3 is dedicated to a derivation of the softmax decision rule from imposing an information theoretic cost on any sequential decision making strategy.

4.A.1 Random Utility Derivation of Softmax

The following argument appears in Luce and Suppes (1965), who in turn attribute it to E. Holman and A.A.J. Marley. The probability that action a has maximal value U_a can be

obtained by integrating over all possible maximal values:

$$\begin{aligned}
\text{Probability that } \mathsf{U}_a \text{ is maximal} &= \int_{-\infty}^{\infty} \big[P(\mathsf{U}_a = \mathsf{x}) \cdot \prod_{a' \neq a} P(\mathsf{U}_{a'} \leq \mathsf{x})\big]\, d\mathsf{x} \\
&= \int_{-\infty}^{0} e^{\beta V(a) + e^{\mathsf{x}\beta V(a)}} \cdot \prod_{a' \neq a} \big[\int_{-\infty}^{\mathsf{x}} e^{\beta V(a') + e^{\mathsf{y}\beta V(a')}}\, d\mathsf{y}\big] d\mathsf{x} \\
&= \int_{-\infty}^{0} \big[e^{\beta V(a)} \cdot e^{e^{\mathsf{x}\beta V(a)}} \cdot \prod_{a' \neq a} e^{e^{\mathsf{x}\beta V(a')}}\big]\, d\mathsf{x} \\
&= \int_{-\infty}^{0} \big[e^{\beta V(a)} \cdot \prod_{a'} e^{e^{\mathsf{x}\beta V(a')}}\big]\, d\mathsf{x} \\
&= \int_{-\infty}^{0} \big[e^{\beta V(a)} \cdot e^{\sum_{a'} e^{\mathsf{x}\beta V(a')}}\big]\, d\mathsf{x} \\
&= \frac{e^{\beta V(a)}}{\sum_{a'} e^{\beta V(a')}},
\end{aligned}$$

which is precisely the softmax function.

4.A.2 From Behavioral to Mixed Strategies

Let $\Omega = \bigcup_{t<T} \mathcal{O}_t$, with $\mathcal{O}_t$ all observation histories consisting of exactly t actions (and $t+1$ observations). Define Σ_T to be the set of all functions from Ω to A. Note that a distribution $\mu \in \Delta(\Sigma_T)$ canonically gives a distribution on strategies in Σ_t for any $t < T$ as well. We abuse notation and write $\mu(\varsigma)$ for the sum of probabilities μ assigns to strategies in Σ_T that extend ς.

Recall the definition of μ. For each $\varsigma \in \Sigma_T$:

$$\begin{aligned}
\mu(\varsigma) &= \prod_{\omega \in \mathcal{O}_t : t<T} \sigma_\omega(\varsigma(\omega)) \\
&= \prod_{\omega \in \Omega} \sigma_\omega(\varsigma(\omega)).
\end{aligned}$$

We first need to check that μ is a proper probability distribution, specifically that it sums to unity. The crucial observation is the following identity for all $t < T$:

$$\sum_{\varsigma \in \Sigma_{t+1}} \mu(\varsigma) = \sum_{\varsigma \in \Sigma_t} \mu(\varsigma) \times \Big(\prod_{\omega \in \mathcal{O}_t} \sum_{a \in A} \sigma_\omega(a) \Big) \tag{4.14}$$

The reason this holds is that each function in Σ_t gets extended in all possible ways to a function in Σ_{t+1}. Summing over (the weights of) all such extensions gives the term on the right. Using this identity T times leads to the result:

$$\begin{aligned}
\sum_{\varsigma \in \Sigma_T} \mu(\varsigma) &= \Big(\prod_{\omega \in \mathcal{O}_0} \sum_{a \in A} \sigma_\omega(a) \Big) \times \ldots \times \Big(\prod_{\omega \in \mathcal{O}_{T-1}} \sum_{a \in A} \sigma_\omega(a) \Big) \\
&= \prod_{t<T} \prod_{\omega \in \mathcal{O}_t} \sum_{a \in A} \sigma_\omega(a) \\
&= \prod_{\omega \in \Omega} 1 \\
&= 1,
\end{aligned}$$

since each σ_ω is assumed to be a probability distribution summing to one.

We now show that σ and μ produce the same distribution on histories. The crucial point is that, for any observation history $\omega = o_0, a_0, \ldots, o_{T-1}, a_{T-1}, o_T$, the probability that σ generates each action a_t at observation history $\omega \restriction t$ is the same as the *sum* of probabilities that μ assigns to deterministic strategies ς, each of which generates a_t at $\omega \restriction t$. That is, where Σ^*_T is the set of all those ς such that $\varsigma(\omega \restriction t) = a_t$ for all $t < T$, we need:

$$\prod_{t<T} \sigma_{\omega\restriction t}(a_t) \quad = \quad \sum_{\varsigma\in\Sigma^*_T} \mu(\varsigma). \tag{4.15}$$

To show (4.15) consider a variation on Eq. (4.14), for each $t < T$:

$$\sum_{\varsigma\in\Sigma^*_{t+1}} \mu(\varsigma) \quad = \quad \sum_{\varsigma\in\Sigma^*_t} \mu(\varsigma) \times \sigma_{\omega\restriction t}(a_t) \times \Big(\prod_{\omega'\neq(\omega\restriction t)} \sum_{a\in A} \sigma_{\omega'}(a) \Big), \tag{4.16}$$

where Σ^*_t is defined analagously to Σ^*_T. The argument is the same as for (4.14), except that we remove all variations for the observation histories $\omega \restriction t$, since those are held fixed at a_t. Contuing analogously to the previous argument, using Eq. (4.16) T times, we obtain:

$$\begin{aligned}
\sum_{\varsigma\in\Sigma^*_T} \mu(\varsigma) &= \prod_{t<T} \Big[\sigma_{\omega\restriction t}(a_t) \times \Big(\prod_{\omega'\neq(\omega\restriction t)} \sum_{a\in A} \sigma_{\omega'}(a) \Big) \Big] \\
&= \Big[\prod_{t<T} \sigma_{\omega\restriction t}(a_t) \Big] \times \Big[\prod_{t<T} \prod_{\omega'\neq(\omega\restriction t)} \sum_{a\in A} \sigma_{\omega'}(a) \Big] \\
&= \Big[\prod_{t<T} \sigma_{\omega\restriction t}(a_t) \Big] \times \Big[\prod_{t<T} \prod_{\omega'\neq(\omega\restriction t)} 1 \Big] \\
&= \Big[\prod_{t<T} \sigma_{\omega\restriction t}(a_t) \Big] \times 1 \\
&= \prod_{t<T} \sigma_{\omega\restriction t}(a_t).
\end{aligned}$$

This gives us exactly Eq. (4.15), which completes the proof.

4.A.3 Information Theoretic Derivation of Softmax

The argument here uses the optimization method of Lagrange multipliers (see, e.g., Chapter 5 of Boyd and Vandenberghe 2004). What we want are the values of variables $x_a = \alpha(a)$. We can rewrite the cost as a function $\kappa(\vec{x})$ of the variables x_a in the following way:

$$\begin{aligned}
\kappa(\vec{x}) &= D_{KL}(\alpha \,||\, \delta) \\
&= \sum_a x_a \log\Big(\frac{x_a}{\delta(a)}\Big)
\end{aligned}$$

So we want to choose α to minimize the information theoretic cost $D_{KL}(\alpha \,||\, \delta)$, subject to the overall utility of the system, $\sigma(\vec{x}) = \sum_a x_a \,\mathbb{E}\mathrm{U}(a)$, being equal to some constant c. We also need to ensure that the sum $\sum_a x_a$ is equal to 1, and that each $x_a \geq 0$.

Writing everything in Lagrangian form, the following is what we want to optimize:

$$L(\vec{x}, \beta, \lambda) \quad = \quad \kappa(\vec{x}) - \beta\big(\sigma(\vec{x}) - c\big) + \lambda\big(\sum_a x_a - 1\big).$$

The parameter β will effectively trade off between the constraint of minimizing costs and maximizing utility, while λ will effectively be a component of a normalizing term.

We set the gradients all to 0, which gives a set of $k = |A| + 2$ equations in k unknowns. Solving this will give us our desired values $x_a = \alpha(a)$.

$$\nabla L \quad = \quad \begin{pmatrix} \vdots \\ \frac{\partial \kappa}{\partial x_a} - \beta \frac{\partial \sigma}{\partial x_a} + \lambda \\ \vdots \\ \sigma(\vec{x}) - c \\ \sum_a x_a - 1 \end{pmatrix} \quad = \quad \vec{0}$$

Note that $\frac{\partial \sigma}{\partial x_a} = \mathbb{E}\mathrm{U}(a)$, while $\frac{\partial \kappa}{\partial x_a}$ is equal to $\log\big(\frac{x_a}{\delta(a)} + 1\big)$. The equation for x_a becomes:

$$\log\big(\frac{x_a}{\delta(a)} + 1\big) \quad = \quad -\lambda + \beta\, \mathbb{E}\mathrm{U}(a)$$

Solving for x_a gives:

$$\begin{aligned} x_a &= \delta(a)\, e^{-\lambda + \beta\, \mathbb{E}\mathrm{U}(a)}. \\ &= \frac{1}{e^\lambda}\, \delta(a) e^{\beta\, \mathbb{E}\mathrm{U}(a)} \end{aligned}$$

Note that each x_a is therefore guaranteed to be non-negative.

Since λ is the same for each x_a, the term e^λ is essentially a normalizing constant (or a "partition function" as it is called in physics). The last equation, guaranteeing that the sum $\sum_a x_a$ is equal to 1, implies that

$$\begin{aligned} e^\lambda &= \sum_a \delta(a)\, e^{\beta\, \mathbb{E}\mathrm{U}(a)}, \text{ and thus} \\ \lambda &= \ln(\sum_a \delta(a)\, e^{\beta\, \mathbb{E}\mathrm{U}(a)}). \end{aligned}$$

To determine β we appeal to the remaining equation, which gives

$$c \quad = \quad \frac{1}{e^\lambda} \sum_a \delta(a)\, e^{\beta\, \mathbb{E}\mathrm{U}(a)}.$$

Solving for the only remaining unknown β gives probabilities whereby

$$\alpha(a) \quad \propto \quad \delta(a)\, e^{\beta\, \mathbb{E}\mathrm{U}(a)}.$$

If the default distribution δ was uniform, then $D_{KL}(\alpha \,||\, \delta)$ would be $\log(|A|) - H(\alpha)$, in which case the term $\delta(a)$ could be removed everywhere above. The derivative $\frac{\partial(\log(|A|) - H(\alpha))}{\partial x_a}$ becomes $\log(x_a) + 1$, so the probabilities simplify to:

$$\alpha(a) \quad \propto \quad e^{\beta\, \mathbb{E}\mathrm{U}(a)}.$$

In any case, the central point is that the parameter β and the expected level c of success are intimately related. If we want a higher value of c, we correspondingly have to increase β.

A variation on this derivation appeared in Jaynes (1957) to establish the softmax distribution as that with maximum entropy among those giving a specific mean value.

Chapter 5

Resource Rational Analysis

Researchers interested in descriptive questions about how minds work have numerous tools at their disposal. Behavioral experiments form the core of the cognitive scientific toolkit, but there are also other methods that reveal aspects of mental processing including eye-tracking and response-time studies. Some tools even allow probing the presumptive locus of cognition—the brain—for instance, with imaging techniques (fMRI, EEG, etc.), various types of lesions or external stimulation (TMS, optogenetics, etc.), and measurements of individual cells.

Even with all of these methodological tools available, much about the mind remains opaque, for at least two reasons. The first is simply that the tools are still typically inadequate for adjudicating between competing hypotheses, resulting in a fundamental *identifiability problem* (Pylyshyn, 1984; Anderson, 1990; Yamins and DiCarlo, 2016). A second source of epistemic opacity is the shear complexity of brain and behavior, a feature shared with many other scientific domains (Humphreys, 2009). Even if we had perfect observational and experimental access to brains, we would still face the formidable challenge of uncovering the fundamental logic of the mind—how all the moving parts fit together to produce behavior (Bechtel and Richardson, 2010; Jonas and Kording, 2017). Often it is not even clear what kinds of hypotheses would be sensible to test in the first place. This second challenge stems from the very human resource limitations that are the focus of this monograph (cf. the discussion in Wimsatt 2007).

A founding doctrine for the field of cognitive science is that—at an appropriate level of abstraction—the mind can be understood as solving various informational problems (Newell et al., 1959; Marr and Poggio, 1976; Marr, 1982). This perspective affords a distinctive methodological tool, a potentially powerful way of formulating psychological hypotheses. If we can clarify the problem some cognitive system is solving, then we can use solutions to that problem as guiding hypotheses for the particular way the mind manages to solve them. This is the methodology of *rational analysis*.

Traditional approaches to rational analysis (e.g., Anderson 1990) tended to downplay the role of resource constraints. A number of authors have suggested that resources should instead be front and center in rational analysis, resulting in a methodology of *resource rational analysis*. We have seen several projects in the spirit of this methodology from the previous chapters, e.g., optimal tradeoffs in neural processing (Example 2, §2.3.4), broad resource-related reasons why we might expect polarization in belief formation (§3.5), or randomness is behavioral responses to stimuli (Chapter 4). Some of those approaches (e.g., the latter two) are less concerned with detailed modeling of human cognition than with general conditions under which some type of behavior might be rationalized. Resource rational analysis is motivated by more specific questions about how the human mind works in particular. As an explicit research program it

has only been formulated relatively recently (Lewis et al., 2014; Icard, 2014; Griffiths et al., 2015; Gershman et al., 2015; Lieder and Griffiths, 2020). The present chapter offers a perspective on this research program, illustrated by a number of concrete examples from different arenas of cognition that showcase the wide variety of styles and angles that researchers have employed.[1]

5.1 Rational Analysis

The idea that thought and cognition can be understood in (instrumentally) rational terms is centuries old. Famously, for Aristotle rationality was characteristic of human beings, with much of mentality understood by appeal to its function.

The specific application of probabilistic and decision theoretic tools—of the sort introduced in Chapter 2—for understanding human thought began to appear in the 1950s, with signal detection theory in perception (Tanner and Swets, 1954), statistical approaches to learning and decision making (Luce and Suppes, 1965), and the hypothesis that people could be understood as "intuitive statisticians" (Brunswick, 1957). In a review article, Peterson and Beach (1967) summarized the state of the field by saying, "in general, the results indicate that probability theory and statistics can be used as the basis for psychological models that integrate and account for human performance in a wide range of inferential tasks" (p. 29).

By the 1970s, however, psychologists had amassed dozens of settings where human choice and behavior seemed dramatically at odds with decision-theoretic standards, including cases where people apparently thwart their own self-interests (Tversky and Kahneman, 1974; Slovic et al., 1977). Though some researchers argued that the many suboptimal heuristics and biases under study in this literature merely pinpoint the fault lines of an otherwise very well adapted cognitive system (see the discussion between Gigerenzer 1996 and Kahneman and Tversky 1996), others drew more pessimistic conclusions with "bleak implications for human rationality" (Nisbett and Borgida, 1975, p. 935). The so-called rationality wars have continued (Samuels et al., 2002), with vociferous debates about the precise fault lines of human thought and reasoning, about the general question of whether people are, on the whole, rational, and even whether such questions could be empirically meaningful (e.g., Cohen 1981).

Right at the height of the rationality wars, Anderson (1990) proposed a bold research program aimed at mitigating the identifiability problems in cognitive science, premised on a kind of rationality assumption. Rather than targeting a general notion of rationality applied at the "personal level" (in the sense of Dennett 1969), the target was rather specific cognitive systems. The methodology was laid out in six steps:

1. Precisely specify the goals of the cognitive system.
2. Develop a formal model of the environment to which the system is adapted.
3. Make the minimal assumptions about computational limitations.
4. Derive the optimal behavioral function given items (1)-(3).
5. Examine the empirical literature to see whether the predictions are confirmed.
6. If the predictions are off, iterate.

The methodology was illustrated with four examples of putative cognition systems, for memory retrieval, causal strength induction, general decision making, and categorization.

[1] The presentation of resource rational analysis in this chapter is an updated version of Icard (2014, 2018).

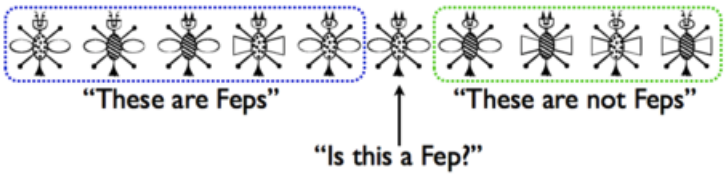


Figure 5.1: The problem of categorization. The image comes, with permission, from the e-book by Goodman and Tenenbaum (2008), and is based on Shafto et al. (2011).

Example 26 (Categorization)**.** Anderson (1990, 1991a) suggests that a fundamental reason we form categories is to make predictions, with an underlying assumption that objects tend to cluster together as a function of their attributes (perceptually salient features, labels, behavior, function, etc.), which makes prediction possible. Thus, suppose an agent has observed some sequence $\vec{o}$ of objects with different combinations of attributes. They then encounter a new object, but the value x of some specific attribute is unknown. The goal is to infer the value of this hidden attribute. See Figure 5.1, where the missing attribute is a linguistic label.

What makes this a categorization problem is that the distribution $p(x|\vec{o})$ is assumed to decompose via *clusterings* z of objects by their features:

$$p(x|\vec{o}) \quad = \quad \sum_{z} p(x|z)\, p(z|\vec{o}). \tag{5.1}$$

That is, the task is to estimate the probability of each clustering z given observations $\vec{o}$, and to determine the probability of an attribute value x given a clustering z. Step (1) of Anderson's methodology is thus to specify the goal of inferring the probability in (5.1).

In step (2), Anderson makes an assumption that these two terms—$p(x|z)$ and $p(z|\vec{o})$—reflect a certain type of environment. The probability $p(x|z)$ of a feature given a clustering takes the form of a so-called beta-distribution: as the number of observations increases, an attribute becomes more likely to the extent that other objects in the same cluster have that attribute. In this respect, objects in the same category tend to be more similar. Meanwhile, the probability $p(z|\vec{o})$ is given by Bayes rule, being proportional to $p(\vec{o}|z)\, p(z)$. The crucial prior $p(z)$ over clusterings implements a kind of "rich get richer" scheme, which turns out to be an instance of the so-called Dirichlet process mixture model (Sanborn et al., 2010). This captures the intuitive idea that objects tend to cluster together into categories.

The precise details of the model are not crucial. The upshot is that, by making these assumptions about the goal of categorization and about the distributional properties of the environment, Anderson was able to retrodict a wide array of empirical phenomena uncovered in the category learning literature, including order effects, prototype effects, the relative ease of learning different categories of Boolean concepts, and several others (Anderson, 1990, 1991a).

As the example illustrates, rational analysis as a methodological tool has a number of potential virtues:

1. **Hypothesis generation**: Proposals for how a cognitive system works are limited to those that solve the presumptive problem.

2. **Unification**: As in Example 26, many disparate empirical phenomena may be seen to arise from the same underlying principle, e.g., optimal prediction.

3. **Rationalization**: Insofar as the assumed task is intuitive, a successful rational analysis may "make sense" of otherwise puzzling behavior.

4. **Explanation**: Aside from the virtues of unification and rationalization, a successful rational analysis may also point the way toward an etiological explanation of the cognitive phenomenon. If a system optimizes some goal, this may well have been caused by a process of optimization, whether at a phylogenetic or ontogenetic scale (or both).

The formulation of cognitive problems as inference problems under uncertainty has been especially influential, with recent decades seeing a "Bayesian boom" (Hahn, 2014). People have been assumed to be "intuitive statisticians" not only in explicit prediction and estimation (Peterson and Beach, 1967; Griffiths and Tenenbaum, 2006), but also in more implicit ways through perception (Ma et al., 2022; Rescorla, 2015), intuitive physics (Battaglia et al., 2013), language understanding (Goodman and Frank, 2016), and many others domains (Oaksford and Chater, 1999; Chater et al., 2006). Here is another example, which goes beyond mere Bayesian inference to encompass the sequential decision making apparatus introduced in §2.2:

Example 27 (Optimal Feedback Control)**.** Imagine a simple motor planning task like reaching for a cup. If we think of A as the (discretized) "basic" movements an agent can undergo, for a given state $s \in S$ the utility $u(s, a) = v(s, a) - k(s, a)$ might be broken down into a difference between the task-related "intrinsic value" $v(s, a)$ of performing a in s—e.g., reflecting the degree to which the task has been successfully completed—and a "cost" term $k(s, a)$ measuring the energetic or metabolic resources required to carry out a in s. Note that this is an "external" cost rather than one that is internal to a deliberative mechanism. The aim is to maximize expected utility: in this case, to achieve the goal (grasping the cup) minimizing the costs.

One way a motor system might try to achieve this is to plan out a trajectory ahead of time. Then all the system needs to do is to execute that plan. In a stable and controllable environment, this may well be adequate. However, motor actions often take place in dynamic environment with many sources of noise, both internal (cf. §4.1.1) and external. What should the system do when forced off its planned trajectory? One possibility would be to plan a new trajectory. Another might be to try and return to the original trajectory.

A different perspective on this problem—first suggested by Todorov and Jordan (2002); see McNamee and Wolpert (2019) for a recent review—is to construe the predicament as one of *control*. Instead of planning out a specific trajectory, we should understand the system as implementing a *strategy* in the sense of Def. 2. As discussed in Chapter 2 (Remark 2), strategies are richer objects than what would be needed for a plan. If one were engineering a motor system, why might it make sense to implement a full strategy? And why might this be a helpful theoretical lens through which to understand the biological motor system? The intuition is that, in a highly noisy environment, embodying a more "reactive" policy may be be necessary. A well adapted agent will have suitable dispositions for a wide range of circumstances, possibly including those they do not expect to face.

Without speculating on how the strategy might be learned (§2.3) or implemented—and thus ignoring computational costs—it turns out that optimal strategies for various motor planning tasks predict empirical motor performance remarkably well. As a typical example, because of the cost term in the utility function, an optimal system will only respond to noise-induced deviations when necessary to achieve the goal. See Fig. 5.2. This is exactly what is found in numerous studies of motor control; see Scott (2012); McNamee and Wolpert (2019).

Despite being a remarkably prodigious research program, Bayesian rational analysis has met considerable resistance, on virtually all fronts. In many individual cases, researchers have questioned the norms associated with different cognitive functions. For instance, in contrast to Anderson's (1991a) analysis of categorization, Murphy (1993) points out that categories serve many other roles in cognition, e.g., in memory retrieval, language comprehension, planning, etc., which do not obviously reduce to attribute prediction.

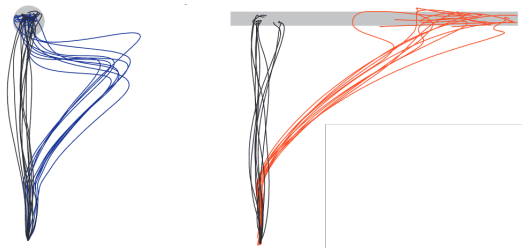


Figure 5.2: In a motor planning task, participants reach for a goal and are forced off path. When the departure frustrates task performance, reaching movements return to the original path (blues lines on left side). By contrast, when the departure is task-irrelevant, they continue with the perturbed path (red lines on the right). Reprinted with permission from Scott (2012).

The very idea that rational (in particular ideal-Bayesian) models would be explanatory has come under criticism. One source of the criticism stems from an expectation that psychological explanation should be *mechanistic* in nature (Jones and Love, 2011), while rational explanations are by design "nearly mechanism-free" (Anderson, 1990, p. 30). At best, therefore, rational analyses offer a preliminary step toward a genuine (mechanistic or etiological) explanation (Danks, 2008; Colombo and Seriès, 2012). A related concern is that the method is under-constrained, such that it is too easy to determine—*post hoc*—a Bayesian model to fit any data (Bowers and Davis, 2012). A third concern is simply that we appear to have a wealth of data showing that, as a matter of fact, behavior is seldom in accord with (Bayesian) rational principles. In other words, the very background assumption of optimality is generally untenable (Simon, 1991; Eberhardt and Danks, 2011; Mandelbaum, 2019).

There is a general reason that this last criticism has particular force. As many authors have stressed, the computations required to behave in a Bayes-optimal manner are almost always intractable (see van Rooij et al. 2019 for extended discussion of this point). Even if we grant that Bayesian inference provides the right normative standard for a given cognitive task, it is simply not possible for a resource-limited system to achieve that standard.

Example 28. Example 26 illustrates the point clearly. Even ignoring the Bayesian inference required to compute $p(z|\vec{o})$ for a specific clustering z, the expression in (5.1) involves a sum over all possible clusterings (i.e., partitions) of the observed objects. As Anderson recognized, the number of clusterings grows extremely quickly, giving the so-called Bell exponential numbers. This makes (5.1) wildly intractable, so much so that he was unable to compute the purportedly optimal posterior to derive its predictions.

This is where step (3) is invoked. Instead of maintaining a distribution over all possible clusterings, Anderson proposed that the mind commits at each step to a particular clustering, deterministically opting for the extension that has the highest posterior probability. In this sense, the approximation is "locally" optimal. Significantly, this version of the model was crucial in accounting for much of the data. For instance, order effects can result from "misleading" presentations of stimuli, while the idealized model in (5.1) would not predict any order effects (recall the discussion of Bayesian inference in §2.1.1, and of probabilistic automata in §3.5).

Subsequently, Sanborn et al. (2010) introduced an alternative, stochastic approximation to the distribution in (5.1), based on the so-called *particle filter*. This noisy alternative provided an even better quantitative fit to existing data, also accounting for individual-level variation. This raises the question of whether the partly randomized alternative could be in any sense better, or more tractable. It seems plausible that one of the strategies discussed in Chapter

4 might supply a positive answer to this question. For instance, a particle filter may be more efficient in memory usage (cf. Levy et al. 2009).

Of course, intractability is not just a feature of probabilistic calculations; it equally applies to other normative frameworks grounded in deduction, abduction, game theory, or indeed any framework that involves minimally complex logic or probability as a central component (recall Remark 9). The same concern applies, for instance, to the optimal control strategy in Example 27. Indeed, failures of optimality in that setting have been argued to be informative about the concrete mechanisms underlying motor control (see, e.g., de Rugy et al. 2012).

Anderson (1990) argued that the need to incorporate computational costs and architectural constraints could be the "Achilles' heel" of the rational approach (p. 32). In keeping with step (3), the methodology is said to be powerful to the extent that resource-related assumptions are minimal. Yet, as Example 28 illustrates, assumptions about resource constraints may well turn out to be crucial. *Resource rational analysis* is animated by the contention that rendering resource costs explicit will not only ameliorate some of the weak points of the methodology, it also has the potential to enhance the promised virtues listed above in 1-4.

5.2 Toward Procedural Rationality

A key motivation for rational analysis is to mitigate the identifiability problems.[2] The space of possible cognitive models will almost always be under-determined by the observational and experimental data we have available. If we can constrain this space further by hypothesizing that the system effectively solves a well-motivated problem, this can help to identify promising candidates that can be further tested and investigated.

In classical rational analysis the problem to be solved is typically understood in substantive terms (cf. Fn. 5): given some observations, the agent needs to take the right action. For instance, the action might be a guess about the hidden state that produced the observations. Given a prior on states and a likelihood describing how states produce observations, the optimal action is to guess a state with maximal posterior probability. Anderson's analysis of categorization (Example 26) is a typical instance. More generally, we might say that classical rational analysis is concerned with strategies (or "policies") in the sense of Def. 6, that is, functions from (sequences of) observations to actions. The example of optimal control (Example 27) would be another paradigmatic instance of classical rational analysis.

By contrast, resource rational analysis characterizes problems in *procedural* terms (§1.2.2). Crucially, this involves potential costs involved in "processing" observations and any calculations required to generate a response. One way of making this idea concrete is to focus not on abstract strategies, but on *programs* π that transform observations into actions—and thus, in effect, implement strategies—while incurring a cost in doing so. We have already seen this idea at play in Chapters 3 and 4. Generalizing, a given history[3] will be associated with some cost to the program of generating this history. Analogous to the determination of expected utility (2.18), we have an *expected* cost $C(\pi)$ of using program π. Abusing notation by letting $\mathbf{U}(\pi)$ be the expected utility of the strategy π that effects (2.18), we arrive at a relatively general version of the fundamental equation (1.1):

$$\mathbf{V}(\pi) \quad = \quad \mathbf{U}(\pi) - C(\pi). \tag{5.2}$$

[2] As Crupi and Calzavarini (2023) emphasize, it is useful to distinguish at least two identifiability problems: aiding in the search for promising models on the one hand, and validating models on the other.

[3] Recall a *history* is a sequence of interactions the agent has with the environment. See §2.4.

For the purpose of resource rational analysis, it is **V** defined on Π—rather than **U** defined on all strategies—that captures the relevant problem to be solved by a given cognitive system. Significantly, the best programs in Π for problem **V** might have rather different characteristics from the best (unconstrained) strategies for **U**.

Resource rational analysis requires strictly more from the researcher. Not only do we need to specify the problem to be solved in broad inferential terms (essentially **U**), we also need to specify a *space* Π *of possible programs*, possibly also with a cost function C on Π. A central concern of resource rational analysis is how to determine Π and C. Recall from §1.3 that there are two broad approaches to this problem, the cost-theoretic and the panoramic. Methodology also varies in the degree of detail they assume at the outset of investigation. At one end of the spectrum among panoramic approaches, the space of programs can be highly constrained by antecedent understanding of the architecture in question. For instance, Π might be restricted to a specific set of mechanisms that all resemble a cascade of neural populations (Example 2, §5.3.2, etc.), or it might be built from a specific set of anatomically grounded operations with alternative programs differing only in the values of numerical parameters (§5.3.1).

At the other end of the spectrum, some cost-theoretic approaches have explored imposing abstract costs on *any possible implementation*, as in Chapter 3 (Def. 11) and Chapter 4 (Defs. 14, 15). For instance, abstract information-theoretic costs can be combined with minimal assumptions about the mechanism mediating between perceptual input and behavioral output (§5.3.3). Slightly less abstract, researchers have ventured proposals for the algorithmic logic underlying cognitive processes, often inspired by computational work in statistics or machine learning (§5.3.4; Griffiths et al. 2015; Gershman et al. 2015; Lieder and Griffiths 2020).

To summarize, a resource rational analysis involves postulation of (up to) three components:

1. **Goals**: a utility function **U** on strategies, characterizing the abstract decision problem involved (Def. 5). This typically exhausts classical approaches to rational analysis.

2. **Programs**: a potential restriction Π of the space of all strategies for problem **U**. In the panoramic approach, Π encompasses all putatively *feasible* strategies; in the cost-theoretic approach programs are measured by:

3. **Costs**: a measure C of resources required to execute different computations, which can be traded off against utility as in Eq. (5.2). Specifically, costs are understood at a suitable level of generality as *opportunity costs*: the value forgone when employing a resource for one purpose rather than any other. There is no demand to stipulate a separate cost function in the panoramic approach. (Recall the discussion in §1.4.)

The result is a relatively conservative revision of Anderson's six step methodology: in step 3 we do not necessarily make the minimal assumptions about computational limitations. Instead, such considerations take center stage and constitute a core component of what is under investigation (cf. Lewis et al. 2014, p. 282).

When it comes to the role of (resource) rational analysis in hypothesize generation (virtue 1 above), the more we can credibly constrain the problem, the more progress we may hope to make on the identifiability problem (cf. Anderson 1990, pp. 29-32). Further constraints on the assumed task—a more demanding decision problem, a more restricted space of programs, or a more discerning cost function—will typically result in fewer good solutions. With fewer good solutions, it is more likely that any particular good solution we identify will capture important aspects of the mental process under investigation. Cao and Yamins (2021) dub this familiar pattern the *contravariance principle*, as the size of the solution space is "contravariant" in (or "anti-monotonically related" to) the difficulty of the problem.

Of course, the principle is applicable only insofar as two requirements are met. First, and most obviously, we should believe that the constraints accurately characterize the cognitive problem in question. Any architectural assumptions built into Π or C should be plausible. This may be established through background knowledge or independent investigation, though in some cases successful (resource) rational analyses have themselves been taken as confirmation that the characterization of the problem was an appropriate one (see, e.g., §5.3.4).

Second, we need to be able to identify at least one good solution (viz., cost-efficient program) to the constrained optimization problem so characterized. That is, we need to be able to generate at least one good hypothesis to test experimentally (step 4 in the methodology). In many cognitively relevant contexts, finding good solutions is a major research challenge in its own right (cf. Zednik and Jäkel 2016). This is certainly true for problems of probabilistic inference (Murphy, 2012), not to mention the problem of efficiently solving POMDPs (§2.4).

The prospects for a resource rational analysis to address the methodological problem of identifiability thus depend in part on navigating these subtle issues. Concerning the other putative *explanatory* virtues, 2-4, we return to them below in §5.4 after first considering four illustrations of resource rational analysis. Together, these four illustrations reveal a wide variety of very different approaches to the general resource rational methodology.

5.3 Illustrations of Resource Rational Analysis

There is a growing body of work that is explicitly labeled as *resource rational analysis*, including significant strands on planning and decision making (Callaway et al., 2022; Ho et al., 2022), probabilistic inference (Vul et al., 2014; Lieder et al., 2017; Dasgupta et al., 2017), causal reasoning (Icard and Goodman, 2015; Bramley et al., 2017), language processing (Gibson et al., 2019; Hahn et al., 2022), moral reasoning (Levine et al., 2023), and more (Nobandegani, 2017; Lieder and Griffiths, 2020). However, on the present understanding of resource rational analysis, much else falls within its purview. Most obviously, it includes work invoking the nearly synonymous concept of *computational rationality* (Lewis et al., 2014; Gershman et al., 2015). It also includes work on efficient coding in perception (e.g., Sims 2016), neural reuse and cognitive control (Shenhav et al., 2017), resource-optimality analyses of neural wiring and organization (e.g., Bullmore and Sporns 2012; Cherniak 2012), and recent research on "goal-driven deep learning models" to understand neural computation (Yamins and DiCarlo, 2016; Schrimpf et al., 2020). While different research strands emphasize different ambitions and dimensions of the methodology, they all adhere to the basic tenets presented in the previous section. Described here are four rather different incarnations of the methodology, moving from very low-level to relatively high-level cognitive phenomena. The first two (§5.3.1, §5.3.2) are examples of course-grained approaches, while the next two (§5.3.3, §5.3.4) are examples of cost-theoretic approaches.

5.3.1 Oculomotor Control in Reading

Lewis et al. (2013) study low-level saccadic control in a sequential reading task, focusing especially on the tradeoff between speed and accuracy (see also Lewis et al. 2014). This is an example of a panoramic approach in which much is already known about the detailed costs and constraints faced by the visual and motor systems. Such details are incorporated by restricting to a highly limited class of possible programs.

Experimental participants are presented with sequences of six four-letter strings, and the task is to determine whether any of the strings is *not* a word. For instance, they might see:

aunt swap hack leil step find

In this case, the goal would be to report that there is a non-word in the list, namely leil. American experimental participants are assumed to scan the list from left to right. There are three possible "actions" at each step:

(i) move eyes to next word (ii) push the "yes" button (iii) push the "no" button

While fixated on a particular string, noisy evidence accumulates about whether the string is a word. An unconstrained program for this task—that is, assuming no costs or resource limitations—would move on to the next word instantaneously, as soon as sufficient perceptual evidence is gathered, and then terminate the trial as soon as a conclusion is guaranteed.

The human oculomotor system, by contrast, works under strict architectural constraints, and much is already known about its various costs and resource bounds. Signals from the brain take a certain amount of time to reach the eye, and there is also delay incurred in saccadic and motor programming. Empirically supported hypotheses about these further parameters can be incorporated into the specification of a program, so that there remain only two free parameters. First, we need to specify the threshold θ_s that determines when a saccade to the next string will be initiated. Second, there is a decision threshold θ_d that is reached either when a non-word is detected, or it is determined that all the strings are words. Thus, a strategy essentially amounts to just a pair of numerical thresholds, (θ_s, θ_d), and the space Π of possible programs can be identified with the set of all such pairs.

A researcher might try to determine these two parameters empirically, e.g., fitting them from data. Instead, inspired by work in computer science on bounded optimality (Russell and Subramanian, 1995)—recall Example 1—Lewis et al. (2013) derive the *optimal* values of θ_s and θ_d given the task and assumptions about architectural constraints. The behavior of this resource-optimal program can then be compared to human performance on the same task. They do this across three different conditions characterized by different speed-accuracy tradeoffs: participants (and the model program) are rewarded for speed and penalized for an incorrect response, and these reward and penalties come in different proportions. In other words, they study three different problem settings—$\mathbf{U}_{\text{speed}}$, $\mathbf{U}_{\text{acc}}$ and $\mathbf{U}_{\text{medium}}$—and test whether people's behavior matches that of the optimal program—i.e., the pair (θ_s, θ_d)—for that problem.

Lewis et al. (2013) find that the resource-optimal program is both qualitatively and quantitatively similar to human behavior across all three conditions. There are at least three particularly noteworthy features of the analysis:

1. Saccadic control appears to be appropriately modulated by the task, viz. the specific speed-accuracy tradeoff ($\mathbf{U}_{\text{speed}}$, $\mathbf{U}_{\text{acc}}$ or $\mathbf{U}_{\text{medium}}$). In other words, the time spent fixated on each word is shown to be sensitive to the costs of inaccuracy vs. the reward for speed.

2. Both human performance and the optimal program are appropriately attuned to the "ecology" of words. As processing becomes more accurate with more familiar words, less time should be spent on them. This is exactly what they find. (Word familiarity in these experiements is approximated by frequencies in natural corpora, and is incorporated into the sequential decision problem formulation.)

3. Crucially, the architectural bounds are necessary to match human performance. The authors compare the resource-optimal program to various unbounded (or less bounded) versions of it, in effect expanding the class Π to include programs that do not respect the cost constraints imposed by the oculomotor system. Such ("unboundedly optimal") programs show strictly worse fit to the human data.

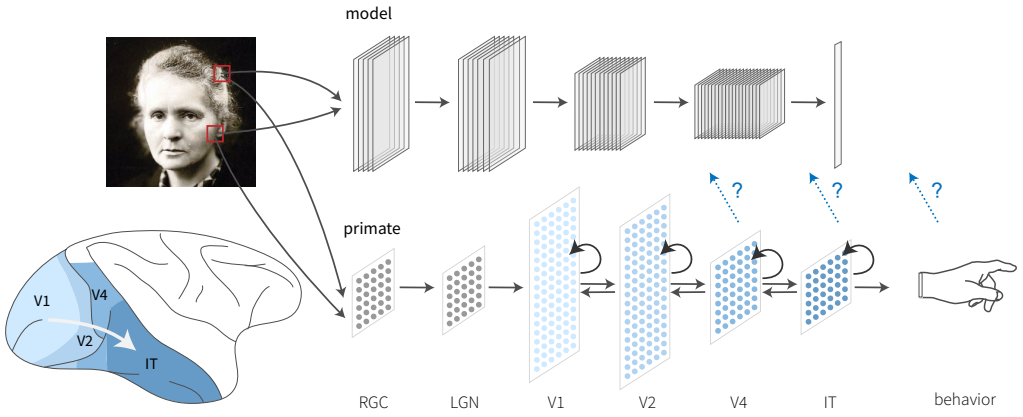


Figure 5.3: The problem of image classification in deep convolutionary neural networks (top) and in primate cortext (bottom). The figure comes from Schrimpf et al. (2020), reproduced with permission, and is based on Yamins and DiCarlo (2016).

Thus, given the various intrinsic costs and constraints faced by the perceptual and motor systems, human saccadic movements appear to be optimized along the remaining degrees of freedom for solving reading (and presumably many other oculomotor) tasks. That is, they seem to be optimal given the existing resource constraints confronting the system.

5.3.2 Modeling the Ventral Stream

The primate ventral stream is composed of a sequence of cortical areas, widely believed to transform visual stimuli into successive representations that ultimately (by area IT) facilitate robust object recognition. A fundamental challenge is to understand the nature of these transformations. Yamins et al. (2014) introduced the possibility of confronting this challenge by exploring *deep convolutional neural networks* (DCNNs) that were trained to perform well at purportedly difficult visual classification tasks (see also Yamins and DiCarlo 2016). Convolutional neural networks were originally inspired by the biology of the mammalian visual system (see Lindsay 2021), and Yamins et al. (2014) restricted attention to a class of DCNNs whose organization bears resemblance to the hierarchy of layers in the ventral stream (Figure 5.3). The hypothesis was that DCNNs which perform relatively well at the classification task might, at some appropriate level of abstraction, be good *models* of the ventral stream.

Like in the previous case (§5.3.1) this is a panoramic approach: the space Π is restricted to a specific class of "biologically plausible" programs, viz. artificial neural networks, with no further need to specify additional costs. The programs differ only in their settings of 57 parameters, including some for architectural properties such as the number of hidden layers. The task $\mathbf{U}$, meanwhile, is given by 8-way classification of 5,760 images: correctly classifying more images means better performance on the task.

A notable feature of this case study is that the *optimal* program in Π for $\mathbf{U}$ cannot always be identified. Instead, high-performing programs are identified by through training, using gradient descent algorithms like backpropagation. Thus, the part of Π to be explored is not stipulated "by hand" but rather determined indirectly by a training process in much the way that modern machine learning models are devised in engineering contexts.

A second notable feature is that, unlike in the previous case (§5.3.1), the goal is not (just)

to predict choice behavior, but to predict neural response patterns in the brain. The express goal is to reveal the mechanism by which the ventral stream transforms perceptual input into a classification decision, in effect promising a "triumphant cascade" (Dennett, 1987) from higher to lower levels of analysis. Much of the methodological work thus focuses on the question of how to assess fit between a (trained) DCNN and neural measurements.

Yamins et al. (2014) showed that neural activity in successive areas of the macaque cortex could be predicted (up to a linear transformation) by the activity of artificial neurons in a high-performing DCNN. Strikingly, such models provide the best quantitative fit to neural activity in multiple areas of visual cortex (V4, IT) of any existing model, better even than models that were specifically trained to predict such activity (Yamins and DiCarlo, 2016). Furthermore, Yamins et al. (2014), and later Schrimpf et al. (2020), identified a strong correlation between performance of a program π on the task and the degree to which π predicts neural responses. Crucially, this correlation is only found among models that have the appropriate structure. Just as in the study from Lewis et al. (2013), "ideal observer" models, which perform the task perfectly but are free from the relevant architectural constraints, provide poorer fit, showing again that both the task ($\mathbf{U}$) and the constraints (the restriction to strategies in Π) are essential.

5.3.3 Policy Compression

Moving from perception to action, consider any ordinary sequential decision problem such as grocery shopping. We might imagine that for a given individual different outcomes of the shopping process might be more or less desirable, perhaps with the "optimal" outcome one that best balances quality of goods with their monetary costs. At the same time, most people are not quite so careful when moving down the shopping aisles. Instead of stopping to inspect each individual brand of crackers—reading nutritional information, considering the price, and so on—people will often simply pick up the first acceptable one they see (Simon, 1955).

In such scenarios people are essentially ignoring differences between states: seeing brand A at the top of the shelf is no different from seeing B. Even though there may be some perceptible differences between them, people pay no attention to them and the action chosen is the same either way. The strategy is simply to take whatever box appears at the top of the shelf.

The same example illustrates another intuitive source of strategy simplification. After visiting the same grocery store many times, traversal through it may become increasingly automated. On a first visit, one might proceed according to a list, deliberately searching for each successive item. But eventually large sequences can be "chunked" so that little thought is necessary when moving from item to item. To the extent that the store's arrangement stays fixed, the chunked (or "compressed") strategy achieves the same outcome, with seemingly less required on the part of the shopper. See §6.2 for more on this type of abstraction.

Lai and Gershman (2021) have argued that these and other related phenomena can be subsumed under a general resource rational principle of *policy compression*. Suppose an agent's predicament (like the shopping example) can be captured by a Markov decision process (MDP; Def. 3). There is a space S of possible states, a space A of possible actions, and if action a is taken in state s, then utility $u(s, a)$ is obtained and there is a probability $q_{s,a}(s')$ of moving from s to new state s', At that point another action is taken. A strategy (or "policy") is a map from states to actions, and the best strategy σ is one that maximizes $\mathbf{U}(\sigma)$ (§2.9).

In order to effect a strategy any physically instantiated agent will need to *encode* states in some way, to provide some physical intermediary between the "percept" s and the generated action a. Recall Fig. 4.4, in which the agent program first applies an encoding e to the state s to obtain a code $c = e(s)$, and then uses this code to generate an action a with some probability $\sigma_c(a)$. On this picture, σ now depends not directly on the state, but on some

(possibly coarsened, or even distorted) representation of it. Empirically, there does seem to be a significant compression of information moving from the cortex—where state information is evidently discerned—to striatum—where action selection is assumed to originate (Lai and Gershman, 2021). The question now is how we should understand the *cost* associated with different degrees of compression, and more abstractly with enacting a given agent program π.

Following the analysis presented in §4.3.2, we might take the cost of implementing program π to be the *mutual information* $I(\mathsf{A}, \mathsf{S})$ between the "default" distribution on actions, $P_\pi(\mathsf{A})$, and the stationary distribution on states, $P_\pi(\mathsf{S})$, or equivalently, by the average KL-divergence from the default action distribution to the situation-specific distribution π_s (Def. 15). As before, the goal is then to maximizing cumulative reward $\mathbf{U}(\pi)$ while minimizing (some factor of) information-theoretic cost:

$$\mathbf{V}(\pi) \quad = \quad \mathbf{U}(\pi) - \frac{1}{\beta} I(\mathsf{A}; \mathsf{S}) \tag{5.3}$$

The parameter β determines how "coding" costs trade off against utility: the higher β is, the less information costs, and the closer the program should be toward maximizing utility. Recall that, in principle, β should reflect the opportunity cost of using cognitive resources that could be put to some other purpose.

The optimal program π^* is one that maximizes $\mathbf{V}(\pi^*)$, which we noted (see Thm 8) has:

$$\pi^*_s(a) \quad \propto \quad P_{\pi^*}(\mathsf{A} = a)\, e^{\beta Q^{\pi^*}(s,a)}. \tag{5.4}$$

Adopting Eq. (5.4) as a working hypothesis for how people engage in sequential decision problems implies several distinctive qualitative patterns. Perhaps most obviously, the optimal program is in general *stochastic*. As long as there are at least two actions with positive utility, they will both be taken with some positive probability. As discussed in Chapter 4, this might be offered as a kind of justification for randomized behavior. Another notable aspect of Eq. (5.4), highlighted by Gershman (2020), is that it implies a kind of *perseveration*: because of the term $P_{\pi^*}(\mathsf{A} = a)$, actions a that are more often chosen are in turn more likely to be chosen going forward, even when it appears suboptimal. Such a pattern is one the oldest observed in the study of human and non-human animal behavior (Thorndike, 1911); here it emerges as part of an optimal solution to the assumed problem of resource rationality.

When it comes to more quantitative comparisons of the optimal program to human behavior, it turns out that solving the system of equations specified by (5.4) is itself a challenging problem. The perseverance term introduces a difficult type of circularity, since the probability of taking a depends on the *stationary* probability of taking a. Lai and Gershman (2021) introduce a learning algorithm for approximating Eq. (5.4), which allows moving to step 5 of the (resource) rational analysis methodology, that is, comparison to existing data. The algorithmic approximation is itself inspired by existing models of the basal ganglia, in effect taking a further architectural constraint into account (see also Gershman and Lai 2021).

The resulting algorithmic proposal was shown to account for a number of subtle quantitative patterns in response times, perseveration, action chunking, and state chunking (Lai and Gershman, 2021). One-shot decisions with large state spaces (e.g., as in Collins and Frank 2012) highlight the need for state chunking, while sequential problems with structured transitions call for action chunking (Lai et al., 2022). Analysis of these tasks showed that people compress for both reasons, as predicted by the model, and that increase in cognitive load leads to yet further compression (a pattern also found in schizophrenic participants; see Gershman and Lai 2021). In cases of action chunking, reaction time decreases, presumably because the successive states do not even need to be encoded. In cases of state chunking, responses appear to be more stochastic, a pattern predicted by decrease in the β parameter.

Thus, in addition to predicting a number of key patterns in choice behavior, the policy compression model also potentially helps *make sense* of much of this behavior, insofar as it involves an intuitive trade off between maximizing reward and minimizing memory costs.

5.3.4 Anchoring-and-Adjustment

One of the classic "heuristics" introduced by Tversky and Kahneman (1974) is the *anchoring-and-adjustment* heuristic, whereby point estimates are inappropriately "anchored" to some irrelevant salient value. For instance, people's estimates for the percentage of African countries in the UN are influenced by the number that comes on a roulette wheel (e.g., whether it comes up 65% or 10%). Strikingly, this pattern reveals itself in judicial sentencing decisions, even when the judges know the anchor was generated randomly (Englich et al., 2006). While there has been much debate about the underlying psychological mechanism, evidence that it is influenced by financial incentives, cognitive load, time pressure, and alcohol consumption suggests that it may have a basis in resource limitations (Epley and Gilovich, 2006).

Lieder et al. (2012, 2017) explore the idea that anchoring may result from a resource rational approximation to probabilistic inference. At a high level the idea is that point estimation can be understood as a problem of inference under uncertainty. As a simplifying idealization it is assumed that what a person knows about a topic can be captured by a probability distribution. For example, each individual maintains a distribution $p(x)$ over possible percentages x of African countries in the UN. However, much of this knowledge is implicit and it may take cognitive work to extract the consequences of it, in particular to identify the value x with highest probability. To turn this into a resource rational analysis we need to say something concrete about what this process of extraction might be, and what costs are associated with it.

It was suggested in Vul et al. (2014)—see also Icard (2016); Sanborn and Chater (2016)—that a fundamental cognitive mechanism for probabilistic calculation is randomized *sampling*. As discussed further in §4.1.3 above, the suggestion is that the mind has an ability of draw a "sample" value roughly in proportion to its probability. The more samples drawn, the more the distribution can be accurately recovered (recall the "Galton box" in Fig. 4.1). Behavioral evidence and neural modeling work are consistent with the idea that the brain encodes some probabilistic information and performs probabilistic calculations by means of something like sampling approximations (e.g., Orbán et al. 2016).

How exactly are samples drawn from an implicit probability distribution? Lieder et al. (2017) notice that one particular sampling algorithm—a "Markov chain Monte Carlo" algorithm known as the *Metropolis method* after Metropolis et al. (1953)—bears a prima facie resemblance to anchoring-and-adjustment. The method works as follows:

Definition 16 (The Metropolis Algorithm)**.** For some number $T \geq 0$ of times steps:

1. Start at time $t = 0$ with an initial value $\widehat{x}_0$, which we might call the *anchor*;
2. At time steps $t+1 \leq T$ suppose the current value is $\widehat{x}_t$. Generate a random perturbation δ, e.g., from a normal distribution with mean 0, and then let:

$$\widehat{x}_{t+1} \quad = \quad \begin{cases} \widehat{x}_t + \delta & \text{if } p(\widehat{x}_t + \delta) > p(\widehat{x}) \\ \widehat{x}_t & \text{otherwise.} \end{cases} \tag{5.5}$$

   It is assumed that making the binary comparison $p(\widehat{x}_t + \delta) > p(\widehat{x})$ is relatively easy.

The output of (i.e., the *sample* from) the algorithm is $\widehat{x}_T$.

Fact 2. As $T \to \infty$, the distribution of samples $\widehat{x}_T$ converges to the true distribution p.

However, for small values of T the method produces *biased* samples, specifically biased toward the starting anchor. The connection to anchoring-and-adjustment is thus apparent. The suggestion is that for point estimation problems people employ a program that works as follows:

1. Identify some initial value, either from a salient plausible example, or perhaps by self-generation (Epley and Gilovich, 2006).

2. For some number of steps, randomly adjust it in the direction of greater plausibility.

For instance, in the original example from Tversky and Kahneman (1974), a participant might take 10%—the number on the roulette wheel—as an anchor and then consider a short sequence of random perturbations that move the number up.

Formalizing this procedure as an instance of the Metropolis algorithm gives our set Π of programs. Fixing a given anchor, the programs in Π differ only in the number of adjustment steps taken, with more steps promising a more accurate guess but incurring a greater cost. We thus arrive at the present version of Eq. (1.2), with π_T the program that runs for T steps:[4]

$$\mathbf{V}(\pi_T) \quad = \quad \mathbf{U}(\pi_T) - \frac{1}{\beta}\, T. \tag{5.6}$$

As β increases it becomes more advantageous to run the algorithm for longer.

Lieder et al. (2017) show that this simple proposal accounts for a large number of observed patterns from the literature on anchoring. At a general level it helps make sense of the influence of factors like financial incentives, cognitive load, and time pressure, insofar as these all impact the tradeoff parameter β, which itself concerns the overall *opportunity costs* involved (§1.4). Other qualitative patterns—such as the increase in bias with more extreme anchors, or the decrease in bias when participants are more knowledgeable about the domain—also easily fall out of the proposal. By fitting the various parameters of the model (e.g., the tradeoff parameter β, the prior p, etc.), Lieder et al. (2017) are also able to provide a close quantitative match to a range of studies on these and other effects from the literature. Subsequent work has expanded this type of account from the uni-dimensional settings typical of point estimation problems to multi-dimensional, combinatorially rich domains (Dasgupta et al., 2017).

5.4 Some Methodological Points

The most formidable challenge to any methodological approach founded on optimality analyses is the risk of triviality. For any behavior whatsoever there will be some analysis on which it is optimal. The approach is effective to the extent that the underlying problem (i.e., $\mathbf{U}$), and the characterization of programs (Π) and their potential costs, are all well motivated and independently plausible. When they are, and when experimental behavior is well predicted by the analysis, this can lend even further credence to those very assumptions that went into the analysis. Not only were they antecedently plausible, but they in concert were sufficient to derive empirical phenomena. As elsewhere, the more specific these predictions are—the less they appear to be derivable from other possible accounts—the stronger this inference will be.

The four examples presented above all seem to enjoy this status. Importantly, in all four cases, both the underlying problem and the incorporation of resource constraints appear to be crucial for deriving the full suite of empirical predictions. As the examples also demonstrate, there is a remarkably wide variety of settings in which resource rational analysis can be applied, spanning numerous domains of perception and cognition. Moreover, these are merely four

[4]The utility is naturally taken to be expected accuracy: $\mathbf{U}(\pi) = -\mathbb{E}_{\hat{x}\sim\pi}\mathbb{E}_{x\sim p}|\hat{x} - x|$. See Lieder et al. (2017).

representative examples of a now large body of work employing similar strategies. It is worth highlighting several points about the general methodology.

Resource rational analysis need not be Bayesian. Chapter 2 presented a normative framework for assessing agents based on Bayesian decision theory. Although we acknowledged possible variants of this approach (§2.1.6), it remains the presumptive framework for understanding what "things going well" might mean. In as far as we are picking up on structure that has, one way or another, sculpted how our minds work, it would not be surprising if some agents in some circumstance solve those fundamental problems by employing some of the very same apparatus that we use from our external perspective, viz. representation of probabilities, desirabilities, costs, and so on (cf. Chapter 6). With that said, there is no presumption that a good (externally characterized) solution π will engage in anything that we would recognize as explicitly "performing Bayesian inference" or "maximizing expected utility," for example.

Resource rational analysis typically cuts across Marr's levels of analysis. Marr's (1982) "computational" level of analysis can roughly be equated with what we have been calling the underlying problem $\mathbf{U}$, or (following Anderson 1990, 1991b) the "goals" of the system. Conceptually, it remains useful to consider problems at this level. After all, abstract inference and decision problems can be easier to understand and analyze than intricately constrained optimization problems. In order to separate $\mathbf{U}$ from Π (and C) in the first place, we need some separate understanding of $\mathbf{U}$. Nonetheless, in contrast to traditional rational analysis, resource rational analysis emphasizes and celebrates "downward glances" (Churchland, 1986) toward the algorithmic and even implementational details that ultimately constrain possible solutions to those problems. In this way, resource rational methodology often invokes operations over concrete representations (cf. §6.1), typically associated with Marr's second (algorithmic) level.

Some authors propose a "Marr level 1.5" in between the computational and algorithmic levels, to capture this stance (Griffiths et al. 2015; see also Peacocke 1986), while others highlight the multiple dimensions at play in talk of levels (e.g., Danks 2008). In any case, distinctive of the methodology is its promiscuity with respect to the traditional levels of analysis.

A successful analysis can help make sense of otherwise puzzling behavior. In part due to the separation of $\mathbf{U}$ on the one hand, and the constraints agents face in dealing with $\mathbf{U}$ on the other, behavior that may seem arbitrary or "kluge-like" can instead emerge as *reasonable* in some meaningful sense (virtue 3 above). Lieder et al.'s (2017) treatment of anchoring-and-adjustment phenomena is a potential example of this. At first glance, anchoring one's estimate toward a number known to be randomly generated appears patently irrational. It may not reflect behavior that we would want to see in a societally consequential role (such as a court judge) either. Nonetheless, it is enlightening to consider the possibility that this behavior arose as part of a good (possibly even resource optimal) solution to a very general and very challenging problem, in this case, estimation from diverse and scattered evidence. In cases where we want to "correct" this behavior—say, because we care more about accuracy than efficiency, compared to the weighting implicit in whatever development lead to the behavior—it may be helpful to pinpoint its potential (resource) rational source.

Resource rational analysis is not guaranteed to succeed. Anderson (1991b) characterized rational analysis as a "high risk, high gain" enterprise (p. 472). It is high risk because the postulation of goals could be entirely wrong. We could simply have misunderstood the fundamental task that some cognitive function is solving. Evidently the primary way we have of verifying this is via poor fit to empirical data. If it turns out to be consistent with a wide range of experimental behavior, then it can at least serves as a *summary* of the behavior in question, whether or not it gives any insight into the etiology of that behavior. It is usually infeasible

to trace the development of a cognitive trait, especially in evolutionary time. In this respect, hypotheses generated by rational analysis are falsifiable, though in practice only weakly so.

Hypotheses generated by resource rational analysis are even higher risk, but also more strongly falsifiable, at least in principle. In addition to the postulation of goals, the researcher attempts to capture genuine physical constraints at some appropriate level of abstraction. These may be antecedently understood (as in the first two examples, §5.3.1, §5.3.2). But in many cases they are merely hypothesized, potentially even as a central target of the analysis. In such cases there are simply more ways to be wrong. Here, mischaracterization can be demonstrated not only through behavioral experiment, but also through subsequent analysis of the relevant substrate, the brain. Causal and interventional analysis of neural systems could be used to falsify architectural claims, and claims about availability or use of a given resource. While such methods remain largely elusive for biological brains, there has been considerable recent progress in causal analysis of artificial neural systems (see, e.g., Geiger et al. 2021, 2023).

Celebrated "explanatory virtues" are on display in the methodology. A chief methodological role for resource rational analysis is to facilitate the search for promising hypotheses. Yet this clearly does not exhaust its scientific interest. Even when the space Π of natural variation is already known—that is, even if we could narrow down our space of hypotheses to precisely those programs that some (possible) agent in our class of interest embodies (an achievement rarely, if ever, attained)—we may still wish to know *why* it has the character that it does. The methodology promises progress on this question as well.

Like more traditional rational analysis, there is a promise of *unification* insofar as many disparate phenomena can be subsumed under some of the same general patterns (Colombo and Hartmann, 2017). This is evident, for instance, in Lai and Gershman's (2021) analysis of policy compression. Their entropic cost term leads to a unified analysis of persevertaion, randomness, action chunking, state chunking, and several other phenomena. Much more broadly, construing many different aspects of cognition instrumentally as solutions to sequential decision problems brings in tow a set of tools and concepts—including from engineering—that help pinpoint deeper structural connections between phenomena that might have otherwise seemed unrelated.

Finally, again in the spirit of traditional rational analysis, there is a hope that a successful resource rational characterization of a phenomenon will at least point the way toward some causal or etiological account of how the cognitive function came to be, whether through learning, evolution, or some combination. As Anderson (1990) expressed the suggestion, "My own sense is that cognition is likely to be one of the aspects of the human species that is most completely optimized and optimized in a clean, simple way so that it will yield to scientific analysis" (p. 29). Furthermore, explanations based on optimality principles may possess their own distinct explanatory virtues, viz. causal generality (see, e.g., Sober 1983). To the extent that cognition does yield to optimality-style explanations and analyses—specifically through principles of resource rationality as presented in this Element—that tempers the perspectival nature of resource rational pronouncements as introduced in Chapter 1. When employed for this purpose, such pronouncements should pinpoint those features of thought that have in fact been subject to a process of optimization, not as a "spandrel" of an unrelated process.

In sum, we should expect resource rational analysis to be an important tool in the larger toolbox of cognitive science. Ultimately, this style of analysis promises rather deep insights about what type of agents we are, insofar as it succeeds in highlighting the many ways our minds are exquisitely tuned to performing tasks under formidable resource constraints. As Nozick (1993) put it, "our view of the world and of ourselves, and our notion of what counts as rational, are in continual interplay" (p. 135).

Chapter 6

Creature Construction

Much of what we know about our own minds comes from careful scientific experimentation and analysis. Often in cognitive science we treat the mind as a kind of "black box," venturing hypotheses about its inner workings that we hope will eventually merge with a bottom-up mechanistic understanding of neurobiology. The four examples of resource rational analysis discussed in the previous chapter all have some of this character.

At the same time, we also know our own minds from a more direct type of acquaintance. Granted that introspection and "self-knowledge" are subject to limitations and illusions (e.g., Carruthers 2011), we nonetheless appear to enjoy a kind of access to the workings of the mind that reveals distinctive insights into its structure. We can have the experience of figuring out what to believe about a topic and subsequently feeling the conviction of the resulting conclusion. We can experience ourselves planning a course of action and then feeling committed to carry out each step of the plan. Such "folk" concepts as belief, plan, intention, and so on, are evidently fundamental to our understand of who we are. They help us navigate the social world, predicting and interpreting each other's behavior (Fodor, 1987b)—a capacity that may even be at the root of our self-understanding—and they play pivotal roles in many of the moral, social, and legal institutions that matter to us most (Malle and Nelson, 2003; Buchak, 2014).

A common approach to the topic of "bounded rationality" takes folk attitudes for granted and then motivates resource-sensitive norms for such attitudes. For instance, authors following Harman (1986) have explored the idea that one ought to avoid "cluttering" one's mind with frivolous or useless beliefs. Meanwhile, authors following Stalnaker (1984) have suggested that much of belief might be usefully "fragmented" in order to solve the aforementioned frame problem (§1.2). Pursuing this kind of approach requires postulating a relatively specific format for the attitudes in question. For instance, for the "clutter" metaphor to make sense, harboring more beliefs must generally demand greater memory usage.

The stance adopted in this Element is slightly different. Though we might grant that our ordinary conceptions of folk notions like belief, desire, intention and so on, may indeed pinpoint real structural features of our minds—and thus at this level we can trust the basic deliverances of introspection—these conceptions remain ambiguous, even equivocal (see, e.g., Schwitzgebel 2010 on belief). As such, they vastly underdetermine many of the architectural features that matter for characterizing resources and their costs. In the spirit of the previous chapter on resource rational analysis, we could instead take the architecture itself as a target of analysis and try to work back to what such structures must be like in order to play the right resource rational role in our lives. Again as in the previous chapter, this endeavor may be informed by what we know about the architectural substrate and its resource capacities. Thus, rather than

taking propositional attitudes and other folk concepts for granted and investigating what is resource rational for an agent characterized in these terms, we rather use resource rationality as a tool for uncovering fundamental psychological building blocks, potentially even vindicating some of our pre-theoretical conceptions.

The present chapter is merely a speculative, preliminary gesture at such a project. Much of the current literature in resource rational analysis—including examples like those in §5.3.3 and §5.3.4—can already be seen as pursuing this line, probing questions about how belief-like states might be generated, or how sequential policies might be implemented. The aim here will be to explore the possibility that many cognitive structures—including attitudes like beliefs and intentions, as well as capacities for metacognition and metareasoning—might themselves be understood as solutions to problems of resource optimization.

Such an endeavor can be taken as a species of "creature construction" in the spirit of Grice (1975b); see also Bratman (2000). Roughly speaking, our task is to sketch "an ascending order of psychological types" (Grice, 1975b, p. 38), such that each successive type is able to solve some fundamental problem that eluded previous agent types in the sequence. This endeavor has a different character from the more quantitative, experiment-driven methodology described in the previous chapter. It is more speculative and typically less exact. Despite this, the hope is that it may help clarify some very general ways in which evident psychological features of human agents emerge from resource rational considerations. To the extent that it succeeds, the strategy may then transition into a methodology closer to resource rational analysis, facilitating the study not only of adult human agents, but also of human pre-verbal infants, non-human animals, and even complex artifacts that display some of the same complex behavioral repertoires as ours. And if successful, our ordinary "folk" concepts will have been an essential guide in hypothesizing aspects of what generates intelligent behavior.

The strongest version of a creature construction argument demonstrates the *inevitability* of certain cognitive traits. That is, one might like to show that almost any suitable process of learning, evolution, or optimization would inevitably result in the psychological feature of interest. Holding fixed enough of the existing architectural background facts, this can sometimes be achieved (see, e.g., §5.3.2). However, especially when it comes to the highest-level cognitive traits, such as propositional attitudes, we may remain content with weaker arguments. Merely showing that the trait plausibly solves a problem of resource rationality would be illuminating.

6.1 Representation

We began our discussion in this Element with a classical picture of an agent embedded in an environment, related to it by perception in one direction, and action in the other (recall Fig. 1.1 and Fig. 2.1). We never presupposed that any of the fundamental aspects of the agent's situation—viz. states S, actions A, probability p, and so on—are represented anywhere inside the "agent box." A virtue of the framework is that it applies to any agent whatsoever, no matter whether (nor how) any of these aspects are encoded.

Many problems, however, appear to be "representation hungry" (Clark and Toribio, 1994), in the sense that solving them requires sufficient internal structure to mediate between perception and action. This is especially salient in "informationally translucent" environments (Sterelny, 2003). It is worth distinguishing a very minimal sense of representation, in which we require no more than the existence of complex internal states to mediate between perceptual input and behavioral output, from the more common, robust conception employed in philosophy and cognitive science. On the latter conception, we only speak of representation when we can associate some reasonably stable "meaning" or "content" with stably identifiable representational "vehicles" (e.g., Shea 2018, Chapter 2). We discuss each in turn.

6.1.1 Internal Structure

What we called a strategy (Def. 2) is a very general description of a *behavior*, mapping every history of environmental interaction to a next action. A good strategy will often require substantial sensitivity to the state (or to observations generated from the state as represented in a POMDP), which will invariably demand physical resources.

One very general way of making this precise uses automata theory. Def. 11 specifies an inherent memory cost for implementing a strategy (see also Def. 12 and Prop. 3), captured by the smallest implementing automaton. Any artifact that is well modeled by a k-state automaton maintains at least k distinct internal states (physical "configurations"), sensitively transitioning from one to another as a function of each new observation. In the physical world such maintenance can only be achieved through energetic consumption, and the pursuit of energy sources introduces an obvious opportunity cost. As we saw in §4.4.2 (especially Example 23), the possibility of undergoing stochastic, or randomized, transitions in one's internal states may mitigate these costs. But even in the probabilistic (non-stationary) setting it will certainly remain the case that good strategies demand a substantial number of distinct internal configurations. The resource rational framework—specifically with automata-theoretic resource costs (Def. 11)—precisely captures a fundamental tradeoff between success of a strategy and the resource costs involved in maintaining complex sequences of internal configurations.

In the Markovian setting (formalized by MDPs, Def. 2.3), sensitivity to observations can be measured by information theoretic cost (Def. 15). Recall that one intuition (§4.3.4) motivating an entropic cost function centers around the *internal coding* of distinct states (Fig. 4.4). The more a strategy diverges from the "default" on average, the longer the codes will need to be (on average) to transform observations into appropriate actions. To the extent that we take this talk of codes seriously, it is all but presupposed that the codes have meaning: they "denote" or "represent" a state, or some aspect or function of the state. This brings us to the philosophical discussion of representational content.

6.1.2 Representational Content

It has long been appreciated that an agent could benefit from "internalizing" aspects of the external world, so that structures internal to the agent's mind could "stand for" external environmental features (Craik, 1943). Some theorists take the stance that few, if any, cognitive behaviors are truly underwritten by representations in this stronger sense (e.g., Hutto and Myin 2013). But of course that depends on how we analyze representational content, and perhaps also what kinds of physical entities can serve as representational vehicles. Without any such constraints, there is a relatively uninteresting sense in which any successful behavior whatsoever implies the existence of an internal "model" of the environment (Thobani, 2023). Achieving a complex behavior may demand a complex suite of distinct internal states—as just discussed in §6.1.1—but that may fall short of the more demanding concept of representation that has been at the center of much philosophical discussion.

This more demanding concept goes beyond an ability to occupy a large number of distinct internal configurations. The claim is that, within the pattern of internal states, there are relatively stable physical vehicles that play an important role in the agent's behavioral repertoire. Specifically, the behavior is made possible by patterns of operations over "contentful" representational vehicles, which "bear exploitable relations to distal features of the environment" (Shea, 2018, p. 36). Such a claim was more or less implied by all of the examples discussed in Chapter 5. The analysis of oculomotor control (§5.3.1) postulates representations of words; models of the ventral stream (§5.3.2) postulate representations of increasingly abstract properties of the visual stimulus; policy compression (§5.3.3) involves encoding features of the state and presum-

ably also actions; and the sampling account of anchoring (§5.3.4), at a minimum, attributes representations of estimated values for a given proposition. Another very simple example has received a great deal of attention in the philosophical literature:

Example 29. In a famous series of experiments, Lettvin et al. (1959) demonstrated that the frog retina incorporates an array of "feature detectors," sensitive to edges, movement, and other changes in incoming light patterns. Some of these cells respond most actively, "when a dark object, smaller than a receptive field, enters that field, stops, and moves about intermittently thereafter" (p. 1951), in other words, when a typical fly—the frog's staple energy source—is immediately present. The excitation of these cells, in turn, elicits a tongue movement, which typically results in the capture and consumption of the fly.

Putting aside subtleties about what precisely is being represented in this pattern of neural activity and why (see Millikan 2023 for a recent discussion), it seems that the frog's nervous system, beginning already in the optic nerve, is effectively *tracking* the presence of—indeed *representing*—(something like) a fly. This plays a fairly obvious role in solving a large-scale sequential decision problem, namely, ensuring sufficient energetic and nutritional sustenance in an environment where relatively dark centripetal movement tends to be food.

What might representation have to do with conserving resources (cf. Schulz 2018)? One way of approaching this question is to compare representational strategies like the one in Example 29 (or those in §5.3) to the way a notional resource-unbounded agent might solve a problem. A common caricature employs the concept of a "lookup table" agent (see, e.g., Block 1981; Maloney and Mamassian 2009 for different uses of this thought experiment). Imagine, for instance, that every pattern of light hitting the frog retina followed a completely independent causal path to the appropriate behavioral response. That is, rather than passing through a general purpose "fly detector," the frog brain would need to embody a separate pathway from every possible light pattern to the appropriate motor output.

Even if this look-up table solution to the problem were physically feasible, it is well known that such a piecemeal strategy causes a problem for learning. Transferring and generalizing beyond past experience depend intimately on the ability to extract invariant features—that is, to form representations—from that experience (Bengio et al., 2013). If we think of "data" as another type of resource—not a *cognitive* resource, but certainly one relevant to what it means for things to "go well" for most agents—then some type of representation appears to be crucial for resource-efficient learning (cf. the discussion of sample complexity in §2.3.3).

It is likely that representational cognitive strategies evolved in large part because their importance for robust learning (Sterelny, 2003). However, even if we put the learning problem to the side, it is an obvious waste of neural resources to encode every stimulus separately, especially when only certain distinctions are behaviorally relevant. This was precisely the intuition motivating the entropic cost function (Def. (15)) discussed in §5.3.3. In the case of the frog, it is simply not relevant at what exact speed and angle the fly is entering the visual field. For such a simple behavioral repertoire, a binary signal (fly present or not) suffices. This advantage in efficiency is even more pronounced in cases where the very same representation (or system of representations) is employed toward many disparate ends (cf. 6.3 below).

Insofar as such a representation groups together many different inputs on the basis of common features, the result can be understood as one type of *abstraction*, a topic in its own right.

6.2 Levels of Abstraction

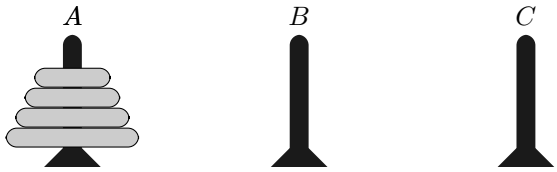
Dewey suggested that, "abstraction is the heart of thought; there is no way—other than accident—to control and enrich concrete experience except through an intermediate flight of

thought with conceptions, relations, abstracta" (quoted in Winther 2014). Much if not all of thought inherently involves abstractions. One of the remarkable capacities of many agents, including human agents, is to employ different *levels* of abstraction in a flexible manner. We are able to move from one level of granularity, to either a finer or a coarser level, often in a way that seems most appropriate to the immediate problem being solved.

The need for increasingly abstract representations in decision making was discussed at some length already in Savage's (1954) treatise on rational choice: "a smaller world is derived from a larger world by neglecting some distinctions between states" (p. 9). The usual way to think of a state abstraction is as a *partition* on the set of states into disjoint subsets, reifying those subsets as new states. For instance, in the discussion of policy compression (§5.3.3), a shopper may ignore distinctions between brands. And in Example 29 from the previous section, the frog's brain effectively ignores distinctions between similar light pattern trajectories.

But virtually anything that can be represented can be represented in a more or less abstract way. In particular it is also true of actions, as the following classic example illustrates:

Example 30 (Tower of Hanoi)**.** Consider the Tower of Hanoi problem, which involves moving a stack of rings on a peg (A in the picture below) to another peg (C in this picture), with the constraint that no larger ring can ever be placed on top of a smaller ring:



This problem can be encoded as an MDP with no uncertainty about the state or noise in the transitions. The states include all legal configurations of rings on pegs, while the "primitive" actions include moving a ring from the top of one stack to the top of another, again provided there is no smaller ring on that peg. In the "goal" state, the rings are all on peg C. Thus, the only state/action pair with positive utility is one that involves moving the smallest ring (from peg A or B) to peg C, when the remaining rings are already appear on C.

Following §2.2, a strategy for this problem would need to specify an action for all $3^4 = 81$ legal configurations. But there is a way of thinking about this problem in terms of more abstract states and actions. Rather than focus on sequences of concrete moves, it is helpful to break this problem down into a kind of recursion, using the "abstracted" action, *move the top N rings on peg X to peg Y*, which we might symbolize as $\texttt{Move}(N, X, Y)$.[1] The entire goal sequence then becomes a single action in this sense, $\texttt{Move}(4, A, C)$, which can itself be broken down into a sequences of simpler instances of the same abstraction:

1. $\texttt{Move}(3, A, B)$; 2. $\texttt{Move}(1, A, C)$; 3. $\texttt{Move}(3, B, C)$.

$\texttt{Move}(3, X, Y)$ can similarly be decomposed, and in general, $\texttt{Move}(N + 1, X, Y)$ is solved by:

1. $\texttt{Move}(N, X, \phi(X, Y))$; 2. $\texttt{Move}(1, X, Y)$; 3. $\texttt{Move}(N, \phi(X, Y), Y)$.

where $\phi(X, Y)$ denotes whatever the remaining peg is other than X and Y. At the "base" case of the recursion, $\texttt{Move}(1, X, Y)$ is again a primitive action.

[1] These are sometimes called *options* in the literature on reinforcement learning (Sutton et al., 1999). Note also that this is an abstraction in Church's (1941) sense of λ-abstraction, since we are considering sequences of operations—in this case, of actions—in a way that *abstracts* from any particular triple of arguments.

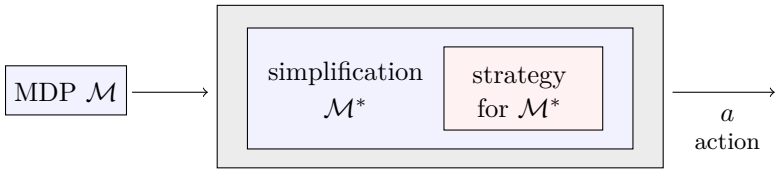


Figure 6.1: Similar to Fig. 4.4, we are now assuming that an agent is encoding some aspects of their deliberative predicament, formalized by $\mathcal{M}$. The simplified representation, $\mathcal{M}^*$, may omit details (abstraction) and it may also *mis*represent aspects of $\mathcal{M}$ (idealization). The figure is inspired by Ho et al. (2022), Fig. 1(c).

Simon (1975) analyzed the memory requirements for carrying out strategies like this one, which could also be studied with the automata theoretic framework discussed in §3.4. While the number of distinct "internal states" (i.e., states of the automaton) grows linearly with the number of rings, it is significantly smaller than that required for a strategy that does not collapse the 3^k many states, or increasingly long sequences of actions, at all. It could thus be encoded with only minimal demands on cognitive resources.

This example is somewhat removed from ordinary cognition—indeed, it is often used as a litmus test for problem solving abilities (see, e.g., Kotovsky et al. 1985). But similar abstractions can be shown to arise routinely in quite ordinary thought (Posner and Keele, 1968), witness the examples discussed in §5.3.3 (action), Example 29 (vision), and elsewhere. Ellis et al. (2023) offer a recent study of abstractions over spaces of programs, including for intuitive physical reasoning, showing how such abstractions can be synthesized from a small number of examples.

It is common in philosophy of science to distinguish abstraction from *idealization*: while abstraction involves omission of details, idealization involves *mis*representation of the assumed facts (Jones, 2005). Abstraction and idealization often go together. For instance, in a study of human planning, Ho et al. (2022) investigate human-like approximations to the transition function in an MDP that mix idealization and abstraction. The approximations are idealizations because the transition probabilities are inaccurate; they are inaccurate because they are themselves built from causal representations that omit detail (cf. Icard and Goodman 2015).

The general problem of abstraction (and of idealization) can be understood in the following way. We suppose a Markov decision process, $\mathcal{M}$—or at least parts of $\mathcal{M}$—could be somehow presented to an agent. Think of, e.g., presenting the Tower of Hanoi problem to a person. The agent's task is to "solve" the MDP, that is, to pursue a strategy that achieves high cumulative reward (in the sense of Eq. (2.9)). Instead of representing every detail of $\mathcal{M}$—all of the states, actions, utility, and transition probabilities—the agent could instead encode a "coarsened" or "idealized" version $\mathcal{M}^*$ of $\mathcal{M}$. See Fig. 6.1. Which $\mathcal{M}^*$ should an agent employ? And more generally, what representations should an agent use when solving a problem?

As already discussed (in §6.1), (sample-)efficient learning demands abstraction. This is true quite independently of any considerations of cognitive or computational resources: the very possibility of bringing one set of experiences (or data) to bear on another grounds out in an ability to identify common structure across those experiences. Much work in learning theory has investigated strategies for abstraction that maximize a learning objective, in model-based approaches to reinforcement learning (see, e.g., Jiang et al. 2015), and even in model-free approaches, witness neural approximate Q-learning discussed above in Remark 6 (§2.3).

But quite apart from its importance for learning, we have seen that level of abstraction trades off against the costs of deliberation: as in Example 30, employing abstractions can render solvable problems that are not otherwise solvable (or are solvable only under severe

cost) by resource limited agents. We have seen alternative ways of measuring the cost or complexity of a deliberative strategy, for instance, using automata theory (§3.4), information theory (§4.3), and so on. However this measurement is done, for agents that can represent aspects of their environment and of their own capacities within that environment, we should expect a significant degree of abstraction to emerge in any resource rational strategy. Often this will be combined with an ability to modulate flexibly among multiple levels of abstraction, depending on the specific problem being solved (cf. §6.5 below).

6.3 Beliefs and Desires

Throughout this Element we have assumed that an agent's predicament—at virtually any scale, from milliseconds to lifespans—can be understood in terms of something like a Markov decision process (Def. 3, or perhaps a generalization of it like a sequential decision problem or a POMDP). These characterize ways our world might be, in terms of states and transitions between them, together with their probabilities, and utilities for each action in each state. We have been assuming that such a characterization captures something fundamental about the kinds of problems we face, and about the nature of (instrumental) rationality.

So far in this chapter we have suggested that resource-constrained solutions to these problems may often necessitate the deployment of internal representations, possibly at various levels of representational granularity. But what exactly should be represented, and how? A natural suggestion is that it could be advantageous to represent aspects of the agent's assumed predicament, that is, to *internalize* aspects of the situation as characterized by an MDP, the putative determinants of what it means for things to "go well" for the agent.[2]

For simple environments that demand simple behavioral repertoires, more rudimentary types of representation may suffice. Example 29 of the frog's "fly detector" exemplifies this situation. The pattern of activation blends both indicative and imperative moods (Millikan, 1989): it indicates the presence of a fly and simultaneously it brings about the relevant behavior in a direct way. For suitably complex environments that appear to demand considerable flexibility in behavior, however, it has often been suggested that maintaining something like beliefs and desires—or else some other types of internal states with both mind-to-world and world-to-mind directions of fit (§1.1)—would be well worth the energetic costs. Belief-like representational states in particular have the following panoply of qualities:[3]

1. In the agent's psychological economy their function is to track aspects of the environment (or potentially of the agent's own internal states).

2. They are potentially *decoupled* from immediate perceptual input. Thus, unlike the fly detector (Example 29), belief-like representations do not have to co-occur, spatially or temporally, with what is being represented.

3. They may reflect information collected from a variety of sensory modalities, and from a variety of different external sources.

4. They tend to play a role in a multitude of different (potential) decision problems, enjoying at least some degree of stability across time and contexts.

[2]Note that this is delicate in light of the perspectival nature of resource rational pronouncements (§1.1). We presume that the characterization plays into a good explanation of how the agent came to exemplify a particular strategy for dealing with their environment, through evolution, learning, and maturation. Cf. §1.3 and §5.4.

[3]This list is inspired by discussions from Godfrey-Smith (1998) and Sterelny (2003). Note that Sterelny admits the terms 'belief-like' and 'desire-like' are "weasilish" (p. ix). Some authors instead employ the terms 'cognitive' and 'conative' for essentially the same distinction (e.g., Schulz 2018).

More may be required to count as a genuine belief state, e.g., a substantial tendency toward integration with other belief-like states, and so on (see, e.g., Bratman 1992). The more qualities associated with them, of course, the more grounds there are for skepticism about their existence as "psychologically real" constructs (Dennett, 1987). As discussed in §1.1, some have viewed beliefs as part of a holistic, rational reconstruction of an agent's behavior without any commitment to corresponding internal states (e.g., Lewis 1974; Davidson 1975). The characteristics listed in (1)-(4), however, encompass a wider array of possible states, including notions like "acceptance" (Stalnaker, 1984; Bratman, 1992) and even less intricate constructs, possession of which may not be exceedingly demanding, but still quite useful.

When it comes to desire-like states—preferences, goals, desires, and others with world-to-mind direction of it—some have supposed that their satisfaction is just what it means for things to "go well" for the agent (cf. §1.1). For human agents in particular, this may even involve a pressure toward coherence with one's *higher order* desires about what first-order desires to have (e.g., Frankfurt 1971). Such views essentially presuppose that an agent has internalized evaluative distinctions among possible actions or outcomes. Questions about resource rationality can then be raised concerning the instrumental attainment of those ultimate aims.

Alternatively, we can again ask the same question about desires as about beliefs: Why would it make (instrumental) resource rational sense to internalize evaluative information about possible outcomes in the first place, even for relatively simple agents? One salient possibility is that it promises flexibility about *what ends to pursue*. This was evident in the relatively sophisticated example of the Tower of Hanoi puzzle (Example 30): a good way of achieving the ultimate goal is by breaking it down into a series of subgoals. Especially when combined with belief-like representations, some potential ends may appear more or less achievable, and flexible pursuit of the most worthwhile ends may demand some way of representing not only what is more or less likely, but also what would be more or less desirable.

The expected utility calculus reviewed in Chapter 2 was of course crafted precisely to systematize the instrumentally rational integration of belief-like and desire-like representations—specifically, of probabilities and utilities. Yet, as emphasized repeatedly, a resource rational agent might succeed in a given environment, achieving high expected utility, by employing a shortcut that bypasses encodings of any of these quantities. In the framework, they are to be understood, first and foremost, as part of an externalist characterization of the task.

The question is thus: at what point do the evident advantages of encoding and maintaining belief-like and/or desire-like representations outweigh the resource costs involved in doing so? Our folk conceptions of ourselves, and perhaps of other animals, would surely suggest that we are, in some sense, past that point.

6.3.1 Probabilities and Desirabilities

In fact, there is evidence of probabilistic information encoded already in low-level perception, where the tradeoff between robust representation and cognitive costs is highly salient. Consider a basic (but potentially very important) task like figuring out whether one can safely walk down a steep ravine. We often have access to multiple sources of information about such matters, including visual, haptic, auditory, and other cues (cf. (3) above). Each one of these cues supplies potential data about, e.g., the grade of the incline or the distance to a drop-off. Moreover, different cues may enjoy different levels of reliability. A large body of research suggests that perceptual systems are able to combine these cues in a statistically optimal manner (see Rescorla 2015; Ma et al. 2022 for reviews). That is, not only does the brain seem to be estimating probabilities for each modality separately; it evidently combines them in a way that is sensitive to their respective levels of reliability, and potential dependencies among them.

Where and how these probabilistic representations are encoded—and even whether they are explicitly encoded at all—remain hotly contested questions (see, e.g., Rahnev et al. 2021; Rescorla 2020). The answers are far from obvious, especially in light of the considerable computational demands on probabilistic inference (van Rooij et al., 2019), and on high-level perceptual inference in particular (Brooke-Wilson, 2023). One possibility, at least for some lower-level perceptual tasks, is that only "sufficient statistics" are explicitly encoded, for instance, the mean and variance, and simple computations are performed over these small numbers of parameters (Ma et al., 2006, 2022). Another possibility, mentioned at several earlier points (§4.1.3, §5.3.4), is that the brain employs a sampling approximation, capable of generating token representations with probability proportional to an assumed "subjective" probability.

Another line of work has revealed evidence of an internal encoding of utility (in the economic sense of (4) from §1.1), particularly in the midbrain dopamine system (Schultz, 2016). However, as highlighted by Gershman and Daw (2012), much of the data on probability and utility is actually consistent with a somewhat deflationary hypothesis according to which the brain is directly (and approximately) encoding *expected utilities* of different possibilities. In other words, there may be no decomposition into probabilities and utilities that have to be combined by taking an expectation (Eq. (2.2) in Chapter 2). The model-free method of Q-learning (§2.3.1) involves such an encoding. The Q-value of an action a in state s (Eq. (2.11)) essentially captures how well an agent could *expect* to fare in the long term by taking a in the current state s. As discussed in §2.3.4, there is evidence of such model-free representations in addition to more goal-sensitive "model-based" representations in the brain, as well as effective adjudication between the two. But even the latter, more sophisticated model-based representations often appear to involve substantial simplifications and shortcuts (see Momennejad et al. 2017 on the so-called successor representation for one notable example).

As we move from relatively low-level "subpersonal" processes to higher ("person-level") cognition, the neural underpinnings of putative representations become even less scrutable. But we again know, without any further ado, that perfect probabilistic calculations are not even within the realm of physical possibility. Moreover, characteristic of high-level cognition is the seemingly endless range of considerations that could be relevant for a given probabilistic judgment, one instance of the notorious frame problem (§1.2). The fact that probabilistic propositions can be considered at multiple levels of conceptual abstraction (§6.2) easily leads to patterns of judgment that violate the axioms of probability (Tversky and Koehler, 1994).

The literature on person-level judgment and decision making is immense, but one recent trend is worth highlighting because of its resonance with the themes explored in this Element. The possibility that probability is partly encoded dynamically in terms of sampling propensities (§4.1.3) makes salient the idea that judgments and decision making under uncertainty will be highly impacted by *what comes to mind*, from memory (e.g., Stewart et al. 2006) or in imagination (e.g., Bear et al. 2020). Moreover, what comes to mind will naturally be impacted by factors that are not directly reflective of likelihoods in the world. When deliberating about what to do, for instance, it would be beneficial for possibilities to come to mind that are not only probable or typical, but also likely to be *good* or *valuable* in the given context (Morris et al., 2019; Bear et al., 2020). What spontaneously comes to mind in humans is often impacted by moral considerations, and this influence turns out to affect not just deliberation about what to do, but also judgments involving concepts that one might have thought were non-moral—perhaps even "purely" probabilistic—like causation (Icard et al., 2017), or whether an agent performed an action *freely* or *intentionally* (Phillips et al., 2015; Phillips and Cushman, 2017).

Were we to design a resource limited agent to perform intricate tasks in complex environments, it does seem that we would want to instill some ability to reason both about what is likely to happen, and about what they want to achieve. But the mechanisms by which these

abilities are deployed will likely serve many functions at once, and this sharing of cognitive resources results in distinctive departures from the austere decision theoretic picture of a clean decomposition and recomposition of probabilities and utilities. To some extent this insight is ascertainable from first principles. Studying judgment and decision making in humans (and non-human animals), for both perception and high-level cognition, makes it particularly vivid.

6.3.2 "All-out" Beliefs

Among the most basic constructs of commonsense psychology is belief. Not partial belief, or credence, but the "all-out" state one is in when, for instance, feeling convinced today is Tuesday, or that this Element is too long. Belief appears not just in folk psychology, but also centrally in epistemology and in much of cognitive and social psychology.

A point of contention in philosophy is how to understand the relationship between (all-or-nothing) beliefs and degrees of belief (or subjective probabilities). Both purportedly have the role of tracking the world, and they are both thought to feed into decision making. On one prominent view, subjective probabilities are most fundamental, with beliefs emerging as useful *abstractions* of those finer-grained representations. The motivation for such abstractions is essentially one of resource limitations. Typical of this perspective is the following quotation from Ross and Schroeder (2014) (see also Holton 2009, Chapter 2):

> If we had infinite cognitive resources, then we'd have no need for an attitude of outright belief by which to guide our actions, for we could reason in an ideal Bayesian manner on the basis of our credences and preferences alone. But such reasoning isn't feasible for cognitively limited agents like us, and so we need an attitude of outright belief or of settling on the truth of propositions, so as to limit what we consider in our reasoning to possibilities consistent with what we have settled on. (p. 286)

The assumption is that something like the picture we have sketched in Chapter 2 does indeed characterize the agent's deliberative predicament, and that an ideal agent would simply deliberate on the basis of those very components, viz. probabilities and utilities (or preferences). Assuming those features of the situation are accessible to the agent (recall our externalist starting point in §1.1), it does seem intuitive that calculating with numerical probabilities could be more difficult than reasoning with presumed settled facts, at least in conscious thought. There are numerous approaches to "extracting" or "abstracting" beliefs from probabilities (e.g., Lin and Kelly 2012; Leitgeb 2017; Mierzewski 2022). Such proposals could be offered as hypotheses for resource rational deliberative architectures, insofar as the resulting suite of operations is in fact more efficient and effective than probabilistic alternatives (cf. Staffel 2019).

Another possibility is that the processes underpinning intuitive instances of belief are the same processes that encode uncertainty (cf. Weisberg 2020). This is perhaps clearest for the analysis of subjective probability in terms of sampling propensities (discussed in the previous section §6.3.1 and in §4.1.3). It may be, for instance, that the ordinary conception of belief tracks just those sampling propensities that are entirely stable across a wide variety of contexts, generating the very same possibility each time.[4] On this account there would be no need of further resource rational justification for beliefs, since they emerge as a special limiting case of an architectural feature that already enjoys such a rationale.

Finally, there are other accounts of belief that ground its primary function not in individual-level decision making and deliberation, but in social and interpersonal phenomena. For instance,

[4]This would be an instance of what Greco (2015) calls the "simple view" of the relation between beliefs and subjective probabilities. Importantly here, the latter are understood mechanistically as sampling propensities.

Buchak (2014) argues that beliefs play a fundamental role in our practices of holding one another blameworthy or accountability, a role for which subjective probabilities are inappropriate. Meanwhile, MacFarlane (2023) suggests that the role of belief is to ground *reason* giving. That is, as he puts it, "to say that an agent believes p is to say, roughly speaking, that the agent treats p as a candidate reason," e.g., in argumentation. While such rationales point in rather different directions, they may also ultimately interface with resource constraints that surface in interactive settings, for instance, involving communication or coordination (cf. §6.6 below).

6.4 Intentions and Plans

It is sometimes suggested that beliefs and desires stand at the core of our commonsense view of agents (e.g., Jeffrey 1965; Lewis 1974). Given the sequential decision making framework characterizing the key structural of what it means for things to "go well" for an agent, we can understand belief-like states and desire-like states as at least roughly tracking probabilities and utilities, respectively. We have been exploring the possibility that a resource rational agent might *internalize* these components, so that they are not just reflexively responding to environmental cues, but consulting "internal models" of the world and their place in it.

Another fundamental component of the sequential decision making framework laid out in Chapter 2 is what we called a *strategy* (Def. 2, or a "policy"). This external specification of an agent's behavioral dispositions highlights an additional feature that could profitably be internalized, namely, *what they will do in the future*. That is, an agent possessing rich internal states—beliefs about the world, complex evaluative representations about what would be more or less desirable, and so on—might draw upon these and other representations to generate *plans*. The idea that future planning underlies much of intelligent behavior was a founding doctrine in the development of cognitive science (Newell et al., 1959; Miller et al., 1960).[5]

Particularly in the seminal work of Bratman (1987) (see also Bratman et al. 1988) on plans and intentions, planning agency is seen as part of a solution to a problem of resources. Perhaps an idealized agent without resource limitations could "precompile" an entire optimal strategy for all possible contingencies, and then simply carry out one action after another.[6] Alternatively, one could simply compute, at each point, what the best action is at that time, under the assumption that one will always be acting optimally at each future point (thus, essentially internalizing the optimal Q-function in Eq. (2.12)).

Neither of these approaches may be available to a resource constrained agent. In between the two extremes of planning ahead for all possible contingencies, and reflexively responding (whether optimally or otherwise) to each incremental decision problem, it makes sense to form *partial plans* that fix some aspects of what one will do, while deferring other details until later. Bratman (1987) highlighted three key features of planning states—and their building blocks, *intentions*—that help conduce to resource rationality:

1. An intention to perform an action a at a time t should, in the normal course, cause the agent to take a at t. For instance, if I form an intention to take out the trash when I arrive, then upon arrival, absent any further deliberation about it, the intention should simply lead me to take out the trash.

[5] Notably, Miller et al. (1960) suggested that plans could be identified with *programs*, understood as specifications of sequences of actions that cause each action in turn. Programs in the resource rational framework specified here (§1.2.2) are importantly different: they may involve many different types of processing, including, e.g., perceptual processing, and they need not explicitly encode sequences of actions.

[6] After recognizing that such a suggestion is "preposterous," Savage (1954) went on to say that it is nonetheless "the proper subject" of decision theory (p. 16), which in turn necessities a focus on "small worlds" (cf. §6.2).

It seems useful to figure out what one will do in a future context, particularly when time and other resources promise to be scarce when the decision is to be realized. Some of the deliberative costs can then offloaded to an earlier time, when cool, reflective thought is perhaps easier to realize. This benefit is only effective to the extent that planning does indeed guide those downstream actions.

2. Plans for the future will only bring about the relevant actions so long as the agent has taken the necessary *means* to render those actions feasible. There is thus a pressure to ensure the means to one's ends are indeed met.

   This pressure toward means-end coherence simultaneously promises resource benefits: in deliberating about what to do, one can all but ignore possibilities that would be inconsistent with the ends one has already committed to pursuing. Bratman et al. (1988) refer to this as a *filtering* procedure. It can help generate simplified, "local" decision problems, eliminating some possibilities altogether, and coarsening the state and action spaces to levels of abstraction appropriate for achieving a given ends (cf. §6.2).

3. As long as plans are partial, there will inevitably be circumstances in which a previously formed plan no longer appears to be a good way of achieving one's ends. Dogmatic adherence to a prior plan goes counter to the supposed aim of maximizing one's ends.

   At the same time, stopping to consider at each point whether one's plan is still the best course threatens the very functions described in (1) and (2). Furthermore, some personal characteristics that are widely admired, such as "grit" (Morton and Paul, 2019), promote resistance to revision even in highly inauspicious circumstances. Although there is a question about when resolve turns to dogma, some degree of *stability* seems essential.

These key features help clarify why we might expect a resource rational creature to maintain something like intentions and planning states in the first place.

The question raised in (3), about when it makes sense for an agent to revise a previously drawn plan, drives to the heart of resource rationality. Following the broad framework introduced by Bratman et al. (1988), a number of researchers have explored concrete revision strategies across different types of environments (Kinny and Georgeff, 1991; Schut et al., 2004; van Zee and Icard, 2015). This is a paradigmatic exercise in comparing the resource rationality of alternative *agent programs* for specific tasks, in precisely the sense we have described in this Element. Findings from toy environments suggest that even very simple reconsideration strategies can achieve optimality (see van Zee and Icard 2015).

Much of the literature on intention reconsideration focuses on more complex phenomena, especially so-called "temptation cases," in which one's judgment about what is best temporarily shifts; recall Example 6. One noteworthy feature of such cases is that they often involve desires that are, generally speaking, good for one to have. After all, being drawn to rich information sources (like the television in Example 6) is surely adaptive. The same goes for many other common temptations, like rich foods and inactivity (i.e., rest). Whereas internalizing desires for these made sense for a niche at an earlier point in human history, such desires may become maladaptive (cf. Sterelny 2003). Moreover, the preferences one has at the point of temptation often disagree sharply with what one wants before and after.

As discussed in §2.1.7, there are numerous responses to such cases, ranging from versions of "resolute choice" (McClennen, 1990; Gauthier, 1994) to "sophisticated choice" (Steele, 2010), to the proposal that one ought to configure oneself so that the possibility of reconsideration does not even seriously arise (Holton, 2009). Although some of these proposals are likely cumbersome to formalize in their entirety, there is a sense in which many of the debates are about what "policies of revision" would be most resource rational for a life overall.

Following the rationales laid out in (1)-(3), there is is widespread agreement that planning should be understood as, "instrumental to the attainment of our objectives, not an objective itself" (Gauthier, 1996, p. 218), where those objectives amount to one's "life going as well as possible" (Gauthier, 1994, p. 690). Temptation cases present situations in which we must, "say no to short-run considerations and yes to "long-run" expected return" (McClennen, 1990, p. 236). What one finds in this literature is thus a range of conjectures about what revision policies will in fact achieve the best "long-run expected return" for a typical human life.[7] Another way of putting this, in context of this Element, is that planning agency, together with specific policies for revision, is offered as part of a specification of an *agent program* that enjoys high expected utility for agents like us, in environments like ours.

Another important rationale for plans and intentions, and their stability, stems again from problems of interpersonal coordination and cooperation (see, e.g., Bratman 1987; Gauthier 1994, among many others). Yet even without moving to the multi-agent setting—and outstanding details about revision policies notwithstanding—there seems ample reason to internalize plans for a sufficiently sophisticated resource-constrained agent.

6.5 Metareasoning

The resource rationality framework is marked by a separation between the underlying task facing an agent, determined by their interaction with their environment, and the efficient use of internal cognitive resources in confronting that task. As remarked at several points (e.g., §1.3, §6.1), this relies on a distinction between what is "inside" versus "outside" the agent. The focus is not just on abstract strategies for solving the task, but on *programs* that implement abstract strategies and also specify the internal structures and processes used in pursuit of the task. In this chapter so far we have been exploring the idea that a program π might involve operations over internal representations, with various possible functional roles in the program (desire-like, belief-life, plan-like, etc.), at various possible levels of abstraction.

It is but a small step to imagine that an agent program might represent aspects of that very program. In other words, the agent may represent not only features of the external environment and their place in it, but also features of their own internal processing. We have already encountered numerous points where some type of *cognitive control* might be relevant:

1. Representing cost (Kool and Botvinick, 2018) and accuracy (Daw et al., 2005) of different learning strategies so as to adjudicate among them (Example 2, §2.3.4);
2. Representing possible levels of abstraction and selecting among them (§6.2), and more generally encoding a problem into an appropriate format (Fig. 6.1);
3. Combining perceptual estimates of distal properties from different modalities (§6.3.1);
4. Internalizing desires about one's own desires (Frankfurt, 1971);
5. Intending that one's intention be robust to reconsideration (Holton, 2009), and more generally determining when to reconsider a prior plan (§6.4).

We can understand *metareasoning* as cognitive control facilitated by representations of other internal representations and operations. Metareasoning in this sense seems to be ubiquitous in human thought, viz. assessing confidence, reasoning about what one does or does not know,

[7] Not everyone in this literature takes planning to be purely instrumental, however. For instance, Bratman (2018) argues that stability of intention plays crucially into a fundamentally valuable kind of "diachronic self-governance" that is threatened, among other times, whenever an agent goes in for temptation.

selecting from among a "toolbox" of problem solving strategies, and so on (see, e.g., Proust 2013; Griffiths et al. 2019). The broad rationale for why it might make sense to spend resources reasoning about how best to use one's resources is intuitive. Particularly as one's mental capacities and available strategies are sufficiently rich and flexible, one inherits the "metaproblem" of determining how those capacities should be brought to bear on a problem, that is, how best to allocate those very limited resources.

Just as information about the world can be relevant to one's decision about which action to take, so information about one's own mind can be relevant to one's decision about which deliberative strategy to employ. Recall that the latter problem is assumed to be captured by balance between expected utility of a strategy and its costs, Eq. (1.1), reproduced here:

$$\mathbf{V}(\pi) \quad = \quad \mathbf{U}(\pi) - C(\pi). \tag{1.1}$$

Supposing there is some space Π of candidate "cognitive strategies," the task is to select among them to maximize Eq. (1.1). Perhaps the simplest way to internalize this problem is to consider the setting of just two relevant options—think or act—a predicament sometimes known as "Hamlet's problem" (Fodor, 1987a).

Example 31 (Hamlet's Problem)**.** Suppose, as in §4.3, there is some *default* behavior δ that an agent would display without any further thought. And suppose there is some deliberative procedure π that the agent could pursue, which might improve ultimate decision quality. We have encountered numerous ways of measuring the cost of π, e.g., by the KL-divergence from δ to π (Def. 14). The agent should engage in further deliberation, i.e., adopt π, to the extent that its cost is outweighed by its expected marginal benefit relative to δ. The quantity that must be strictly positive is sometimes known as the *value of computation* (Horvitz, 1987; Griffiths et al., 2019), by analogy with value of information (e.g., from an experiment):

$$\text{Value of computation} \quad = \quad \big(\mathbf{U}(\pi) - \mathbf{U}(\delta)\big) - C(\pi).$$

This basic predicament is of course faced by agents constantly, and a solution to it can simply be "hard-wired" into an agent architecture. While such "reflexive" strategies enjoy the benefits of model-free architectures (§2.3.1), for all the reasons discussed in this chapter, representational strategies that involve reasoning explicitly about crucial determinants of the problem—in this case, resource costs and benefits of alternative strategies—may in fact be preferable.

As acknowledged in §1.3, determining the optimal solution to Eq. (1.1) is not in general a task that an agent can solve through deliberation. By the time deliberation has begun one has already effectively ruled out any programs that are incompatible with that start.[8] This is especially pronounced in Example 31. Nonetheless, in settings where the relevant space Π is sufficiently separate from the current context of deliberation, this tension need not arise. In the tradition of resource rational analysis (Chapter 5), a variety of studies have demonstrated that key phenomena involving metareasoning in humans can be understood in resource rational terms (e.g., Cushman 2020; Milli et al. 2021; Callaway et al. 2022, among many others).

In sum, for agents with suitably complex internal lives, the question arises of how to deploy one's array of possible cognitive strategies in an effective way that fits the task at hand. Metareasoning emerges as a potential solution to that predicament, when the costs of such "reasoning about reasoning" are outweighed by the benefits. This is already enough to justify

[8]Of course, some programs may be "self-ratifying" in the sense that the program somehow "endorses" its own execution (somewhat analogous to Jeffrey's 1965 notion of ratifiability). But note that there is a tension here if the selection process is costly. Similar to the argument for Prop. 4 (failure of the Folk Theorem for finite automata), one would presumably prefer a simpler program that skipped the unnecessary "preprocessing" step.

metareasoning from a resource rational perspective. But here again there are proposals that locate at least some of its function in the multi-agent setting. For instance, Shea et al. (2014) suggest that sharing confidence reports between minds helps facilitate complex coordination and cooperation. This leads to the last section of this chapter.

6.6 Communication, Coordination and Beyond

We could expand the list of important cognitive phenomena that can be understood, at least in part, in resource rational terms, to include mind-wandering (e.g., Christoff et al. 2011), moral judgment (e.g., Levine et al. 2023), trust (e.g., Nguyen 2022), counterfactual thought (e.g., Icard et al. 2018), emotion (e.g., Huys and Renz 2017), and many others. A repeating theme throughout our discussion so far is the possibility that some of what makes these cognitive constructs adaptive stems from their role in interpersonal activities, such as communication and coordination. It is worth briefly reflecting here on how resource rationality can accommodate, and be adapted to, the multi-agent setting.

The orientation in this Element has been decision theoretic, focusing on a single decision maker. As discussed in §3.1.1, there is controversy about how to assess rational behavior in (competitive) strategic multi-agent scenarios, with Nash equilibrium (and its relatives) sometimes offered as an alternative to maximization of expected utility. Chapter 3 showed how resources and their costs could be easily incorporated into game theory; in fact, the automata theoretic perspective has been explored most thoroughly in that setting. We also encountered a number of instances where incorporation of resource costs invalidates some seminal results on the subject, e.g., existence of equilibria (Example 12) and the Folk Theorem (Prop. 4).

Much of what makes human cultural achievement possible is our ability to engage in rich practices of cooperation and social learning, passing cumulative knowledge from one generation to the next (e.g., Boyd et al. 2011). How can these more cooperative forms of multi-agent interaction be captured in the resource rational framework? As a basic starting point it makes sense to postulate a single assessment function $\mathbf{U}$ that depends on the behavior of multiple agents. That is, unlike the strategic context where each agent's separate utility depends on what they both do (Eq. (3.1)), but like the "team reasoning" account of Gold and Sugden (2007) mentioned in §3.1.2, the function $\mathbf{U}(\sigma_0, \sigma_1)$ will assign a single number measuring how well the (here, just two) agents coordinate to achieve some joint task.

A paradigmatic example of cooperative joint activity, evidently essential for cultural accumulation, is linguistic communication. It turns out that a prominent approach to natural language pragmatics, the Rational Speech Act model (see, e.g., Goodman and Frank 2016, essentially formalizing Grice 1975a), can be understood in precisely the same resource rational framework as the one discussed in §4.3 for single-agent decision making.

Example 32 (Rational Speech Acts)**.** Imagine a speaker, knowledgeable about some underlying state s, and a listener who will update their view on the state given what the speaker says. In addition to some states S with distribution p (known to the speaker), there are some possible messages ("utterances") M that the speaker could produce. Thus, the listener will maintain a distribution $\lambda(s|m)$, while the speaker will maintain an utterance distribution $\sigma(m|s)$.

How might we measure success of a communicative exchange. Recall from Example 19 that a natural utility for guessing state s given an observation m is $\log \lambda(s|m)$: guessing more likely states is better. We might also assume—in the spirit of resource rationality—that there could be a cost $c(m)$ for sending message m, for instance, higher if m is longer or more difficult for the listener to process. So the utility of an exchange in which the sender produces message m and the listener infers the state is s can be taken as $u(m, s) = \log \lambda(s|m) - c(m)$. The overall

expected utility for a pair σ, λ of behaviors can then be rendered:

$$\mathbf{U}(\sigma, \lambda) \quad = \quad \sum_{s,m} p(s)\, \sigma(m|s)\, u(m, s).$$

Moving even closer to resource rationality, Zaslavsky et al. (2020) suggest a further cost term for the communication system as a whole. Let S be the random variable with distribution p, and define a "stationary" utterance distribution M so that $P(\mathsf{M} = m) = \sum_s p(s)\sigma(m|s)$. This is the long-term frequency of uttering m, averaged over states (analogous to Eq. (4.10)).

The mutual information $I(\mathsf{S}; \mathsf{M})$ between these random variables measures the average KL-divergence from the "stationary" utterance distribution M to the state-specific speaker distribution $\sigma(m|s)$ (recall Def. 15). As in the discussion of policy compression (§4.3.4, §5.3.3), this can be viewed as the average additional information that needs to be conveyed when communicating about each state, or alternatively as the amount of information that utterances tend to carry about the underlying state. There is an obvious sense in which a more informative signaling system will be costlier in this sense. Thus taking mutual information as the cost we arrive at a multi-agent version of Eq. (1.2):

$$\mathbf{V}(\sigma, \lambda) \quad = \quad \mathbf{U}(\sigma, \lambda) - \frac{1}{\beta}\, I(\mathsf{S}; \mathsf{M}).$$

By an argument similar to that for Theorems 6 and 8 (Appendix 4.A.3), one can show (see Zaslavsky et al. 2020, Supplementary Material) that the optimal pair of distributions satisfies:

$$\begin{aligned} \sigma(m|s) \quad &\propto \quad P(\mathsf{M} = m)\, e^{\beta u(m,s)} \\ \lambda(s|m) \quad &\propto \quad \sigma(m|s)\, p(s). \end{aligned}$$

Note that λ is essentially a Bayesian listener, while σ is perfectly analogous to the optimal action distribution in Eq. (4.12), derived in Theorem 8. A large body of empirical phenomena is predicted by models of communication between speakers and listeners grounded in these two distributions (see Goodman and Frank 2016 for a review).[9]

In fact, efficiency appears to be ubiquitous in language, at nearly all levels of linguistic analysis (Gibson et al., 2019). Much current research employs the same information theoretic apparatus as Example 32 (and §4.3), conceiving of speech as operating through a "noisy channel" in the sense of Shannon (1948). The communicative aim of language is to transmit as much information along this channel as efficiently as possible, mirroring the signature resource rational tradeoff (Eq. (1.1)). In Example 32 the cost is measured by mutual information, but it can also be measured in terms of working memory necessary for listener comprehension (which can also be approximated by mutual information between words), minimum description length in a hypothesized "language of thought" (Kemp and Regier, 2012; Steinert-Threlkeld, 2021), and others. This general framework has been used to explain linguistic universals pertaining to color terms, kinship terms, word order, morphological marking and redundancy, degree of ambiguity, and many other phenomena (see Gibson et al. 2019 for a review).

While this line of work emphasizes group-level resource rationality around communication (and sometimes learning), there is also work targeting important ways that language may be

[9] Strictly speaking, the usual Rational Speech Act model employs an iterative approximation to these distributions, starting with a "literal" listener (or speaker) and building up from there. Furthermore, the ("perseverence") term $P(\mathsf{M} = m)$ is usually not included. Zaslavsky et al. (2020) show that the model without this term emerges if we replace mutual information with conditional entropy, in effect ignoring the entropy of M (cf. Corollary 2), though there may be empirical reason for including the perseverence term.

resource optimal at the individual level. For instance, it has been suggested that natural human languages achieve near optimality in the way that they links "sound and meaning," mediating between thought and its expression (e.g., Chomsky 2005). There are extensive debates about which of the many functions of language are responsible for its evolution (Christiansen and Kirby, 2003). While obviously important, one of the attractive features of the resource rational framework is that it focuses less on these specific explanatory questions (though it may of course feed into them, cf. §5.4). Many different aspects of language can be resource rational in numerous different respects, and the evidence so far does in fact favor such a pluralistic stance.

Chapter 7

Conclusion and Outlook

Resource rationality, as understood in this Element, is grounded in relatively traditional tools of decision theory, viz. probabilities, utilities, and so on. We employ these tools as an external means of assessing the instrumental rationality of a wide array of agents across different kinds of environments. The framework applies equally to sophisticated intentional agents with rich mental lives, to subsystems of those agents, and even to non-living artifacts, so long as they can be parsed as "doing something."

Distinctive of resource rationality is its separation between what is "inside" and what is "outside" the agent (Figs. 1.1, 2.1, 4.4, 6.1). The primary role of the traditional decision theoretic tools is to assess external *behavior*, that is, interactions between the agent and their environment, captured formally by a *strategy*. The framework does not impose assumptions about what is inherently good or bad for the agent, but rather accommodates a broad range of such assumptions in the form of a utility function (§1.1).

The concept of a *program* summarizes the relevant internal structure of the agent. A program has two key features: it consumes resources and thus incurs costs, and it implements a strategy. In other words, a program is just a costly implementation of a strategy. Resource costs are understood as the utility foregone, which the same resources could instead have been used to achieve (§1.4). In practice, however, costs are determined by some measurable quantity that ideally stands in for true opportunity costs. Having thus specified both how to assess (the strategy implemented by) a program π on the task, and the (pseudo-)cost of π, the overall value of the program is given by a weighted sum, Eq. (1.2), repeated here:

$$\mathbf{V}(\pi) \quad = \quad \mathbf{U}(\pi) - \frac{1}{\beta}\widetilde{C}(\pi).$$

Meanwhile, on so-called panoramic approaches it suffices simply to compare all "feasible" programs by how well they perform on the underlying task (§1.3).

Two recurring analyses of cost can be found throughout this Element, both of which purport to measure a kind of inherent resource demand for implementing a strategy:

> **Automata-theoretic costs:** Every strategy can be implemented by some (possibly finite-state) automaton. We can measure the inherent cost of the strategy by the size (number of states) of the smallest implementing automaton (Def. 11). This captures a lower bound on the amount of memory required to carry out the strategy, insofar as the agent must be able to occupy at least this number of distinct internal configurations.
>
> **Information-theoretic costs:** Relative to a "default" strategy (which can be taken as the stationary action distribution in the sequential case), we can measure the cost of a

> strategy σ by the (average, in the sequential case) relative entropy, or "KL-divergence," from the default strategy to σ (Defs. 14, 15). In the sequential case this becomes the mutual information between the action distribution and the state distribution. Intuitively, this measures the degree to which the strategy tracks distinctions in the world, again highlighting intuitions about memory requirements on *any* implementation of the strategy.

A natural question is how these two analyses relate to one another. As anticipated in Remark 3, it is possible to have arbitrarily high automata-theoretic cost with no information-theoretic cost whatsoever, because we can have complex strategies that depend in no way on the state (or observations of it). These will simply be sequences of actions (like in Example 25).

In the other direction, for any stationary strategy we will merely need an automaton with as many (machine) states as there are actions that receive positive probability in some (world) state. All strategies with the same set of positive-probability actions will have the same automata-theoretic complexity, ranging from those with minimal information-theoretic complexity (because all states are ignored) to those with maximal information-theoretic complexity (given by the entropy of the state distribution, in the case of a one-to-one correspondence between states and actions). Thus, in general, the two analyses are orthogonal.

Another analysis of cost appeared in §5.3.4, viz. number of iterations in a Markov chain for drawing a Monte Carlo sample. The broader literature includes many others, such as the amount of energy consumed in maintaining ion gradients (Niven and Laughlin, 2008), the volume of connections between neural cells (Cherniak, 2012), the number of samples from a generative model (Vul et al., 2014), description length in an abstract language of thought (Piantadosi et al., 2016), the number of steps in a Turing machine program (van Rooij et al., 2019), and many others. On the view propounded here (§1.4), these should all be understood as approximations to opportunity costs, which may be more or less apt for different purposes.

The resulting framework promises numerous uses, including the typical roles for a theory of rationality, helping to clarify ways in which forms of deliberation and behavior may be adaptive or improvable. In contrast to more idealized rational frameworks that in effect posit, "infallible powers of calculation and independence from the institutional and ideological context we inhabit" (O'Neill, 1987, p. 56), resource rationality accommodates and even highlights the various quirks and limitations of concretely embodied agents. The framework is distinctive to the extent that we are able to codify important architectural features of the agents under study.

Resource rationality is also singular in the way it blurs the boundary between the normative and the descriptive. The focus of much of this Element has been its use in identifying and clarifying the very architectural features that could themselves serve as input to the framework. Resource rational analysis, as a scientific methodology (Chapter 5), aims to draw descriptive conclusions from normative assumptions (viz. that some way of solving a problem is a good one), and there is even some promise of shedding light on very general structural aspects of human thought and cognition (Chapter 6). The deliverances of this enterprise can then feed back into further resource rational pronouncements that take this finer structure into account. We always begin (as Stalnaker 2010 memorably put it in a different context) in the middle.

By design, resource rationality is perspectival in its pronouncements, embracing a dependence on how we choose to characterize an agent's predicament. This includes both the underlying task and the relevant features of the agent's deliberative capacities and proclivities. Our ability to limn the structure of rational thought and action thus relies on identifying useful abstractions and idealizations to frame our investigation. In other words, our goal as theorists is to achieve a resource rational balance between fidelity to the facts and theoretical tractability. Somewhat perversely, idealizing agents' abilities often affords analytical simplicity, which makes finding the right balance particularly acute. It is perhaps a testament to its scope that such resource-sensitive theoretical tradeoffs fall within the purview of the framework itself.

Bibliography

Ackerman, N. L., Freer, C. E., and Roy, D. M. (2019). On the computability of conditional probability. *Journal of the ACM*, 66(3):1–40.

Ahmed, A. (2021). *Evidential Decision Theory*. Cambridge University Press.

Anderson, J. R. (1990). *The Adaptive Character of Thought*. Lawrence Earlbaum Associates, Inc.

Anderson, J. R. (1991a). The adaptive nature of human categorization. *Psychological Review*, 98(3):409–429.

Anderson, J. R. (1991b). Is human cognition adaptive? *Behavioral and Brain Sciences*, 14:471–484.

Andersson, H. and Herlitz, A. (2022). *Value Incommensurability: Ethics, Risk, and Decision-Making*. Routledge Studies in Ethics and Moral Theory.

Annas, J. (2011). *Intelligent Virtue*. Oxford University Press.

Anscombe, G. E. M. (1957). *Intention*. Harvard University Press.

Aronowitz, S. (2021). Exploring by believing. *Philosophical Review*, 130(3):339–383.

Arpaly, N. and Schroeder, T. (2012). Deliberation and acting for reasons. *Philosophical Review*, 121(2):209–239.

Arrow, K. J. (2004). Is bounded rationality unboundedly rational? Some ruminations. In Augier, M. and March, J. G., editors, *Models of a Man: Essays in Memory of Herbert A. Simon*, page 47–55. MIT Press.

Åström, K. J. (1965). Optimal control of Markov processes with incomplete state information. *Journal of Mathematical Analysis and Applications*, 10:174–205.

Aumann, R. J. (1981). Survey of repeated games. In Böhm, V. and Nachtkamp, H. H., editors, *Essays in Game Theory and Mathematical Economics*, pages 11–42. Bibliographisches Institut Mannheim/Wien/Zürich.

Aumann, R. J. (1987). Correlated equilibrium as an expression of Bayesian rationality. *Econometrica*, 55(1):1–18.

Aumann, R. J. (1997). Rationality and bounded rationality. In Hart, S. and Mas-Colell, A., editors, *Cooperation: Game-Theoretic Approaches*, pages 219–231. Springer.

Azar, M. G., Osband, I., and Munos, R. (2017). Minimax regret bounds for reinforcement learning. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, pages 263–272.

Bareinboim, E., Forney, A., and Pearl, J. (2015). Bandits with unobserved confounders: A causal approach. In Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc.

Barlow, H. B. (1961). Possible principles underlying the transformation of sensory messages. *Sensory Communication*, 1(1):217–234.

Battaglia, P. W., Hamrick, J. B., and Tenenbaum, J. B. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences*, 110(45):18327–18332.

Baumol, W. J. and Quandt, R. E. (1964). Rules of thumb and optimally imperfect decisions. *American Economic Review*, 54(2):23–46.

Bear, A., Bensinger, S., Jara-Ettinger, J., Knobe, J., and Cushman, F. (2020). What comes to mind? *Cognition*, 194:104057.

Bechtel, W. and Richardson, R. C. (2010). *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. MIT Press, second edition.

Bellman, R. (1957). *Dynamic Programming*. Princeton University Press.

Bengio, Y., Courville, A., and Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828.

Bennett, C. H. (1982). The thermodynamics of computation—a review. *International Journal of Theoretical Physics*, 21(12):905–940.

Bialek, W., Nemenman, I., and Tishby, N. (2001). Predictability, complexity, and learning. *Neural Computation*, 13:2409–2463.

Binmore, K. (2015). Why all the fuss? The many aspects of the Prisoner's Dilemma. In *Peterson (2015)*, pages 16–34.

Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. Clarendon Press.

Blackwell, D. (1970). On stationary policies. *Journal of the Royal Statistical Society. Series A*, 133(1):33–37.

Blackwell, D. and Dubins, L. (1962). Merging of opinions with increasing information. *The Annals of Mathematical Statistics*, 33(3):882–886.

Block, N. (1981). Psychologism and behaviorism. *The Philosophical Review*, 90(1):5–43.

dal Bó, P. and Fréchette, G. R. (2018). On the determinants of cooperation in infinitely repeated games: A survey. *Journal of Economic Literature*, 56(1):60–114.

Bowers, J. S. and Davis, C. J. (2012). Bayesian just-so stories in psychology and neuroscience. *Psychological Bulletin*, 138(3):389–414.

Boyd, R., Richerson, P. J., and Henrich, J. (2011). The cultural niche: Why social learning is essential for human adaptation. *Proceedings of the National Academy of Sciences*, 108(2):10918–10925.

Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press.

Bramley, N. R., Dayan, P., Griffiths, T. L., and Lagnado, D. A. (2017). Formalizing Neurath's ship: Approximate algorithms for online causal learning. *Psychological Review*, 124(3):301–338.

Bramson, A., Grim, P., Singer, D. J., Berger, W. J., Sack, G., Fisher, S., Flocken, C., and Holman, B. (2017). Understanding polarization: Meanings, measures, and model evaluation. *Philosophy of Science*, 84(1):115 – 159.

Bratman, M. E. (1987). *Intention, Plans, and Practical Reason*. Harvard University Press.

Bratman, M. E. (1992). Practical reasoning and acceptance in a context. *Mind*, 101(401):1–15.

Bratman, M. E. (2000). Valuing and the will. *Philosophical Perspectives*, 14:249–265.

Bratman, M. E. (2018). *Planning, Time, and Self-Governance: Essays in Practical Rationality*. Oxford University Press.

Bratman, M. E., Israel, D. J., and Pollack, M. E. (1988). Plans and resource-bounded practical reasoning. *Computational Intelligence*, 4(4):349–355.

Briggs, R. A. (2015). Costs of abandoning the sure-thing principle. *Canadian Journal of Philosophy*, 46(4-5):827–840.

Briggs, R. A. (2019). Normative theories of rational choice: Expected utility. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Fall 2019 edition.

Brooke-Wilson, T. (2023). How is perception tractable? *Philosophical Review*. Forthcoming.

Broome, J. (1991). *Weighing Goods*. Wiley.

Broome, J. (2021). Reasons and rationality. In Knauff, M. and Spohn, W., editors, *The Handbook of Rationality*, pages 129–136. MIT Press.

Brunel, N. and Nadal, J.-P. (1998). Mutual information, Fisher information, and population coding. *Neural Computation*, 10(7):1731–1757.

Brunswick, E. (1957). Scope and aspects of the cognitive problem. In Gruber, H., Hammond, K. R., and Jessor, R., editors, *Contemporary Approaches to Cognition*, pages 5–31. Harvard University Press.

Buchak, L. (2013). *Risk and Rationality*. Oxford University Press.

Buchak, L. (2014). Belief, credence, and norms. *Philosophical Studies*, 169(2):285–311.

Buchak, L. (2022). Normative Theories of Rational Choice: Rivals to Expected Utility. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Summer 2022 edition.

Bullmore, E. and Sporns, O. (2012). The economy of brain network organization. *Nature Reviews Neuroscience*, 13(5):336–349.

Bulsara, A. R. and Zador, A. (1996). Threshold detection of wideband signals: A noise-induced maximum in the mutual information. *Physical Review E*, 54(3):2185–2188.

Callaway, F., van Opheusden, B., Gul, S., Das, P., Krueger, P. M., Griffiths, T. L., and Lieder, F. (2022). Rational use of cognitive resources in human planning. *Nature Human Behavior*, 6(8):1112–1125.

Camerer, C. F. (2003). *Behavioral Game Theory: Experiments on Strategic Interaction*. Princeton University Press.

Camerer, C. F. and Ho, T.-H. (2015). Behavioral game theory experiments and modeling. In Young, H. P. and Zamir, S., editors, *Handbook of Game Theory with Economic Applications*, volume 4, pages 517–573. North Holland.

Cao, R. and Yamins, D. L. K. (2021). Explanatory models in neuroscience: Part 2 – constraint-based intelligibility.

Carr, J. R. (2022). Why ideal epistemology? *Mind*, 131(524):1131–1162.

Carruthers, P. (2011). *The Opacity of Mind: An Integrative Theory of Self-Knowledge*. Oxford University Press.

Cesa-Bianchi, N., Gentile, C., Lugosi, G., and Neu, G. (2017). Boltzmann exploration done right. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS)*, page 6287–6296.

Chater, N., Tenenbaum, J. B., and Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Sciences*, 10(7):287–291.

Cherniak, C. (1986). *Minimal Rationality*. MIT Press.

Cherniak, C. (1994). Philosophy and computational neuroanatomy. *Philosophical Studies*, 73(2/3):89–107.

Cherniak, C. (2012). Neural wiring optimization. In Hofman, M. A. and Falk, D., editors, *Progress in Brain Research*, volume 195, pages 361–371. Elsevier.

Chiappori, P.-A., Levitt, S., and Groseclose, T. (2002). Testing mixed-strategy equilibria when players are heterogeneous: The case of penalty kicks in soccer. *American Economic Review*, 92(4):1138–1151.

Chomsky, N. (2005). Three factors in language design. *Linguistic Inquiry*, 36(1):1–22.

Christiansen, M. H. and Chater, N. (2016). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, 39:1–72.

Christiansen, M. H. and Kirby, S. (2003). *Language Evolution*. Oxford University Press.

Christoff, K., Gordon, A., and Smith, R. (2011). The role of spontaneous thought in human cognition. In Vartanian, O. and Mandel, D. R., editors, *Neuroscience of Decision Making*, page 259–284. Psychology Press.

Church, A. (1941). *The Calculi of Lambda Conversion*. Annals of Mathematics Studies. Princeton University Press.

Churchland, P. S. (1986). *Neurophilosophy: Toward a Unified Science of the Mind-Brain*. MIT Press.

Clark, A. and Toribio, J. (1994). Doing without representing? *Synthese*, 101(3):401–431.

Cohen, L. J. (1981). Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences*, 4:317–370.

Collins, A. G. E. and Frank, M. J. (2012). How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. *European Journal of Neuroscience*, 35(7):1024–1035.

Colombo, M. and Hartmann, S. (2017). Bayesian cognitive science, unification, and explanation. *The British Journal for the Philosophy of Science*, 68:451–484.

Colombo, M. and Seriès, P. (2012). Bayes in the brain—on Bayesian modelling in neuroscience. *The British Journal for Philosophy of Science*, 63:697–723.

Cook, J. and Lewandowsky, S. (2016). Rational irrationality: Modeling climate change belief polarization using Bayesian networks. *Topics in Cognitive Science*, 8(1):160–179.

Craik, K. (1943). *The Nature of Explanation*. Cambridge University Press.

Crupi, V. and Calzavarini, F. (2023). Critique of pure Bayesian cognitive science: A view from the philosophy of science. *European Journal for Philosophy of Science*, 13(28).

Cushman, F. (2020). Rationalization is rational. *Behavioral and Brain Sciences*, 43:e28.

Danks, D. (2008). Rational analyses, instrumentalism, and implementations. In Chater, N. and Oaksford, M., editors, *The Probabilistic Mind: Prospects for Bayesian Cognitive Science*, pages 59–75. Oxford University Press.

Danks, D. and Eberhardt, F. (2011). Keeping Bayesian models rational: The need for an account of algorithmic rationality. *Behavioral and Brain Sciences*, 34(4):197–197.

Dann, C., Lattimore, T., and Brunskill, E. (2017). Unifying PAC and regret: Uniform PAC bounds for episodic reinforcement learning. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS)*, page 5717–5727.

Dasgupta, I., Lampinen, A. K., Chan, S. C. Y., Creswell, A., Kumaran, D., McClelland, J. L., and Hill, F. (2022). Language models show human-like content effects on reasoning.

Dasgupta, I., Schulz, E., and Gershman, S. J. (2017). Where do hypotheses come from? *Cognitive Psychology*, 96:1–25.

Daskalakis, C., Goldberg, P. W., and Papadimitriou, C. H. (2009). The complexity of computing a Nash equilibrium. *Communications of the ACM*, 52(2):89–97.

Davidson, D. (1975). Hempel on explaining action. *Erkenntnis*, 10(3):239–253.

Daw, N. D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, 8(12):1704–1711.

Dennett, D. C. (1969). *Content and Consciousness*. Routledge.

Dennett, D. C. (1987). *The Intentional Stance*. MIT Press.

Dewey, J. (1896). The reflex arc concept in psychology. *Psychological Review*, 3(4):357–370.

Dorst, K. (2023). Rational polarization. *Philosophical Review*. Forthcoming.

Eagle, A. (2005). Randomness is unpredictability. *The British Journal for the Philosophy of Science*, 56(4):749–790.

Eberhardt, F. and Danks, D. (2011). Confirmation in the cognitive sciences: The problematic case of Bayesian models. *Minds and Machines*, 21(3):389–410.

Edgeworth, F. Y. (1879). The hedonical calculus. *Mind*, (4):394–408.

Ellis, K., Wong, L., Nye, M., Sablé-Meyer, M., Cary, L., Pozo, L. A., Hewitt, L., Solar-Lezama, A., and Tenenbaum, J. B. (2023). Dreamcoder: Growing generalizable, interpretable knowledge with wake–sleep Bayesian program learning. *Philosophical Transactions of the Royal Society A*, 381(2251):1–18.

Elster, J. (1983). *Sour Grapes: Studies in the Subversion of Rationality*. Cambridge University Press.

Englich, B., Mussweiler, T., and Strack, F. (2006). Playing dice with criminal sentences: The influence of irrelevant anchors on experts' judicial decision making. *Personality and Social Psychology Bulletin*, 32(2):188–200.

Epley, N. and Gilovich, T. (2006). The anchoring-and-adjustment heuristic. *Psychological Science*, 17(4):311–318.

Estes, W. K. (1959). The statistical approach to learning. In Koch, S., editor, *Psychology: A Study of a Science*, volume 2, pages 380–491. McGraw-Hill.

Faisal, A. A., Selen, L. P. J., and Wolpert, D. M. (2008). Noise in the nervous system. *Nature Reviews Neuroscience*, 9:292–303.

Feynman, R. P. (1998). *Lectures on Computation*. Addison-Wesley.

de Finetti, B. (1937). La prévision: Ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré*, 7:1–68.

de Finetti, B. (1974). *Theory of Probability*, volume 1. Wiley, New York.

Fodor, J. A. (1987a). Modules, frames, fridgeons, sleeping dogs, and the music of the spheres. In *Pylyshyn (1987)*, pages 139–149.

Fodor, J. A. (1987b). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. MIT Press.

Frankfurt, H. G. (1971). Freedom of the will and the concept of a person. *Journal of Philosophy*, 68(1):5–20.

Frederick, S., Loewenstein, G., and O'Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature*, 40(2):351–401.

Friedman, M. and Savage, L. J. (1948). The utility analysis of choices involving risk. *Journal of Political Economy*, 56(4):279–304.

Fudenberg, D. and Maskin, E. (1986). The Folk Theorem in repeated games with discounting or with incomplete information. *Econometrica*, 54(3):533–554.

Gaifman, H. and Liu, Y. (2018). A simpler and more realistic subjective decision theory. *Synthese*, 195(10):4205–4241.

Galton, F. (1889). *Natural Inheritance*. MacMillan.

Gauthier, D. (1994). Assure and threaten. *Ethics*, 104(4):690–721.

Gauthier, D. (1996). Commitment and choice: An essay on the rationality of plans. In Farina, F., Hahn, F., and Vannucci, S., editors, *Ethics, Rationality, and Economic Behaviour*, pages 217–245. Oxford University Press.

Geiger, A., Lu, H., Icard, T., and Potts, C. (2021). Causal abstractions of neural networks. In Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W., editors, *Advances in Neural Information Processing Systems*.

Geiger, A., Wu, Z., Potts, C., Icard, T., and Goodman, N. D. (2023). Finding alignments between interpretable causal variables and distributed neural representations. Ms., Stanford University.

Genewein, T., Leibfried, F., Grau-Moya, J., and Braun, D. A. (2015). Bounded rationality, abstraction, and hierarchical decision-making: An information-theoretic optimality principle. *Frontiers in Robotics and AI*, 2(27):1–24.

Gershman, S. J. (2020). Origin of perseveration in the trade-off between reward and complexity. *Cognition*, 204:104394.

Gershman, S. J. and Daw, N. D. (2012). Perception, action, and utility: the tangled skein. In Rabinovich, M., Friston, K., and Varona, P., editors, *Principles of Brain Dynamics: Global State Interactions*, pages 293–312. MIT Press.

Gershman, S. J., Horvitz, E. J., and Tenenbaum, J. B. (2015). Computational rationality: A converging paradigm for intelligence in brains, minds and machines. *Science*, 349:273–278.

Gershman, S. J. and Lai, L. (2021). The reward-complexity trade-off in schizophrenia. *Computational Psychiatry*, 5(1):38–53.

Gershman, S. J., Markman, A. B., and Otto, A. R. (2014). Retrospective revaluation in sequential decision making: A tale of two systems. *Journal of Experimental Psychology: General*, 143:182–194.

Ghavamzadeh, M., Mannor, S., Pineau, J., and Tamar, A. (2015). Bayesian reinforcement learning: A survey. *Foundations and Trends in Machine Learning*, 8(5-6):359–483.

Gibson, E., Futrell, R., Piantadosi, S. P., Dautriche, I., Mahowald, K., Bergen, L., and Levy, R. (2019). How efficiency shapes human language. *Cognition*, 23(5):389–407.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky (1996). *Psychological Review*, 103:592–596.

Gigerenzer, G. and Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science*, 1:107–143.

Gigerenzer, G. and Selten, R. (2001). Rethinking rationality. In Gigerenzer, G. and Selten, R., editors, *Bounded Rationality: The Adaptive Toolbox*, pages 1–12. MIT Press.

Gigerenzer, G., Todd, P. M., and The ABC Research Group (2000). *Simple Heuristics that Make Us Smart*. Oxford University Press.

Gilboa, I. and Samet, D. (1989). Bounded versus unbounded rationality: The tyranny of the weak. *Games and Economic Behavior*, 1(3):213–221.

Gilboa, I. and Schmeidler, D. (2001). *A Theory of Case-Based Decisions*. Cambridge University Press.

Gillespie, J. (1977). Natural selection for variances in offspring numbers—a new evolutionary principle. *American Naturalist*, 111:1010–1014.

Glimcher, P. W. (2005). Indeterminacy in brain and behavior. *Annual Review of Psychology*, 56:25–56.

Godfrey-Smith, P. (1998). *Complexity and the Function of Mind in Nature*. Cambridge University Press.

Gold, N. and Sugden, R. (2007). Collective intentions and team agency. *Journal of Philosophy*, 104(3):109–137.

Goldman, A. I. (1986). *Epistemology and Cognition*. Harvard University Press.

Good, I. J. (1950). *Probability and the Weighing of Evidence*. Charles Griffin & Co.

Good, I. J. (1952). Rational decisions. *Journal of the Royal Statistical Society. Series B*, 14(1):107–114.

Goodman, N. D. and Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Science*, 20:818–829.

Goodman, N. D. and Tenenbaum, J. B. (2008). Probabilistic Models of Cognition. `http://v1.probmods.org/`. Accessed: 2022-29-9.

Gopnik, A. (2020). Childhood as a solution to explore–exploit tensions. *Philosophical Transactions of the Royal Society B*, 375:20190502.

Greco, D. (2015). How I learned to stop worrying and love probability. *Philosophical Perspectives*, 29:179–201.

Grice, H. P. (1975a). Logic and conversation. In Cole, P. and Morgan, J., editors, *Syntax and Semantics*, volume 3. Academic Press.

Grice, H. P. (1975b). Method in philosophical psychology (from the banal to the bizarre). *Proceedings and Addresses of the American Philosophical Association*, 48:23–53.

Griffiths, T. L., Callaway, F., Chang, M. B., Grant, E., Krueger, P. M., and Lieder, F. (2019). Doing more with less: Meta-reasoning and meta-learning in humans and machines. *Current Opinion in Behavioral Sciences*, 29:24–30.

Griffiths, T. L., Lieder, F., and Goodman, N. D. (2015). Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in Cognitive Science*, 7:217–229.

Griffiths, T. L. and Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition. *Psychological Science*, 17(9):767–773.

Hahn, M., Futrell, R., Levy, R., and Gibson, E. (2022). A resource-rational model of human processing of recursive linguistic structure. *Proceedings of the National Academy of Sciences*, 119(43):1–9.

Hahn, U. (2014). The Bayesian boom: Good thing or bad? *Frontiers in Psychology*, 5.

Halpern, J. Y. and Pass, R. (2015). Algorithmic rationality: Game theory with costly computation. *Journal of Economic Theory*, 156:246–268.

Halpern, J. Y., Pass, R., and Seeman, L. (2014). Decision theory with resource-bounded agents. *Topics in Cognitive Science*, 6(2):245–257.

Harman, G. (1986). *Change in View*. MIT Press.

Harremoës, P. (2001). Binomial and Poisson distributions as maximum entropy distributions. *IEEE Transactions on Information Theory*, 47(5):2039–2041.

Harsanyi, J. C. (1982). Subjective probability and the theory of games: Comments on Kadane and Larkey's paper. *Management Science*, 28(2):120–124.

Hellman, M. and Cover, T. (1970). Learning with a finite memory. *The Annals of Mathematical Statistics*, 41(3):765–782.

Hellman, M. and Cover, T. (1971). On memory saved by randomization. *The Annals of Mathematical Statistics*, 42(3):1075–1078.

Hitchcock, C. (2016). Conditioning, intervening, and decision. *Synthese*, 193(4):1157–1176.

Ho, M. K., Abel, D., Correa, C. G., Littman, M. L., Cohen, J. D., and Griffiths, T. L. (2022). People construct simplified mental representations to plan. *Nature*, 606:129–136.

Hobson, A. (1969). A new theorem of information theory. *Journal of Statistical Physics*, 1(3):383–391.

Holton, R. (2009). *Willing, Wanting, Waiting*. Oxford University Press.

Hopcroft, J. E. and Ullman, J. D. (1979). *Introduction to Automata Theory, Languages, and Computation*. Addison-Wesley, 1st edition.

Horvitz, E. (1987). Reasoning about beliefs and actions under computational resource constraints. In *Proceedings of the 3rd Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 429–444.

Hume, D. (1739). *A Treatise of Human Nature*. Oxford University Press, Oxford.

Humphreys, P. (2009). The philosophical novelty of computer simulation methods. *Synthese*, 3(169):615–626.

Hurley, S. (2001). Perception and action: Alternative views. *Synthese*, 129:3–40.

Huttegger, S. M. (2017). *The Probabilistic Foundations of Rational Learning*. Cambridge University Press.

Hutto, D. D. and Myin, E. (2013). *Radicalizing Enactivism: Basic Minds without Content*. MIT Press.

Huys, Q. J. and Renz, D. (2017). A formal valuation framework for emotions and their control. *Biological Psychiatry*, 82(6):413–420.

Ibeling, D., Icard, T., Mierzewski, K., and Mossé, M. (2023). Probing the quantitative-qualitative distinction in probabilistic reasoning. *Annals of Pure and Applied Logic*.

Icard, T. (2014). Toward boundedly rational analysis. In Bello, P., Guarini, M., McShane, M., and Scassellati, B., editors, *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*, pages 637–642.

Icard, T. (2016). Subjective probability as sampling propensity. *Review of Philosophy and Psychology*, 7(4):863–903.

Icard, T. (2018). Bayes, bounds, and rational analysis. *Philosophy of Science*, 85(1):79–101.

Icard, T. (2020). Calibrating generative models: The probabilistic Chomsky-Schützenberger hierarchy. *Journal of Mathematical Psychology*, 95.

Icard, T. (2021). Why be random? *Mind*, 130(517):111–139.

Icard, T., Cushman, F., and Knobe, J. (2018). On the instrumental value of hypothetical and counterfactual thought. In *Proc. 40th Annual Meeting of the Cognitive Science Society*.

Icard, T. and Goodman, N. D. (2015). A resource-rational approach to the causal frame problem. In *Proceedings of the 37th Annual Conference of the Cognitive Science Society*.

Icard, T., Kominsky, J. F., and Knobe, J. (2017). Normality and actual causal strength. *Cognition*, 161:80–93.

Jaakkola, T., Jordan, M. I., and Singh, S. P. (1994). On the convergence of stochastic iterative dynamic programming algorithms. *Neural Computation*, 6(6):1185–1201.

Jara-Ettinger, J., Gweon, H., Schulz, L. E., and Tenenbaum, J. B. (2016). The naïve utility calculus: Computational principles underlying commonsense psychology. *Trends in Cognitive Sciences*, 20(8):589–604.

Jaynes, E. T. (1957). Information theory and statistical mechanics. *The Physical Review*, 106(4):620–630.

Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press.

Jeffrey, R. C. (1965). *The Logic of Decision*. McGraw-Hill.

Jiang, N., Kulesza, A., and Singh, S. (2015). Abstraction selection in model-based reinforcement learning. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning*, ICML'15, page 179–188.

Jin, C., Allen-Zhu, Z., Bubeck, S., and Jordan, M. I. (2018). Is Q-learning provably efficient? In *Proceedings of the 32nd International Conference on Neural Information Processing Systems (NeurIPS)*, page 4868–4878.

Jonas, E. and Kording, K. P. (2017). Could a neuroscientist understand a microprocessor? *PloS Computational Biology*, 13(1).

Jones, M. and Love, B. C. (2011). Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences*, 34(4):169–231.

Jones, M. R. (2005). Idealization and abstraction: A framework. *Poznań Studies in the Philosophy of the Sciences and the Humanities*, 86(1):173–218.

Joyce, J. M. (2009). *The Foundations of Causal Decision Theory*. Cambridge University Press.

Kadane, J. B. and Larkey, P. D. (1982). Subjective probability and the theory of games. *Management Science*, 27:113–120.

Kaelbling, L. P., Littman, M. L., and Cassandra, A. R. (1998). Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101:99–134.

Kahneman, D., Slovic, P., and Tversky, A. (1982). *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press.

Kahneman, D. and Tversky, A. (1996). On the reality of cognitive illusions: A reply to Gigerenzer's critique. *Psychological Review*, 103:582–591.

Kalai, E. (1990). Bounded rationality and strategic complexity in repeated games. In Ichiishi, T., Neyman, A., and Tauman, Y., editors, *Game Theory and Applications*, pages 131–157. Academic Press.

Karlin, S. and Taylor, H. (1975). *A First Course in Stochastic Processes*. Academic Press.

Kemp, C. and Regier, T. (2012). Kinship categories across languages reflect general communicative principles. *Science*, 336(6084):1049–1054.

Kerkez, B., Gruden, C., Lewis, M., Montestruque, L., Quigley, M., Wong, B., Bedig, A., Kertesz, R., Braun, T., Cadwalader, O., Poresky, A., and Pak, C. (2016). Smarter stormwater systems. *Environmental Science and Technology*, 50(14):7267–7273.

Kiesewetter, B. (2017). *The Normativity of Rationality*. Oxford University Press.

Kinny, D. N. and Georgeff, M. P. (1991). Commitment and effectiveness of situated agents. In *Proceedings of the 12th International Joint Conference on Artificial Intelligence*, pages 82–88.

Kitcher, P. (1990). The division of cognitive labor. *Journal of Philosophy*, 87(1):5–22.

Klein, C. (2018). Mechanisms, resources, and background conditions. *Biology and Philosophy*, 33(36):1–14.

Knight, F. (1934). "The Common Sense of Political Economy" (Wicksteed Reprinted). *Journal of Political Economy*, 42(5):660–73.

Kool, W. and Botvinick, M. (2018). Mental labour. *Nature Human Behavior*, 2(12):899–908.

Kotovsky, K., Hayes, J., and Simon, H. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology*, 17(2):248–294.

Kuhn, H. W. (1950). Extensive games. *Proceedings of the National Academy of Sciences*, 36(10):570–576.

Lai, L. and Gershman, S. J. (2021). Policy compression: An information bottleneck in action selection. In Federmeier, K. D., editor, *Psychology of Learning and Motivation*, volume 74, pages 195–232. Elsevier.

Lai, L., Huang, A. Z. H., and Gershman, S. J. (2022). Action chunking as policy compression. PsyArXiv.

Leitgeb, H. (2017). *The Stability of Belief: How Rational Belief Coheres with Probability*. Oxford University Press.

Lettvin, J. Y., Maturana, H. R., McCulloch, W. S., and Pitts, W. H. (1959). What the frog's eye tells the frog's brain. *Proceedings of the IRE*, 47(11):1940–1951.

Levine, S., Chater, N., Tenenbaum, J. B., and Cushman, F. (2023). Resource-rational contractualism: A triple theory of moral cognition.

Levy, R., Reali, F., and Griffiths, T. L. (2009). Modeling the effects of memory on human online sentence processing with particle filters. *Advances in Neural Information Processing Systems*, 21:937–944.

Lewis, D. K. (1974). Radical interpretation. *Synthese*, 23:331–344.

Lewis, R. L., Howes, A., and Singh, S. (2014). Computational rationality: Linking mechanism and behavior through bounded utility maximization. *Topics in Cognitive Science*, 6(2):279–311.

Lewis, R. L., Shvartsman, M., and Singh, S. (2013). The adaptive nature of eye movements in linguistic tasks: How payoff and architecture shape speed-accuracy tradeoffs. *Topics in Cognitive Science*, 5(3):1–30.

Lieder, F. and Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, 43:1–60.

Lieder, F., Griffiths, T. L., and Goodman, N. D. (2012). Burn-in, bias, and the rationality of anchoring. *Proceedings of the 26th International Conference on Neural Information Processing Systems (NeurIPS)*, pages 2699–2707.

Lieder, F., Griffiths, T. L., Huys, Q. J. M., and Goodman, N. D. (2017). The anchoring bias reflects rational use of cognitive resources. *Psychonomic Bulletin and Review*, 25(1):322–349.

Lin, H. and Kelly, K. T. (2012). Propositional reasoning that tracks probabilistic reasoning. *Journal of Philosophical Logic*, 41(6):957–981.

Lindsay, G. W. (2021). Convolutional neural networks as a model of the visual system: Past, present, and future. *Journal of cognitive neuroscience*, 33(10):2017–2031.

Lipton, R. J. and Markakis, E. (2004). Nash equilibria via polynomial equations. In Farach-Colton, M., editor, *LATIN 2004: Theoretical Informatics*, pages 413–422. Springer Berlin Heidelberg.

Lord, E. (2018). *The Importance of Being Rational*. Oxford University Press.

Luce, R. D. (1963). Detection and recognition. In Luce, R. D., Bush, R. R., and Galanter, E., editors, *Handbook of Mathematical Psychology*, pages 103–189. Wiley.

Luce, R. D. and Suppes, P. (1965). Preference, utility, and subjective probability. In Luce, R. D., Bush, R. R., and Galanter, E. H., editors, *Handbook of Mathematical Psychology*, volume 3, pages 249–410. Wiley.

Ma, W. J., Beck, J. M., Latham, P. E., and Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, 9:1432–1438.

Ma, W. J., Kording, K., and Goldreich, D. (2022). *Bayesian Models of Perception and Action: An Introduction*. MIT Press.

Ma, W. J. and Woodford, M. (2020). Multiple conceptions of resource rationality. *Behavioral and Brain Sciences*, 43:30–31.

MacFarlane, J. (2023). Belief: What is it good for? *Erkenntnis*. Forthcoming.

Machado, A. (1993). Learning variable and stereotypical sequences of responses: Some data and a new model. *Behavioral Processes*, 30:103–130.

Mainen, Z. F. and Sejnowski, T. J. (1995). Reliability of spike timing in neocortical neurons. *Science*, 268:1503–1506.

Malle, B. F. and Nelson, S. E. (2003). Judging *Mens Rea*: The tension between folk concepts and legal concepts of intentionality. *Behavioral Sciences and the Law*, 21:563–580.

Maloney, L. T. and Mamassian, P. (2009). Bayesian decision theory as a model of human visual perception: Testing Bayesian transfer. *Visual Neuroscience*, 26:147–155.

Mandelbaum, E. (2019). Troubles with Bayesianism: An introduction to the psychological immune system. *Mind and Language*, 34(2):141–157.

Marr, D. (1982). *Vision*. W.H. Freeman and Company.

Marr, D. and Poggio, T. (1976). From understanding computation to understanding neural circuitry. MIT A.I. Memo 357.

Mattsson, L.-G. and Weibull, J. W. (2002). Probabilistic choice and procedurally bounded rationality. *Games and Economic Behavior*, 41:61–78.

Matějka, F. and McKay, A. (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*, 105(1):272–298.

Maynard Smith, J. and Price, G. (1973). The logic of animal conflict. *Nature*, 146:15–18.

McClennen, E. (1990). *Rationality and Dynamic Choice*. Cambridge University Press.

McCulloch, W. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 7:115–133.

McDonnell, M. D. and Abbott, D. (2009). What is stochastic resonance? Definitions, misconceptions, debates, and its relevance to biology. *PLoS Computational Biology*, 5(5):1–9.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In Zarembka, P., editor, *Frontiers in Econometrics*, pages 105–142. Academic Press.

McNamee, D. and Wolpert, D. M. (2019). Internal models in biological control. *Annual Review of Control, Robotics, and Autonomous Systems*, 2:339–364.

Meek, C. and Glymour, C. (1994). Conditioning and intervening. *The British Journal for the Philosophy of Science*, 45:1001–1021.

Megiddo, N. (1994). On probabilistic machines, bounded rationality, and computational complexity. In Meggido, N., editor, *Essays in Game Theory: In Honor of Michael Maschler*. New York: Springer Verlag.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092.

Mierzewski, K. (2022). Probabilistic stability, AGM revision operators and maximum entropy. *The Review of Symbolic Logic*, 15(3):553–590.

Miller, G. A., Galanter, E., and Pribram, K. H. (1960). *Plans and the Structure of Behavior*. Henry Holt & Co.

Millgram, E. (1991). Harman's hardness arguments. *Pacific Philosophical Quarterly*, 72(3):181–202.

Millgram, E. (2009). *Ethics Done Right: Practical Reasoning as a Foundation for Moral Theory*. Cambridge University Press.

Milli, S., Lieder, F., and Griffiths, T. L. (2021). A rational reinterpretation of dual-process theories. *Cognition*, 217:104881.

Millikan, R. G. (1989). Biosemantics. *Journal of Philosophy*, 86:281–297.

Millikan, R. G. (2023). Teleosemantics and the frogs. *Mind & Language*. Forthcoming.

Mills, C. W. (2005). "Ideal theory" as ideology. *Hypatia*, 20(3):165–184.

Minsky, M. (1967). *Computation: Finite and Infinite Machines*. Prentice Hall.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., and Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, 518:529–533.

Momennejad, I., Russek, E. M., Cheong, J. H., Botvinick, M. M., Daw, N. D., and Gershman, S. J. (2017). The successor representation in human reinforcement learning. *Nature Human Behavior*, 1:680–692.

Mongin, P. (2000). Does optimization imply rationality? *Synthese*, 124(1):73–111.

Morris, A., Phillips, J., and Cushman, F. (2019). How we know what not to think. *Trends in Cognitive Sciences*, 23(12):1026–1040.

Morton, J. M. (2017). Reasoning under scarcity. *Australasian Journal of Philosophy*, 95(3):543–559.

Morton, J. M. and Paul, S. K. (2019). Grit. *Ethics*, 129(2):175–203.

Murphy, G. L. (1993). A rational theory of concepts. *The Psychology of Learning and Motivation*, 29:327–359.

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press.

Musslick, S. and Cohen, J. D. (2021). Rationalizing constraints on the capacity for cognitive control. *Trends in Cognitive Sciences*, 25(9):757–775.

Nachbar, J. and Zame, W. R. (1996). Non-computable strategies and discounted repeated games. *Economic Theory*, 8:103–122.

Namkoong, H., Keramati, R., Yadlowsky, S., and Brunskill, E. (2020). Off-policy policy evaluation for sequential decisions under unobserved confounding. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H., editors, *Advances in Neural Information Processing Systems*, volume 33, pages 18819–18831.

Narens, L. and Skyrms, B. (2020). *The Pursuit of Happiness: Philosophical and Psychological Foundations of Utility*. Oxford University Press.

Nash, J. (1950). Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences*, 36(1):48–49.

von Neumann, J. (1966). *Theory of Self-Reproducing Automata*. University of Illinois Press.

von Neumann, J. and Morgenstern, O. (1953). *Theory of Games and Economic Behavior*. Princeton University Press.

Newell, A., Shaw, J. C., and Simon, H. A. (1959). Report on a general problem-solving program. In *Proceedings of the International Conference on Information Processing*, pages 256–264.

Neyman, A. (1985). Bounded complexity justifies cooperation in the finitely repeated prisoners' dilemma. *Economic Letters*, 19(3):227–229.

Nguyen, C. T. (2022). Trust as an unquestioning attitude. In Gendler, T. S., Hawthorne, J., and Chung, J., editors, *Oxford Studies in Epistemology*, volume 7, pages 214–244. Oxford University Press.

Nisbett, R. E. and Borgida, E. (1975). Attribution and the psychology of prediction. *Journal of Personality and Social Psychology*, 32(5):932–943.

Niven, J. E. and Laughlin, S. B. (2008). Energy limitation as a selective pressure on the evolution of sensory systems. *The Journal of Experimental Biology*, 211:1792–1804.

Nobandegani, A. S. (2017). *The Minimalist Mind: On Minimality in Learning, Reasoning, Action, and Imagination*. PhD thesis, McGill University.

Nozick, R. (1977). On Austrian methodology. *Synthese*, 36(3):353–392.

Nozick, R. (1993). *The Nature of Rationality*. Princeton University Press.

Oaksford, M. and Chater, N., editors (1999). *Rational Models of Cognition*. Oxford University Press.

Okasha, S. (2016). On the interpretation of decision theory. *Economics and Philosophy*, 32(3):409–433.

Okasha, S. (2018). *Agents and Goals in Evolution*. Oxford University Press.

O'Neill, O. (1987). Abstraction, idealization and ideology in ethics. In Evans, J. D. G., editor, *Moral philosophy and Contemporary Problems*, pages 55–69. Cambridge University Press.

Orbán, G., Berkes, P., Fiser, J., and Lengyel, M. (2016). Neural variability and sampling-based probabilistic representations in the visual cortex. *Neuron*, 92:530–543.

Orr, H. A. (2007). Absolute fitness, relative fitness, and utility. *Evolution*, 61(12):2997–3000.

Ortega, P. A. (2010). *A Unified Framework for Resource-Bounded Autonomous Agents Interacting with Unknown Environments*. PhD thesis, University of Cambridge.

Ortega, P. A. and Braun, D. A. (2013). Thermodynamics as a theory of decision-making with information-processing costs. *Proceedings of the Royal Society of London, A*, 469(2153).

Osband, I. and Van Roy, B. (2017). Why is posterior sampling better than optimism for reinforcement learning? In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, pages 2701–2710.

Papadimitriou, C. H. and Tsitsiklis, J. N. (1987). The complexity of Markov decision processes. *Mathematics of Operations Research*, 12(3):441–450.

Paul, L. A. (2014). *Transformative Experience*. Oxford University Press.

Peacocke, C. (1986). Explanation in computational psychology: Language, perception and level 1.5. *Mind & Language*, 1(2):101–123.

Peterson, C. R. and Beach, L. R. (1967). Man as an intuitive statistician. *Psychological Bulletin*, 68(1):29–46.

Peterson, M., editor (2015). *The Prisoner's Dilemma*. Cambridge University Press.

Pettigrew, R. (2019). *Choosing for Changing Selves*. Oxford University Press.

Pettit, P. (1984). Satisficing consequentialism. *Proceedings of the Aristotelian Society*, 58:165–176.

Phillips, J. and Cushman, F. (2017). Morality constrains the default representation of what is possible. *Proceedings of the National Academy of Science*, 114(18):4649–4654.

Phillips, J., Luguri, J. B., and Knobe, J. (2015). Unifying morality's influence on non-moral judgments: The relevance of alternative possibilities. *Cognition*, 145:30–42.

Piantadosi, S. T., Tenenbaum, J. B., and Goodman, N. D. (2016). The logical primitives of thought: Empirical foundations for compositional cognitive models. *Psychological Review*, 123(4):392–424.

Piccione, M. and Rubinstein, A. (1997). On the interpretation of decision problems with imperfect recall. *Games and Economic Behavior*, 20(1):3–24.

Pilgrim, C., Sanborn, A., Malthouse, E., and Hills, T. T. (2022). Confirmation bias emerges from an approximation to Bayesian reasoning. 10.31234/osf.io/jzct8.

Pollock, J. (2006). *Thinking About Acting: Logical Foundations for Rational Decision Making*. Oxford University Press.

Posner, M. I. and Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, 77(3):353–363.

Proust, J. (2013). *The Philosophy of Metacognition: Mental Agency and Self-Awareness*. Oxford University Press.

Putnam, H. (1967). Psychological predicates. In Capitan, W. H. and Merrill, D. D., editors, *Art, Mind, and Religion*. Pittsburgh University Press.

Pylyshyn, Z. W. (1984). *Computation and Cognition*. MIT Press.

Pylyshyn, Z. W., editor (1987). *The Robot's Dilemma*. Ablex.

Rahnev, D., Block, N., Denison, R. N., and Jehee, J. (2021). Is perception probabilistic? Clarifying the definitions. PsyArXiv. doi:10.31234/osf.io/f8v5r.

Ramsey, F. P. (1928). A mathematical theory of saving. *Economic Journal*, 38(4):543–559.

Ramsey, F. P. (1931). Truth and probability. In Braithwaite, R. B., editor, *Foundations of Mathematics and Other Logical Essays*. Martino Fine.

Rawls, J. (1971). *A Theory of Justice*. Belknap.

Rescorla, M. (2015). Bayesian perceptual psychology. In Matthen, M., editor, *The Oxford Handbook of the Philosophy of Perception*, pages 694–716. Oxford University Press.

Rescorla, M. (2020). A realist perspective on Bayesian cognitive science. In Nes, A. and Chan, T., editors, *Inference and Consciousness*, pages 40–73. Routledge.

Rescorla, M. (2021). On the proper formulation of conditionalization. *Synthese*, 198(3):1935–1965.

Rich, P., Blokpoel, M., de Haan, R., and van Rooij, I. (2020). How intractability spans the cognitive and evolutionary levels of explanation. *Topics in Cognitive Science*, 12(4):1382–1402.

Robbins, L. (1934). Remarks upon certain aspects of the theory of costs. *Economic Journal*, 44(173):1–18.

Rogers, T. T. and McClelland, J. L. (2006). *Semantic Cognition: A Parallel Distributed Processing Approach*. MIT Press.

van Rooij, I., Blokpoel, M., Kwisthout, J., and Wareham, T. (2019). *Cognition and Intractability*. Cambridge University Press.

Ross, J. and Schroeder, M. (2014). Belief, credence, and pragmatic encroachment. *Philosophy and Phenomenological Research*, 88(2):259–288.

Rubinstein, A. (1998). *Models of Bounded Rationality*. MIT Press.

de Rugy, A., Loeb, G. E., and Carroll, T. J. (2012). Muscle coordination is habitual rather than optimal. *Journal of Neuroscience*, 32(21):7384–7391.

Rule, J. S., Tenenbaum, J. B., and Piantadosi, S. T. (2020). The child as hacker. *Trends in Cognitive Science*, 24(11):900–915.

Rumelhart, D. E., McClelland, J. L., and The PDP Research Group (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. MIT Press.

Russell, D. F., Wilkens, L. A., and Moss, F. (1999). Use of behavioural stochastic resonance by paddle fish for feeding. *Nature*, 402:291–294.

Russell, S. and Subramanian, D. (1995). Provably bounded-optimal agents. *Journal of Artificial Intelligence Research*, 2:1–36.

Ryle, G. (1949). *The Concept of Mind*. University of Chicago Press.

Samuels, R., Stich, S., and Bishop, M. (2002). Ending the rationality wars: How to make disputes about human rationality disappear. In Elio, R., editor, *Common Sense, Reasoning and Rationality*, pages 236–268. Oxford University Press.

Samuelson, P. A. (1971). The "fallacy" of maximizing the geometric mean in long sequences of investing or gambling. *Proceedings of the National Academy of Sciences*, 68(10):2493–2496.

Sanborn, A. N. and Chater, N. (2016). Bayesian brains without probabilities. *Trends in Cognitive Sciences*, 20(12):883–893.

Sanborn, A. N., Griffiths, T. L., and Navarro, D. J. (2010). Rational approximations to rational models: Alternative algorithms for category learning. *Psychological Review*, 117(4):1144–1167.

Satz, D. and Ferejohn, J. (1994). Rational choice and social theory. *Journal of Philosophy*, 91(2):71–87.

Savage, L. J. (1954). *Foundations of Statistics*. Dover, 1972, 2nd revised edition.

Scanlon, T. M. (1998). *What We Owe to Each Other*. Harvard University Press.

Schoemaker, P. J. H. (1991). The quest for optimality: A positive heuristic of science? *Behavioral and Brain Sciences*, 14(2):205–215.

Schrimpf, M., Kubilius, J., Hong, H., Majaj, N. J., Rajalingham, R., Issa, E. B., Kar, K., Bashivan, P., Prescott-Roy, J., Geiger, F., Schmidt, K., Yamins, D. L. K., and DiCarlo, J. J. (2020). Brain-score: Which artificial neural network for object recognition is most brain-like? *bioRxiv*.

Schultz, W. (2016). Dopamine reward prediction-error signalling: A two-component response. *Nature Reviews Neuroscience*, 17:183–195.

Schulz, A. W. (2018). *Efficient Cognition: The Evolution of Representational Decision Making*. MIT Press.

Schulz, E. and Gershman, S. J. (2019). The algorithmic architecture of exploration in the human brain. *Current Opions in Neurobiology*, 55:7–14.

Schut, M., Wooldridge, M., and Parsons, S. (2004). The theory and practice of intention reconsideration. *Journal of Experimental & Theoretical Artificial Intelligence*, 16(4).

Schwitzgebel, E. (2010). Acting contrary to our professed beliefs or the gulf between occurrent judgment and dispositional belief. *Pacific Philosophical Quarterly*, 91(4):531–553.

Scott, S. H. (2012). The computational and neural basis of voluntary motor control and planning. *Trends in Cognitive Science*, 16(11):541–549.

Sen, A. K. (1977). Rational fools: A critique of the behavioral foundations of economic theory. *Philosophy & Public Affairs*, 6(4):317–344.

Shafto, P., Kemp, C., Mansinghka, V., and Tenenbaum, J. B. (2011). A probabilistic model of cross-categorization. *Cognition*, 120(1):1–25.

Shalizi, C. R. and Crutchfield, J. P. (2001). Computational mechanics: Pattern and prediction, structure and simplicity. *Journal of Statistical Physics*, 104(3-4):817–879.

Shannon, C. (1948). A mathematical theory of information. *Bell System Technical Journal*, 27:379–423.

Shea, N. (2018). *Representation in Cognitive Science*. Oxford University Press.

Shea, N., Boldt, A., Bang, D., Yeung, N., Heyes, C., and Frith, C. D. (2014). Supra-personal cognitive control and metacognition. *Trends in Cognitive Sciences*, 18(4):186–93.

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., and Botvinick, M. M. (2017). Toward a rational and mechanistic account of mental effort. *Annual Review of Neuroscience*, 40:99–124.

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T., Simonyan, K., and Hassabis, D. (2018). A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science*, 362(6419):1140–1144.

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

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, 63(2):129–138.

Simon, H. A. (1975). The functional equivalence of problem solving skills. *Cognitive Psychology*, 7:268–288.

Simon, H. A. (1976). From substantive to procedural rationality. In Kastelein, T. J., Kuipers, S. K., Nijenhuis, W. A., and Wagenaar, G. R., editors, *25 Years of Economic Theory*, pages 65–86. Springer.

Simon, H. A. (1983). *Reason in Human Affairs*. Stanford University Press.

Simon, H. A. (1991). Cognitive architectures and rational analysis: Comment. In VanLehn, K., editor, *Architectures for Intelligence: The 22nd Carnegie Mellon Symposium on Cognition*, pages 25–39. Lawrence Earlbaum Associates, Inc.

Sims, C. R. (2016). Rate-distortion theory and human perception. *Cognition*, 152:181–198.

Skyrms, B. (1980). *Causal Necessity: An Pragmatic Investigation of the Necessity of Laws*. Yale University Press.

Skyrms, B. (1990). *The Dynamics of Rational Deliberation*. Harvard University Press.

Slovic, P., Fischhoff, B., and Lichtenstein, S. (1977). Behavioral decision theory. *Annual Review of Psychology*, 28:1–39.

Smart, J. J. C. (1956). Extreme and restricted utilitarianism. *Philosophical Quarterly*, 6(25):344–354.

Sober, E. (1983). Equilibrium explanation. *Philosophical Studies*, 43:201–210.

Sober, E. (2001). The two faces of fitness. In Singh, R. S., Krimbas, C. B., Paul, D. B., and Beatty, J., editors, *Thinking about Evolution*, volume 2, pages 309–321. Cambridge University Press.

Staffel, J. (2017). Should I pretend I'm perfect? *Res Philosophica*, 94(2):301–324.

Staffel, J. (2019). How do beliefs simplify reasoning? *Noûs*, 53:937–962.

Stalnaker, R. C. (1984). *Inquiry*. MIT Press.

Stalnaker, R. C. (1991). The problem of logical omniscience, I. *Synthese*, 89(3):425–440.

Stalnaker, R. C. (2010). *Our Knowledge of the Internal World*. Oxford University Press.

Steele, K. S. (2010). What are the minimal requirements of rational choice?: Arguments from the sequential-decision setting. *Theory and Decision*, 68(4):463–487.

Steinert-Threlkeld, S. (2021). Quantifiers in natural language: Efficient communication and degrees of semantic universals. *Entropy*, 23:1335.

Sterelny, K. (2003). *Thought in a Hostile World: The Evolution of Human Cognition*. Blackwell.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103(2684):677–680.

Stewart, N., Chater, N., and Brown, G. D. (2006). Decision by sampling. *Cognitive Psychology*, 53:1–26.

Stich, S. P. (1990). *The Fragmentation of Reason: Preface to a Pragmatic Cognitive Evaluation*. MIT Press.

Still, S. and Precup, D. (2012). An information-theoretic approach to curiosity-driven reinforcement learning. *Theory in Biosciences*, 131:139–148.

Strehl, A. L., Li, L., Wiewiora, E., Langford, J., and Littman, M. L. (2006). PAC model-free reinforcement learning. In *Proceedings of the 23rd international conference on Machine learning (ICML)*, page 881–888.

Sutton, R. S. and Barto, A. G. (1998). *Reinforcement Learning*. MIT Press.

Sutton, R. S., Precup, D., and Singh, S. (1999). Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence*, 112(1):181–211.

Taber, C. S. and Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, 50(3):755–769.

Tanner, W. P. and Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review*, 61(6).

Thobani, I. (2023). A triviality worry for the Internal Model Principle. Manuscript, Stanford University.

Thoma, J. (2018). Temptation and preference-based instrumental rationality. In Bermúdez, J. L., editor, *Self-Control, Decision Theory, and Rationality: New Essays*, page 27–47. Cambridge University Press.

Thoma, J. (2019). Risk aversion and the long run. *Ethics*, 129:230–253.

Thoma, J. (2021). Folk psychology and the interpretation of decision theory. *Ergo*, 7(34):904–936.

Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294.

Thorndike, E. L. (1911). *Animal Intelligence: Experimental Studies*. Macmillan Press.

Thorstad, D. (2023). Why bounded rationality (in epistemology)? *Philosophy and Phenomenological Research*. Forthcoming.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, 34(4):273–286.

Tishby, N. and Polani, D. (2011). Information theory of decisions and actions. In Cutsuridis, V., Hussain, A., and Taylor, J. G., editors, *Perception-Action Cycle*, pages 601–636. Springer.

Todorov, E. (2007). Optimal control theory. In Doya, K., Ishii, S., Pouget, A., and Rao, R. P. N., editors, *Bayesian Brain: Probabilistic Approaches to Neural Coding*, pages 268–298. MIT Press.

Todorov, E. and Jordan, M. I. (2002). Optimal feedback control as a theory of motor coordination. *Nature Neuroscience*, 5(11):1226–1235.

Turing, A. M. (1936). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, s2-42:230–265.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59:433–460.

Tversky, A. (1969). Intransitivity of preferences. *Psychological Review*, 76(1):31–48.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131.

Tversky, A. and Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review*, 101(4):547–567.

Ullmann-Margalit, E. (2006). Big decisions: Opting, converting, drifting. *Royal Institute of Philosophy Supplements*, 58:157–172.

Veblen, T. (1898). Why is economics not an evolutionary science? *The Quarterly Journal of Economics*, 12:373–397.

Vredenburgh, K. (2020). A unificationist defence of revealed preferences. *Economics & Philosophy*, 36(1):149–169.

Vul, E. (2010). *Sampling in Human Cognition*. PhD thesis, MIT.

Vul, E., Goodman, N. D., Griffiths, T. L., and Tenenbaum, J. B. (2014). One and done? Optimal decisions from very few samples. *Cognitive Science*, 38(4):699–637.

Vul, E. and Pashler, H. (2008). Measuring the crowd within. *Psychological Science*, 19(7):645–647.

Vulcan, N. (2000). An economist's perspective on probability matching. *Journal of Economic Surveys*, 13(1):101–118.

Wedgwood, R. (2002). Internalism explained. *Philosophy and Phenomenological Research*, 65(2):349–369.

Weirich, P. (2004). *Realistic Decision Theory: Rules for Nonideal Agents in Nonideal Circumstances*. Oxford University Press.

Weisberg, J. (2020). Belief in psyontology. *Philosophers' Imprint*, 20(11):1–27.

Whitehead, A. N. (1911). *An Introduction to Mathematics*. Holt.

Wilson, A. (2014). Bounded memory and biases in information processing. *Econometrica*, 82(6):2257–2294.

Wimsatt, W. C. (2007). *Re-Engineering Philosophy for Limited Beings: Piecewise Approximations to Reality*. Harvard University Press.

Winther, R. G. (2014). James and Dewey on abstraction. *The Pluralist*, 9(2):1–28.

Yamins, D. L. K. and DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortext. *Nature Neuroscience*, 19(3):356–365.

Yamins, D. L. K., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., and DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, 111(23):8619–8624.

Zaffora Blando, F. (2022). Bayesian merging of opinions and algorithmic randomness. *The British Journal for the Philosophy of Science*. Forthcoming.

Zaslavsky, N., Hu, J., and Levy, R. P. (2020). A rate-distortion view of human pragmatic reasoning.

Zednik, C. and Jäkel, F. (2016). Bayesian reverse-engineering considered as a research strategy for cognitive science. *Synthese*, 193(12):3951–3985.

van Zee, M. and Icard, T. (2015). Intention reconsideration as metareasoning. In *NeurIPS Workshop on Bounded Optimality and Rational Metareasoning*.

Zermelo, E. (1913). Über eine Anwendung der Mengenlehre auf die Theorie des Schachspiels. In *Proceedings of the Fifth Congress Mathematicians*, page 501–504. Cambridge University Press.

Zhi-Xuan, T., Mann, J., Silver, T., Tenenbaum, J., and Mansinghka, V. (2020). Online Bayesian goal inference for boundedly rational planning agents. In *Advances in Neural Information Processing Systems*, volume 33.

Zollman, K. J. (2023). On the normative status of mixed strategies. In Augustin, T., Cozman, F. G., and Wheeler, G., editors, *Reflections on the Foundations of Probability and Statistics: Essays in Honor of Teddy Seidenfeld*, pages 207–240. Springer.