# Complexity and Its Measurement*

Ryan Oprea†

October 8, 2024

**Abstract**

In this essay, I propose an operationalization of the word "complexity" for the social sciences, drawing on ideas from computer science and cognitive science. I show that this operationalization predicts a series of behavioral responses to complexity, and argue that we can use these predicted regularities to measure complexity and empirically detect its influence on human behavior. Motivated by these behavioral predictions, I review a number of empirical techniques that have been used in the experimental literature to measure and detect complexity. I explain how these techniques have been used to empirically isolate the influence of complexity from other motivational, epistemic and cognitive drivers of behavior. Finally, I review a recent literature that attempts to measure complexity by studying what characteristics of tasks and algorithms predict behavioral signatures of complexity.

In this essay, I motivate and review empirical methods for measuring and detecting the influence of *complexity* on human behavior. There is growing evidence in the behavioral sciences that complexity is a fundamental driver of human behavior, shaping how humans process information, reason through problems, compute quantities, value options, trade off opportunities, evaluate policies and comply with and understand social rules. Although few social scientists would disagree that "complexity" is an important influence on behavior, it is a term that has often been only hazily defined in the social science literature and, as a result, has often been used casually and inconsistently when framing empirical work. I will argue that complexity can be given a reasonably crisp general definition and that this definition can allow us to measure and detect the influence of complexity and use it as an explanatory variable in a principled way.

I begin in Section 1 by defining complexity, loosely adapting the operationalization used in computer science: "complexity" is a name for the resource cost of using a *procedure* (i.e. an

*I am grateful to Alex Rees-Jones for suggesting that I write this essay and for his tireless support and insightful comments throughout the process of writing it. I am also grateful to Daniel Bartels, Peter Bossaerts, Benjamin Enke, Thomas Graeber, Alex Imas, Kirby Nielsen, Adam Sanjurjo and Alex Whitefield for incredibly helpful, constructive comments on earlier drafts. This research was supported by the National Science Foundation under Grant SES-1949366.

†Economics Department, University of California, Santa Barbara, Santa Barbara, CA, 93106, roprea@gmail.com.

algorithm or cognitive process) to process information when solving a problem or making a decision. For instance, in order to reason through a problem, aggregate sources of information, trade off determinants of value etc. a decision maker has to expend scarce time and energy, tie up limited working memory and attention and put forth subjectively unpleasant effort. These expenses – of time, energy, memory, attention, subjective well-being etc. – are what we mean when we say that a procedure is complex. Building on this procedural definition, a *task* is complex if the *procedure required to correctly or optimally solve that task* is complex: to whatever degree the information processing procedure required to properly perform the task is expensive in terms of time, memory, attention, effort etc., we say the task itself is complex.

There are two immediate implications of this definition that I think are useful for shaping the way we think about measurement. First, complexity is neither a perfectly subjective notion (e.g., a description of the capacities of the decision-maker) nor a perfectly objective notion (e.g., a description of the properties of the task), but is instead a "between-ness" notion that focuses attention on the subjective costs of information processing induced by the objective features of the task. This means that complexity is *both* a way of describing the difficulty of optimally solving a *task* (which we will call "task complexity"), and a way of describing the costs of what a decision maker actually does when attempting to solve it ("procedural complexity"), which need not be optimal. Confusion between these two uses of the term – i.e. the complexity of the procedure required to *properly* approach a task vs. the complexity of the procedure *actually* used to approach the task – can lead to serious errors in thinking about the measurement of complexity. Second, depending on the context, complexity can be (and often is) measured in terms of a number of distinct cognitive resources including time, memory space, attention and costly effort. This gives rise to variations on the notion such as time complexity, space complexity, cost complexity etc. which are measurable in different ways and sometimes produce different predicted effects (in Section 1.3 I organize many of these variations as a lifeline for the confused).

In Section 1.1, I sketch a simple descriptive model, that describes the tradeoffs between the benefits that accrue from using an optimal procedure to tackle a problem, and the complexity costs of using such a procedure. I use this simple framework in Section 1.2 to describe the key implications of complexity for interpeting and explaining human behavior. Definitionally, complexity produces decision costs, which are both (i) of direct importance for the social scientist (e.g., because they impact welfare) and (ii) sometimes can be directly measured. Less directly, complexity distorts how people go about solving problems: if the subjective costs of using an *optimal* procedure to solve a task is too high (if the task is too complex), the decision-maker may be tempted or forced to use a simpler, suboptimal procedure instead. Because of this, complexity can cause people to make serious mistakes for perfectly "rational" reasons.[1] Similarly, complexity can distort behavior

[1] Throughout this essay, when I refer to a choice as "optimal" I mean that it is optimal according to a rational choice model that ignores the cognitive costs of decision making altogether; when I refer to a choice as a "mistake" I mean that it is a mistake relative to an optimal choice defined in this way. This usage may seem odd given that one of the key implications of complexity is that, once the costs and constraints of decision-making are taken into

by inducing people to *avoid* complexity, either by systematically avoiding complex options or information within decisions, or by self-selecting out of complex decisions altogether. These efforts to avoid complexity often produce distinctive patterns of *adaptive* behavior – distinctive responses to seemingly superficial features of the behavioral environment that are hard to account for in the absence of the distorting influence of complexity.

In Section 2, I use these behavioral effects of complexity to motivate a number of empirical techniques the literature has developed to measure and detect its influence on behavior. First, in Section 2.1 I review a series of *metrics* of complexity: measurable empirical outcomes that quantiatively or qualitatively measure the influence of complexity. One class of metrics (Section 2.1.1) aims to measure complexity *directly* by measuring the resource costs including willingness-to-pay to avoid a task, response times in a task and biometric measures of biological responses to tasks. I emphasize the complications involved in interpreting these kinds of measures. Another class of metrics (Section 2.1.2) aims instead to measure *behavioral responses* to complexity, including the use of obviously non-optimal procedures or clear evidence of outright mistakes. A third type of complexity metric (Section 2.1.3) relies on subjects' elicited beliefs about the optimality of their own choices or about the relative complexity of tasks.

Next (Section 2.2), I review experimental treatment manipulations that can be used to test for the influence of complexity on a behavior. One type of intervention (Section 2.2.1) manipulates features of the task in ways that only influence information processing costs, and examines whether this changes behavior – evidence that the original behavior was influenced by complexity. Another (Section 2.2.2) attempts instead to manipulate the cognitive resources available to the decision maker either by artificially constraining the subject's working memory or time available for the task or by giving the subject decision aids that automate processing; to the extent such manipulations reduce or intensify a behavior we have some evidence that behavior was a response to complexity. Finally, manipulations to the rewards of the task (Section 2.2.3) can sometimes be used for the same purpose: to the degree a behavior disappears when rewards increase, we have some evidence that that behavior was rooted in an effort to economize on complexity costs.

In Sections 3 and 4, I put these measurement approaches to work, illustrating how they have been used to study complexity and its effects in the recent literature. In Section 3, I examine the use of these techniques for simply *detecting* the influence of complexity and highlight that this requires the researcher to distinguish complexity from other competing explanatory forces including

account, behavior that looks suboptimal might in fact be optimal and behavior that looks like a mistake may be no mistake at all. Indeed, many important literatures and conceptual frameworks related to complexity including bounded rationality (e.g. Simon 1955), resource rationality (e.g. Lieder & Griffiths 2020), ecological rationality (e.g. Todd & Gigerenzer 2012) and adaptive decision making (e.g. Payne et al. 1993) emphasize this very point. In using these terms therefore it is important to emphasize that I am not questioning the broader rationality of the heuristic procedures decision makers use, or their effectiveness in economizing on decision-making costs. Rather I use these words as a shorthand for the natural benchmark against which it is typically most useful to judge the effects of complexity in the social sciences.

preferences and beliefs subjects bring with them from outside of the lab. In Section 3.1 I illustrate how a range of techniques from Section 2 have been used to distinguish complexity responses from expressions of subjects' true preferences over outcomes. In Section 3.2 I discuss some design principles for preventing uncontrolled beliefs from confounding the measurement of complexity and in Section 3.3 I briefly discuss the relationship between complexity and other, related cognitive frictions.

Finally, in Section 4.2 I discuss the more ambitious enterprise of using these methods not merely to detect but to "map" complexity – to identify general regularities that predict what generates complexity responses in humans. I group this literature into two branches of research. One, discussed in Section 4.1 attempts to to map complexity at the task level, by measuring which coarse task characteristics tend to predictably generate evidence of complexity. The goal is to identify regularities that allow us to anticipate when a task is likely to be complex, using descriptive features of the task rather than the procedure required to solve it. The other branch, discussed in Section 4.2, attempts instead to describe the sources of complexity at the deeper *procedural* level, by measuring what characteristics make one algorithm or choice procedure more complex than another, generically. This literature starts by choosing a formal language for describing algorithms (e.g., finite state machines, Turing machines, elementary information processing units) and then measuring how metrics of those formal descriptions predict variation in evidence of complexity in behavior. By doing this, the literature hopes to predict in a domain-general way what makes tasks, institutions and formal rules complex based on the decision procedures those tasks, institutions and formal rules require of the decision maker. In Section 4.2.4 I discuss related efforts to study human behavior under the lens of computational complexity theory from computer science.

Before launching into the substantive material, two caveats about this essay are in order. First and most importantly, it bears emphasizing that this is a methodological essay, focused on describing approaches to operationalization and measurement and illustrating using some salient examples. It is *not* an overview of the relevant literature and the papers I've selected to illustrate these methods are neither exhaustive nor balanced. Indeed, my selection is biased both in terms of availability (I am a behavioral economist, so my examples are oversampled from that field) and recency (I have deliberately biased my selection of examples to some extent towards recent work). This incompleteness was unavoidable given space constraints and the vast multi-decade and multi-field enterprise, spanning cognitive science, neuroscience, psychology, computer science and behavioral economics, that is relevant to complexity as I've defined it. To misread this essay's sparse collection of examples as a systematic overview of this rich body of research would do a great injustice to the literature. Second, because this essay is focused on methodological issues related to measurement, I will for the most part be forced to ignore many important theoretical and conceptual lines of discussion that are highly relevant to complexity – and with them many important empirical results related to these discussions. For instance, I will not for the most part review a rich, long-running line of research in cognitive science applying and adapting notions of computational complexity to the modeling of human cognition (e.g. van Rooij 2008, van Rooij et al. 2019). I will likewise be forced to largely ignore highly relevant work on cognitive architectures from cognitive psychology (e.g. Kotseruba & Tsotsos 2020), work modeling the bases of subjective effort costs from cognitive psychology and neuroscience (e.g. Kurzban et al. 2013, Shenhav et al. 2017, Kool & Botvinick 2018) and numerous proposed frameworks describing the consequences of imperfect information processing (e.g. Stewart et al. 2006) from psychology and cognitive science. Finally, I will only be able to superficially discuss a rich, emerging literature in behavioral economics modeling responses to complexity, examining topics like rational inattention (e.g. Sims 2003, Maćkowiak et al. 2023), salience (e.g. Bordalo et al. 2012, 2022), representational simplification (e.g. Gabaix 2014), noisy cognition (e.g. Woodford 2020) and aversion to thinking (e.g. Ortoleva 2013, Alaoui & Penta 2022).[2] Needless to say, readers interested in complexity are encouraged to thoroughly explore this rich conceptual terrain.

# 1 What Is Complexity?

In standard usage, "complexity" refers to the difficulty of solving a problem, making a correct decision or properly analyzing information.[3] To operationalize the term, we can hardly do better than to look to computer science which has by far the most developed body of research studying complexity and the sharpest formal definitions of the word. There, "complexity" is defined simply as the resource burden, usually denominated in runtime or memory, on a computational device of implementing an algorithm (e.g. Arora & Barak 2009). Thus, complexity is simply a word to describe the cost to a computational device of utilizing a procedure to solve a problem. In computer science, a problem (e.g., an inference, a decision, a computation) is "complex" to whatever extent the simplest (i.e., least costly) algorithm capable of correctly solving the problem is complex.[4]

This definition suggests a tidy operationalization of the term for the behavioral sciences: *complexity* is the cost, suffered by a decision maker (DM), when implementing a *procedure* (a rule, algorithm or cognitive process) to solve a problem. A choice or inference problem, which we will usually call a "task," is *complex* to whatever degree solving it correctly requires the DM to implement a costly procedure. This definition is built on the recognition that human behaviors, no less than those of a computer, are shaped by a"procedural layer" that lies between the descriptive primitives of a problem and a behavioral response: in order to come to a decision or form a belief, a human decision maker must first process the information embedded in the problem and this information processing is often costly both subjectively (i.e. it is hedonically unpleasant) and

[2] See Enke (2024) for an overview of the recent "cognitive turn" in behavioral economics which reviews many of these models and relevant empirical findings in some depth.

[3] Merriam-Webster defines "complex" in its adjectival form as: "hard to separate, analyze, or solve."

[4] See Bossaerts & Murawski (2017) for an excellent, short introduction to the computer science approach to complexity, that focuses on its applicability to the behavioral sciences. See van Rooij (2008) and van Rooij et al. (2019) for in-depth discussions of the application of these ideas to cognitive science (the "tractable cognition thesis"), emphasizing some of the adaptations of standard computational complexity theory that have proved useful for understanding human cognition. We discuss this literature in more depth in Section 4.2.4.

objectively (i.e., it consumes constrained resources like working memory[5] and time).[6]

In operationalizing complexity this way we are taking a cue from computer science but there are several respects in which we will specialize our usage when applying the idea to humans relative to machines.

First, while the complexity of a task in computer science is usually a description of the costs of *implementing* an optimal procedure, in humans there is an additional cost to consider: humans, unlike computers, often must also *identify* an optimal procedure from a possibly large set of possibilities before implementing it. In some tasks the costs of finding an optimal procedure, which we will call the *complexity of identification*, can be as large or larger than the costs of actually executing an optimal procedure, which we call the *complexity of implementation*.[7] Thus, complexity (as we will use the term) includes both the costs of implementation and identification.[8]

Second, in computer science, complexity costs are typically denominated in terms of objective resources like time or working memory, giving rise to notions of "time complexity" (complexity measured in terms of the number of time-expending operations), or "space complexity" (complexity measured in terms of commitment of working memory). In humans, both of these notions of complexity continue to be relevant: in many contexts human performance will be constrained by limitations on time available for making a choice and sharp limitations on human working memory capacity, in addition to important constraints on cognitive control (Posner & Snyder 1975, Shiffrin & Schneider 1977) and attentional resources.[9] But human behavior is also influenced by the di-

[5]Working memory is, roughly, the cognitive system that stores pieces of information for the purpose of manipulation during problem solving. It is usually thought to be severely constrained in humans.

[6]This information processing can include both (i) the task of forming a *representation* or mental model of the problem (a set of cognitive symbols the mind will operate on when tackling the problem) and (ii) the task of transforming this set of symbols into a decision by procedurally manipulating them in order to "solve" the task. This has led some authors to emphasize a distinction between "representational complexity" and "computational complexity" when taxonomizing the cognitive difficulties of solving a task (Shenhav et al. 2017, Ba et al. 2024). As we will discuss more below, this is sometimes a useful distinction to make because one of the key ways decision-makers simplify information processing is to ignore or underweight relevant information in the problem, economizing on their use of mental resources like working memory by simplify their representation of the problem.

[7]On the other hand, in many important behavioral settings the complexity of identification is negligible: complying with a formal rule likr a law, regulation or agreement, or following a recipe can be very complex to implement, but the procedure is trivially simple to identify since it is a "given" component of the problem for the DM, just as an algorithm is for a computer.

[8]Some approaches model this identification process as the problem of solving a mental optimization problem that involves maximizing the value of computation or VOC, counterfactually reasoning directly about the tradeoffs involved before selecting a procedure to implement (Lieder et al. 2014). This approach thus conceives of the problem of identification as a second implementation problem, this time involving the implementation of a potentially costly meta-reasoning prcoedure. Other approaches emphasize that DMs typically have a "toolkit" (Payne et al. 1993, Gigerenzer & Selten 2001) of often-used procedures (e.g., heuristics) at the ready, that they match to the task possibly based on their usefulness in past choice context. See Rieskamp & Otto (2006) for an influential formalization of this approach that models the selection of a procedure as a problem of reinforcement learning.

[9]See Loewenstein & Wojtowicz (2024) for a recent in-depth discussion of the concept of *attention*, emphasizing its relevance for the social sciences, with a focus on economics. Loewenstein & Wojtowicz (2024) emphasize (i)

rect hedonic costs (i.e., the mental suffering or effort costs) required to implement computationally intensive procedures (Shenhav et al. 2017, Kool & Botvinick 2018). This hedonic preference for avoiding difficult procedures is a particularly important aspect of complexity in humans, since humans can choose whether to avoid the expenses involved in optimal information processing. In some settings decision makers may be perfectly *capable* of using optimal procedures (i.e., have available the objective resources like time and working memory needed to perform the optimal procedure), but would *prefer* not to because of the subjective costs (e.g., strain or suffering) required. In this sense, complexity, instead of simply overwhelming available cognitive resources, often produces an analogue to an economic cost that the DM must trade off against the potential rewards from making optimal choices when choosing a procedure. Because it is rooted in effort costs and bears strong analogies to economic costs, we will call this aspect of complexity "cost complexity."[10,11]

Finally, in computer science it is often natural to think of information processing algorithms as involving deterministic, well-specified rules. This needn't be true, particularly in humans, and so we will define procedures more broadly to include any process a decision maker might use to map the descriptive primitives of a problem into a behavior or belief. Procedures, as we will use the word, may be highly deliberate and they may be highly algorithmic or "rule-like," just like a typical algorithm implemented by a computer, but they need not be either. A procedure may flow from "gut instinct" rather than careful thought, and it may be highly noisy or even random rather than structured and predictable.[12] Likewise, while computers tend to implement pre-specified algorithms, humans may sometimes construct and adjust their algorithms throughout the process of decision-making and this too is consistent with the way we will use the word "procedure."[13]

This operationalization of complexity has the virtue that it is broad enough to allow us to

that attention is a constrained cognitive resource, following Kahneman (1973), but (ii) that it is also a composite of a number of constitutive cognitive functions like working memory and cognitive control, that we will often treat separately in this chapter. Because attention is a resource required for decision making and because it is constrained, much of the literature on attention in, e.g., economics, is also a literature on complexity as I define the term here.

[10]Recent work in cognitive psychology and neuroscience has hypothesized that aversive mental effort costs of this sort are rooted in *cognitive control* – the deployment of mental resources to over-ride default or automatic behavioral responses to stimuli (Kurzban et al. 2013, Shenhav et al. 2017, Kool & Botvinick 2018). Because the resources required of this kind of executive control are sharply constrained, they are phenomonologically represented to the decision maker as an aversive cost.

[11]Of course, these three types of complexity ("time complexity", "space complexity" and "cost complexity") are related to one another, and all three are related to costs and constraints associated with the use of attention. For instance, the opportunity cost of time, given limitations on the deployment of attention at any given moment in time, contribute to cost complexity as we've defined it. Likewise, attention is ultimately rooted in part in working memory (Loewenstein & Wojtowicz (2024)), and taxing either working memory or attention may produce hedonic costs.

[12]There is nothing that prevents an algorithm, even for a computer, from having random components and there are decision settings in which the optimal computationally simple algorithm is in fact noisy (Kalai & Solan 2003). What's more, the use of procedures that are noisy and inconsistent are likely to be sometimes cognitively cheaper than precise and consistent choices and therefore represent an important tool for economizing on the costs of decision making.

[13]See Aloui and Penta (2022) for a thoughtful discussion of how procedures and their costs can be understood and modeled in such cases as stopping problems.

apply the notion to a wide range of behavioral contexts and to consider – and potentially model and predict – how the notion is related across these contexts. For example: (i) to say an inference problem is complex is to say that it is costly to correctly interpret relevant information; (ii) to say a calculation is complex is to say that it is costly to perform the necessary mathematical operations; (iii) to say a valuation task is complex, is to say that it is costly to aggregate the descriptive elements of the problem to understand its value; (iv) to say an optimization task is complex is to say that finding an optimum is costly to find; (v) to say a system is complex is to say that it is costly to predict the system's behavior; (vi) to say that a formal rule or law is complex is to say that complying with it is costly, etc.

To summarize, "complexity" is a way of describing and operationalizing the familiar idea that (i) information processing is costly or constrained for decision-makers, and (ii) that these costs or constraints are shaped by the requirements of the task at hand. Its virtue relative to other closely related concepts in the literature is that it is explicitly a "betweeness" notion – a way of describing the *relationship* between (i) the cognitive demands of a task and (ii) the cognitive capacities of the decision-maker.[14] It links (i) and (ii) by situating costs in the sequence of cognitive acts (the "procedure") the decision-maker uses when solving a problem.[15] In particular, a *procedure* is said to be "complex" to whatever degree it is costly in terms of some cognitive resource, while a *task* is "complex" to whatever degree the simplest procedure for optimally processing its information is complex (costly). This linked, dual usage of the term – linking a notion of procedural complexity to a notion of task complexity – means we can, in principle, use the same set of concepts to (i) measure the average complexity of a task, (ii) empirically rank the average complexities of tasks and (iii) identify when people simplify a task or, conversely, make a task more complex than it really is. Because of its focus on the procedural link between the capacities of the decision maker and the difficulty of the task environment, it is a particularly useful approach for describing and measuring the influence of cognitive costs and constraints in the social sciences.

### 1.1 A Descriptive Model

To fix ideas and set up the problem of measurement, consider the following simple descriptive model. A decision maker (DM) is given some information about a task (the "primitives" of the task) and chooses an action or set of actions $a \in A$, where $A$ is a menu of actions available in the

[14]Another natural way of understanding how "complexity" relates to other related notions, is via Marr's celebrated taxonomy of levels of explanation in cognitive science (Marr 1982). Marr argues that cognition can be studied at (i) the computational level (in terms of the computations required of the task), (ii) the algorithmic level (describing the representations and procedures used in the computation) or (iii) the implementation level (describing the physical way the brain or computational hardware implements those algorithms). To describe and measure behavior under the lens of "complexity" is to study behavior at Marr's algorithmic level, and to focus attention on the resource costs that arise at that level of explanation.

[15]Because it is a betweenness notion, it is a measurement designed to describe the hinge in Herbert Simon's famous *behavioral scissors*: "Human rational behavior is shaped by a scissors whose two blades are the structure of task environments and the computational capabilities of the actor" (Simon 1990).

task. The DM's motivation for taking an action is described by a reward function $\pi : A \mapsto \mathbb{R}$, that maps her choice of action into some index of value (i.e., monetary payment or utility), where a higher value of $\pi(a)$ is more preferred by the DM. Thus, formally, the DM's problem is to choose an $a$ that maximizes $\pi(a)$, given the information available to her. The DM's optimal action, $a^*$, maximizes $\pi$, given the actions available in $A$.

In order to make *any* choice, optimal or not, a DM must choose and implement some *procedure* that transforms the primitives of the problem into a choice. To formalize this idea, assume the DM does not choose an action directly, but instead makes a prior choice about the rule or procedure, $p \in P$, she will use to arrive at an action $a$. Let $\alpha : P \mapsto A$ be a function that maps the DM's procedure into her eventual choice of an action. An optimal procedure $p^*$ is a procedure that produces an optimal action: $\alpha(p^*) = a^*$.

We define *complexity* as the cost of identifying and implementing a procedure. We summarize this with a cost function $c$ that maps the use of a procedures into a cost. One natural way of modeling this (as "cost complexity") is to denominate the costs hedonically, in terms of the motivating rewards of the choice problem, $\pi$, and to say the DM's choice problem is not to choose an action $a$ to maximize $\pi(a)$, but instead to choose a procedure that maximizes the rewards from the action it produces, net of the costs expended to implement or identify that procedure:

$$\max_p \quad \pi(\alpha(p)) - c(p) \tag{1}$$

For very complex tasks (or in very constrained choice contexts), these costs may be effectively infinite. For instance, a DM simply cannot perform 1000 addition problems in the space of 10 seconds, making it impossible (infinitely costly) to implement a procedure aimed at correctly solving 1000 addition problems in the time given. In some contexts it will therefore be more natural to describe the cost in terms of a sharply constrained resource like time, working memory, attention or cognitive control given a budget of that resource $R$, and to stipulate that the cost cannot exceed this constraint:

$$\begin{aligned} &\max_p \quad \pi(\alpha(p)) \\ &\text{s.t.} \quad c(p) \leq R \end{aligned} \tag{2}$$

Let $p^{**}$ be a procedure that solves (1) or (2). Clearly, because of the influence of complexity, $p^{**}$ need not coincide with $p^*$. For instance, in problems best described by (1), the net costs of $p^*$ may be high enough relative to its contribution to rewards to motivate the DM to choose a cognitively cheaper and possibly less accurate procedure instead. Or, in problems of type (2), the working memory or free attention available to the DM or the time allotted to the task may be insufficient to implement $p^*$, *forcing* the DM to choose a less resource-intensive procedure instead. Because of this, complexity costs can distort procedural choice and thus can induce behavior that fails to maximize $\pi$ (i.e., $\alpha(p^{**}) \neq a^*$).

We will say that a *procedure* $p$ is complex to whatever degree $c(p)$ is positive (i.e. the procedure produces costs), and more complex the larger $c(p)$ is. A *task* is complex to whatever degree $c(p^*)$ is positive (i.e. the procedure that generates an optimal action produces costs), and more complex the larger $c(p^*)$ is. We will follow this terminology (and often refer respectively to "procedural complexity" and "task complexity") in the remainder of the paper.

**Interpreting the model.** To close, this model is deliberately crude, and therefore shouldn't be taken too seriously as a model of cognition. The model's purpose is purely to describe the minimal conceptual components necessary to discuss complexity in a clear way, to motivate its main behavioral consequences, and to help the reader keep track of the conceptual moving parts involved in measurement. The cost of this simplicity is that, as the reader will see, it is difficult to discuss some of the effects of complexity clearly in terms of the model.[16] However, I hope the reader will agree that the ease of comprehension afforded by this simplicity is worth the cost of the resulting incompleteness. Equally importantly, it bears emphasizing that although I have formally framed (1) and (2) as optimization problems, I have done this purely in order to describe the tradeoffs faced by an agent facing a complex task (i.e., to describe what the agent would in some sense prefer to do in the face of cognitive costs). By describing the agent's problem this way, I *do not* mean to claim that agents always optimally negotiate these tradeoffs by choosing a $p^{**}$ that maximizes either of these expression.[17] Rather, I mean to suggest directional tendencies away from $p^*$ towards the maxima of (1) and (2) resulting from the DM's attempts to adapt to complexity. For most of what follows in the essay it is sufficient to merely assume that $p^{**}$, to whatever degree it differs from $p^*$, lowers costs ($c(p^{**}) < c(p^*)$) by sacrificing rewards ($\pi(\alpha(p^*)) < \pi(\alpha(p^{**}))$).

[16]Perhaps the most severe limitation of the model is its coarse treatment of $\pi$ and the fact that the model is not rich enough to directly articulate the possibility that complexity can cause $\pi$ to be mis-perceived or perceived with uncertainty, producing an additional source of distortions. This shortcoming will be felt most when motivating complexity Effect 4 (complexity aversion) in the following discussion. In a fuller model in the same spirit, designed to accomodate this possibility, we might for instance break down the procedure into sequential stages of representation and computation in which the DM makes a separate costly choice at each step. This would allow, for instance, for complexity to prevent the DM from correctly understanding $\pi$ and for this to produce the kind of residual uncertainty that may be responsible for complexity aversion (DeClippel et al. 2024). It would also allow us to repair another limitation of the model. The as-if assumption that the DM successfully economizes on cognitive costs, rules out the possibility that her inability to produce accurate beliefs about $\pi$ causes her to *overcomplicate* the problem, choosing a procedure that is unecessarily complex. A richer model that separates out representation from computation would be able to describe this possibility.

[17]After all, as Lipman (1995) points out there is little reason to believe that this higher order optimization problem is any easier than the original one!

## 1.2 Behavioral Consequences of Complexity

Complexity has a number of effects on behavior that make it relevant to the social sciences. These behavioral effects also serve as the basis for the measurement approaches discussed in Section 2, so we will discuss them in some detail here. We will group these into five distinct effects.

**Effect 1** ***(Direct Costs)*** *Even if the complexity of the task is not large enough to induce distortions in behavior (i.e. even if $p^{**} = p^*$ and $\alpha(p^{**}) = a^*$), it produces costs for the DM that (i) matter normatively because of their influence on the DM's welfare and (ii) are sometimes directly measurable.*

The first and most direct effect of complexity is $c(p)$ itself: costs to decision-making that are typically ignored in rational choice, but that are potentially quite important for key tasks in the social sciences like judging policy or institutions. For instance, to the extent that people suffer subjective effort costs (or time costs) from institutional rules – e.g., from complying with a complicated rule inside of an organization; navigating red tape in applying for a permit, social welfare benefits or university admissions; assessing the behavior requirements of a law, etc. – the welfare consequences of policies, contracts or organizational structures may be very different than they would be were we to ignore these direct costs. More importantly for behavioral experiments, in some contexts we can measure these costs directly by eliciting subjective costs or by measuring resource costs like decision time, giving us a direct empirical window into complexity.

**Effect 2** ***(Procedural Distortion)*** *If there exists a suboptimal procedure $p^{**} \neq p^*$, such that either (i) $c(p*) - c(p^{**}) > \pi(p*) - \pi(p^{**})$, under problem (1); or (ii) $c(p^{**}) \leq R < c(p^*)$ under problem (2), the DM may select this simpler, suboptimal procedure.*

Second, if the complexity of the optimal procedure is excessively large, or if there exists a simpler suboptimal procedures that does not result in too large of a reduction in rewards, the DM may be forced to or elect to use a *simpler-than-optimal* procedure instead. In particular in our cost complexity formulation (1), if there is a simplified, suboptimal procedure $p^{**}$ that, relative to optimal procedure $p^*$, generates a reduction in cognitive costs ($c(p*) - c(p^{**})$) that is *greater than* its reduction in expected rewards ($\pi(p*) - \pi(p^{**})$)), the DM may be *incentivized to* choose it instead of the optimal procedure in order to economize on complexity costs. Likewise, under our resource constraint formulation (2), if the optimal procedure is infeasible given cognitive resource stock $R$ (if $c(p^*) > R$), the DM may be *forced to* use a simpler suboptimal procedure $p^{**}$ for which $c(p^{**}) \leq R$. Notably, these simplification procedures might include a number of behaviors that are regularly documented in the behavioral literature including ignoring features of the problem that matter for making optimal choices, restructuring the problem into a form that is solvable using a less costly procedure, using cues from the environment such as defaults, anchors or others' behavior to guide choice, re-using well-practiced procedures that have performed well in the past, or using imprecise, noisy decision rules that economize on the costs of attention and precision. In some choice contexts, these simplified procedures – or markers of the use of simplified procedures

– can be observed or elicited and directly compared to procedures that are consistent with optimal choice. Because it is not always clear, ex ante, what makes one procedure more complex then another, this also motivates an important strand of empirical research on complexity: the search for formal characteristics of procedures that predict complexity costs at the algorithmic level (see Section 4.2).

**Effect 3** ***(Mistakes)*** *Due to the procedural distortions described in Effect 2, complexity may cause the DM to take an action other than the optimal one:* $\alpha(p^{**}) \neq a^*$*. This effect is sometimes measurable even when Effect 2 is not.*

Third, because complexity can induce the use of simpler-than-optimal procedures, it can induce (apparent) *mistakes*: actions that differ from the optimal action given the task's rewards, i.e., that are suboptimal when considering only $\pi(\alpha(p))$ while ignoring $c(p)$. This reduction in decision quality is often the primary effect of complexity of interest in behavioral experiments. Understanding complexity and its role in driving mistakes is important for two primary reasons. First, if we understand the root of the mistake at the procedural level, we can understand why the mistake occurred, in what contexts it will occur and predict future mistakes. Second, understanding the role complexity-derived mistakes play in choice can be important for understanding to what degree we can infer tastes and preferences and therefore welfare judgements from choices, a major use for experiments in contemporary behavioral research. In some choice contexts mistakes can be directly inferred from behavior, but in many others there are serious confounds that produce barriers to this type of inference. As a result, an important strand of research, discussed in Section 3, has focused on developing techniques for separating complexity-derived mistakes from other explanations for behavior.

**Effect 4** ***(Aversion)*** *If the complexity of the choice problem is sufficiently large relative to rewards, the DM may choose to avoid aspects of the decision problem or the entire decision problem altogether (i.e., not take any action) in order to avoid expending procedural costs.*

Fourth, because complexity makes the process of decision-making costly, it may have the effect of causing DMs to make efforts to avoid complex tasks or options. One way they may do this is by opting out of a choice altogether via self-selection. In terms of our descriptive model, if avoiding the task comes bundled with an outside net reward $\beta$ and $\pi(\alpha(p)) - c(p) < \beta$ for all $p$, the DM will have a motivation to self-select out of the task and realize that outside reward rather than complete the task. This is probably a major behavioral effect of complexity in the field, where complexity likely induces decision makers to avoid important choice contexts like voting, investing, entering into contracts, filing for government benefits, starting businesses etc. By contrast, in most experimental contexts, subjects are not given overt opportunities to self-select out of choice – we usually "force" subjects to answer questions in behavioral experiments. Nonetheless, self-selection may be very important in experiments, because many common behaviors may be instances of informal self-selection on the part of subjects. For instance default effects (in which decision makers comply with experimental default decisions), hysteresis (in which decision makers repeat previous choices), imitation (in which decision makers copy what they observe others doing), base-rate neglect (in which decision makers ignore base rates in Bayesian updating tasks and simply report a signal's precisions back to the experimenter), randomization (in which the decision maker chooses from the menu stochastically) and many other classic biases seemingly involve avoiding cognitive effort and decision-making altogether using a cue from their environment or history.[18]

Another way complexity may influence behavior via Effect 4 is by causing the DM to systematically avoid complex *components* of or *options* in the task. Indeed, some experimental results on complexity have been interpreted as growing out of "complexity aversion" in which the DM avoids investing in or exposing herself to complex options. For instance an important strand of the literature on lottery complexity often interprets the effects of complexity on measured certainty equivalents or lottery choices as evidence that subjects preferentially avoid more relative to less complex lotteries (e.g. Huck & Weizsacker 1999, Bernheim & Sprenger 2020, Puri 2023). Complexity aversion is less easily explicable in terms of the rewards articulated in a descriptive model as simple as ours, but may arise from aversion to uncertainty. If complexity generates an inability or unwillingness to fully assess the value of, e.g., a lottery, it may leave the DM in a state of uncertainty regarding the object's value, making the object less attractive to them.[19] DeClippel et al. (2024) provide experimental evidence using several different kinds of choice tasks for this mechanism, showing (i) that subjects who are uncertain about the value of, e.g., lotteries place less value on them and (ii) that valuations for these uncertain subjects, and only these subjects, are predicted by independent evidence of ambiguity aversion.

**Effect 5** ***(Adaptation)*** *Features of the choice problem and choice context that influence $c$ or $R$ (and by extension $p^{**}$ and thus $a^{**}$) but do not influence $\pi$ (and by extension $p^*$ and thus $a^*$), influence $a$.*

Perhaps the most distinctive consequence of complexity, is that it predicts that seemingly irrelevant features of tasks will have significant effects on behavior – that decision makers will *adapt* their behavior to characteristics of the environment that don't seem relevant to the overt rewards of the task (Payne et al. 1993). First, and most obviously, characteristics of tasks that increase the cost of the optimal procedure, $c(p^*)$, can influence behavior through effects 2-4. In particular, the procedures required to achieve a reward $\pi$ can change as a function of details of the task. For instance, increasing the number of pieces of information in the description of the the problem that needs to be processed or held in working memory in order to make an optimal decision may increase the cost of the task's optimal procedure, $p^*$, and may also impact the costs required to achieve other benchmark rewards as well. The same sorts of costs can likewise intensify subjects' motives to avoid complexity by relying on cues from the choice environment or by avoiding selection

[18] This kind of self-selection can be viewed as an extreme version of procedural distortion (Effect 2).

[19] The reason this kind of account is difficult to describe in our simple model is that it depends on an articulation of how beliefs differ across simple vs. complex components of the problem, which are not described in sufficient detail in a model as coarse as ours. See Footnote 16 for a description of how the model could be enrichened to explicitly describe this kind of effect.

of complex options.

As a result, two decision problems with seemingly-similar reward functions and optimal actions can generate very different behavior due to the distorting influence of complexity. These effects can often be quite subtle, operating at the procedural level, meaning even facially similar choice environments can sometimes induce very different rates of optimality (Payne et al. 1993, Gigerenzer et al. 2015). Likewise they can produce sometimes severe failures of predicted *procedural invariance*, producing seemingly different measures of preferences in theoretically isomorphic choice environments (Friedman et al. 2017, 2022, Beauchamp et al. 2020, Bauermeister et al. 2018, Oprea 2024, Freeman & Mayraz 2019, Freeman et al. 2019). See Section 4.1 for a discussion of a literature examining these types of effects. Second, changes to the decision environment or the mental state of agents that influence the cognitive or temporal resources available to a DM – i.e., that modulate the resource budgets of the DM, $R$ – can produce changes in the procedures the DM uses and as a result the quality of decisions. Leveraging these sensitivities to produce treatment variables has therefore proved an important tool for detecting and measuring complexity (see Section 2.2.2). Third, changes to the framing of the decision problem can influence complex choices, because they sometimes give the DM easy alternatives to sinking procedural costs. For instance, providing subjects with defaults, anchors or irrelevant information, the researcher can make it easy for the DM to avoid complexity by self-selecting out of decision-making and linking her behavior to external cues instead. Finally, changes to superficial aspects of the decision problem can influence behavior in complex settings because doing so can alter the automatic responses the DM is forced to rely on when she is unable (or unwilling) to use sufficient attention or cognitive control to override these automatic responses. For instance, when decision makers use strategies that economize on attentional resources, their decisions become susceptible to salience effects (Bordalo et al. 2012, 2022), causing their behavior to be shaped by bottom-up attentional processes that are influenced by factors that have little to due with the optimal response (Bohren et al. 2024).

### 1.3 Recap

To organize the preceding material and better set up the problem of measurement, we close this section by reviewing some of the important terms and distinctions often used to break down and taxonomize complexity. To frame this, we can think of the relevant terminology and key distinctions made in the literature as representing answers to a series of *distinct* questions about the nature and effects of complexity in the researcher's application. Though answering these questions is often unnecessary when modeling or measuring complexity, answering some of them can often help researchers better design their experiments and interpret their findings.

**WHAT** cognitive resources do we think are being taxed when we call a procedure or task "complex." Or, in terms of our model, in what medium is $c(p)$ denominated? Commonly invoked resources include working memory, cognitive control, time and attention, any of which might be strained or overwhelmed when implementing a procedure to solve a problem. Answers to the

"WHAT" question gives rise to resource-based ways of differentiating types of complexity, giving rise to specializations of the term such as **time complexity** (runtime expenditures) and **space complexity** (working memory expenditures), the two most common varietals discussed in computer science. One could easily extend this descriptive construction to discuss variations like attentional complexity (for procedures that tax scarce attentional resources) or control complexity (for procedures that task scarce reservoirs of cognitive control).

**WHERE** do the complexity costs arise in cognition: in the act of representing information to one's self or in the act of performing computations on those representations? The distinction between representation (the act of gather and storing cognitive symbols in memory) and computation (the act of manipulating those cognitive symbols during information processing) is a classic one in cognitive science and can sometimes be more useful for differentiating types of complexity than taxonomies describing the primitive cognitive resource taxed (Thagard 2005). Following this we can describe **representational complexity** as the cost (often denominated in terms of working memory or attention) of representing the problem to one's self or more generally of remembering and attending to a set of cognitive symbols required for computation, while **computational complexity** is the cost (often denominated in effort or time) of manipulating these symbols in order to arrive at an answer to a question or make a decision (Ba et al. 2024).

**WHEN** in the process of cognition do the complexity costs arise? In some problems complexity costs arise in the implementation of the procedure (**the complexity of implementation**) while in other cases the costs arise in the meta-task of identifying which procedure is optimal (**the complexity of identification**) prior to implementation. Or in other words, is the difficulty of the task rooted in the fact that it is difficult to reason towards an optimal procedure (complexity of identification), or simply difficult to perform that procedure (complexity of implementation)?

**WHY** will the DM be compelled to respond behaviorally to complexity? As we've emphasized there are two broad possible answers. One is that the DM is *forced* to respond to complexity because the resource costs of the optimal procedure, $c(p^*)$ exceed the cognitive resources available to the decision-maker, $R$, as in the version of the DM's problem described by equation (2) above. Another is that the DM *chooses* to use a suboptimal procedure, because the direct effort costs (or indirect opportunity costs of time) of the optimal procedure, $c(p^*)$ are not worth the rewards, $\pi(p^*)$, it produces, a'la equation (1). When we describe complexity as a negotiable cost (a'la equation (1)), we will often refer to it as **cost complexity** because in such cases we describe and measure complexity in terms of the (often financial) medium in which the rewards in the task are denominated rather than a specific cognitive resource.[20]

**HOW** are we applying the term "complexity": at the level of the procedure or at the level of the

[20] That is, sometimes instead of focusing attention on the specific cognitive resources taxed in cognition (e.g., space or time complexity) or the cognitive acts generating the costs (e.g., representational or computational complexity), resource costs can be collectively described hedonically as costs, denominated in the same reward medium of the task (e.g., money).

task? As we've emphasized, sometimes it is natural to apply "complexity" to the costs required of a specific procedure (**procedural complexity**), e.g., the procedure a DM actually uses, $c(p)$. Other times it is instead natural to apply the same term to the costs *required to optimally* solve the problem (**task complexity**), $c(p^*)$. As we've discussed task complexity is simply the procedural complexity of the procedure required to optimally solve the task. Avoiding confusion between these two applications of the word "complexity" is particularly important for the task of measurement.

Running through these questions prior to modeling or measuring complexity can often be extremely useful for pin-pointing both the hypothesized cognitive mechanisms complexity operates through and for predicting effects of complexity on behavior. Indeed, the efficacy of some of the measurement tools discussed in the next Section *depend* on the answers to some of these questions. That said, it is rarely necessary to answer *all* of these questions with any precision, and in many contexts it will be difficult to do so ex ante. Nonetheless, considering and making some of these distinctions can often clarify what aspects of complexity are being invoked and measured, influencing both the credibility of the experimental design and the interpretation of the results.

# 2 Measurement Tools

In this section we will review some of the key empirical tools behavioral researchers have developed for measuring and detecting complexity and its influence on behavior. We will take special care to show how these methods derive directly from our operationalization of complexity and the primary behavioral effects we motivated in Section 1.2. We will break these tools into two categories. First, in Section 2.1 we review a number of behavioral and process metrics that have been deployed in the literature for measuring complexity, sometimes indirectly via its secondary effects. Second, in Section 2.2, we discuss several types of *treatment manipulations* that have been used to detect the influence of complexity, often by varying characteristics of the choice problem in such a way as to alter $c(p)$ or $R$, without altering $\pi(p)$. These approaches are summarized in Table 1 where we also summarize (i) which complexity effect from Section 1.2 the measurement approach connects to and (ii) which element of our descriptive model from Section 1.1 it measures or manipulates.

## 2.1 Metrics

We begin by reviewing experimental outcomes that have been used to measure or detect complexity and discuss how they connect to the complexity effects described in Section 1.2. These metrics can be grouped into three categories: direct measures of the costs of decision-making (motivated by complexity Effect 1), measures of behavioral responses to complexity (motivatedby Effects 2-4) and measures derived from specially elicited beliefs (motivated by Effects 1 and 3).

| Metrics | | |
|---|---|---|
| *Cost Metrics* | | |
| Subjective cost | Effects 1, 3 | elicit $c(p)$ in money |
| Response time | Effect 1 | measure $c(p)$ in response time |
| Biometrics | Effect 1 | measure biological response to $c(p)$ |
| *Behavioral Metrics* | | |
| Procedural measurement | Effect 2 | measure $p$ |
| Choice inconsistency | Effect 2 | measure consistency of $a$ over repetitions |
| Mistakes | Effect 3 | test whether $a = a^*$ |
| *Belief-based Metrics* | | |
| Beliefs about optimality | Effect 3 | elicit probability $a = a^*$ |
| Subjective ranking | Effect 1 | rank $c(p^*)$ across tasks |
| **Treatment Instruments** | | |
| *Problem Characteristic Manipulations* | | |
| Manipulating task features | Effect 5 | manipulate $c(p^*)$ |
| Manipulating procedure/description | Effect 5 | manipulate $c(\cdot)$ |
| Removing availability of alternative mechanisms | Effect 5 | manipulate $\pi(\cdot)$ |
| *Cognitive Resource Manipulations* | | |
| Automating information processing | Effect 5 | manipulate $c(\cdot)$ |
| Constraining cognitive resources | Effect 5 | manipulate $R$ |
| *Reward Manipulations* | | |
| Varying rewards | Effect 5 | manipulate $\pi(\cdot)$ |

Table 1: Measurement tools reviewed in Section 2. The second column shows the complexity effect from Section 1.2 the tool relies on. The third column highlights what part of the descriptive model in Section 1.1 it measures or manipulates.

### 2.1.1 Direct Cost Metrics

Procedural costs can sometimes be experimentally measured, giving us metrics that are closest to true *measurements* of complexity. Importantly, this type of metric generally *does not* measure the complexity of the *task* (the resources required to optimally or correctly solve the task), but instead the complexity of the *procedure* the subject actually used (or expects to use). Because an important effect of complexity is to cause people to substitute to simpler-than-optimal procedures (complexity Effect 2), these two objects can be quite different from one another. As a result, failing to recognize that these measurement approaches measure $c(p)$, *not* $c(p^*)$ can lead to serious confusions in interpretation. These metrics are differentiated by the type of resource cost they are meant to measure, and we review three varieties.

**Subjective Cost**. Researchers can measure subjective effort costs ("cost complexity" a'la equation 1), by eliciting them directly in the same reward medium as the decision task, $\pi$, typically money.

By eliciting subjects' willingness to pay (WTP) to *avoid* a task – or their willingness to accept (WTA) to be exposed to a task – we can measure the subjective cost, $c(p)$, the subject expects to suffer in the process of identifying and implementing a procedure, if assigned the task. This kind of measurement can be conducted using standard methods like the Becker-Degroot-Marschak (BDM) method (Becker et al. 1964), multiple price lists or by having the subject choose between the task and varying payment amounts over a sequence of binary decisions. Because this type of design relies on the subject's desire to avoid the costs of a task, its measurement relies to some extent on complexity Effect 3, aversion.

Importantly, this approach (like other approaches from this subsection) doesn't measure the complexity of the *task* – the cost of the *optimal* choice $c(p^*)$ – but instead the costs the subject expects to incur given the procedure she expects to employ, $c(p^{**})$, where $p^{**}$ needn't be $p^*$. In other words it measures what we have called *procedural complexity* not *task complexity.* This means the measure doesn't actually answer the question researchers are often most interested in: how costly would it be to optimally solve the task, and to what degree subjects are economizing on those cost in their behavior.

Except in special designs discussed below (Section 4.2), this kind of metric can be difficult to interpret because the researcher often doesn't directly observe the procedure $p$ used by the subject and so often can't tell what procedure the measured cost is attached to. Even if this were not a problem, the procedure the subject expects to use in a task is endogenously selected by the subject, making WTP/WTA potentially biased measures of the average costs produced by that procedure in the population.[21] However, in specially designed experiments – which I call "assigned procedure" designs – designed specifically to map the structure of procedural costs (a literature reviewed in Section 4.2 below), these kinds of measures can be very valuable. An example is Oprea (2020) who assigns subjects procedures (algorithmic rules) to implement rather than standard decision tasks, allowing him to measure the costs $c(p)$ of a sequence of specific, identifiable procedures. After assigning subjects a series of rules and paying them to implement those rules, he uses the BDM mechanism to elicit subjects' willingness-to-accept to be asked to implement each rule again in the future. This yields a rather direct measures of the cost complexity of a series of exogenously assigned procedures.

**Response Time**. Another direct measure of complexity costs is response time (RT): the clock time required for the DM to make and submit a choice. Response time is closely related to *time complexity*, one of the key types of complexity discussed in computer science, linked to one of the key resources computers and humans alike must expend in the process of making choices and inferences. In many decision settings, the subject's time budget is a major constraint on her ability to identify and implement an optimal procedure, making decision time a natural alternative to monetary cost for measuring complexity. Even when time isn't constrained in an experiment, the

[21] In particular, if the costs of any given procedure vary across subjects, we should expect the costs of procedures actually used to be oversampled from the lower range of the cost distribution.

opportunity cost of time is an important input into the subjective cost of implementing a procedure, and so time is one of the variables traded off against the rewards of a problem when selecting a procedure.

Although RT is a centrally important variable in the behavioral sciences (see Spiliopoulos & Ortmann (2018) for a recent review) – and is often extremely valuable for understanding the procedures subjects use to tackle experimental tasks – like subjective cost measurements, it measures procedural complexity, not task complexity: when we measure RT in a decision task, we are unavoidably measuring the time complexity not of the task, $c(p^*)$ denominated in time, but of the (possibly suboptimal) procedure, $c(p)$, the subject has elected to use in the task. Because of this, as with subjective cost, we cannot easily use RT to reliably measure how time-complex a *task* is: how much time would be required to use an optimal procedure in the task. As a result, RT is a difficult to interpret measure of complexity except in special design (discussed in Section 4.2). For example, it is possible for RT to be non-monotonic in a task's complexity if the difficulty of optimizing in very hard tasks induces subjects to divert to particularly time-simple procedures (e.g., to simply "give up" and randomize).[22]

Instead we can use RT to measure how complex the procedure a subject actually used is *for that subject*. This is for example how Wilcox (1993) – an early paper in behavioral economics that used RT as a measure of what we call complexity – interprets RT. Consistent with the idea that RT measures the sophistication of the subject's chosen procedure, there is some evidence suggesting that there is a negative relationship between RT and clear mistakes in problems that require sophisticated procedures (e.g. Rubinstein 2013, 2016), but a positive relationship in some simpler settings (e.g. Alós-Ferrer et al. 2016).[23] Even when used to measure procedural complexity, RT suffers from some of the same observability and endogeneity problems discussed above for elicitations of subjective costs – we often don't know what procedure subjects are using and even when we do, their time costs are potentially downward biased by their endogenous selection. Thus, as with studies focused on measures of subjective costs, researchers using RT to study procedural complexity (Bettman et al. 1990, Oprea 2020) have avoided these problems by directly assigning procedures to subjects and studying their time complexity, often in enterprises aimed at mapping the cost functions of procedures (Section 4.2).

**Biometrics**. Finally, researchers have used biological measures to infer the costs of behavior, typically when attempting to measure subjective effort costs. For instance, pupil dilation, measurable using eye-tracking devices, is well-established as a metric of cognitive effort and is correlated with

[22]Gonçalves (2024) uses a sequential sampling model to theoretically study the relationship between RT and complexity and shows that RT should, indeed, be expected to be non-monotonic in task complexity for just this reason.

[23]The complicated relationship between RT and decision-quality is elegantly organized by sequential sampling models like the drift-diffusion model (DDM), a workhorse model in psychology that has strong neuroscientific foundations and that has been increasingly applied to decision-making (Fudenberg et al. 2018, Krajbich et al. 2014, Webb 2018, Gonçalves 2024).

activation of areas of the brain related to motivated attention allocation (van der Wel & van Steenbergen 2018). Skin conductance and heart rate measures have been used to measure stress which may be related to cognitive effort and thus complexity costs (Lempert & Phelps 2013). Finally, EEG and fMRI data may produce evidence of cognitive difficulty (Shenhav et al. 2017). Once again, these types of metrics measure the complexity of the procedure the subject used (procedural complexity), not the complexity of the task itself, the costs that would be required to optimally perform the task. Thus, these types of measures should be treated with the same care and interpreted with the same caution as these other measures.

### 2.1.2 Behavioral Metrics

An alternative to measuring complexity costs directly (via Effect 1) is to measure their primary effects on behavior (via Effects 2-4). Because this alternative approach typically involves directly *comparing* subjects' actual behavior to benchmarks of optimal behavior, it (unlike direct cost measures) produces metrics of *task complexity*, the complexity of the optimal procedure. Indeed, this is the major difference in the use case of these two types of metrics (direct cost vs. behavioral metrics): while the direct cost metrics discussed above measure *procedural complexity*, the behavioral metrics reviewed in this section are typically aimed at measuring *task complexity*.

**Procedural Measurement**. The most direct effect of complexity on behavior is to induce DMs to adopt simpler-than-optimal procedures (Effect 2). Researchers can sometimes measure the procedures subjects use directly and compare them to optimal procedures, using the gap between the two as a measure of complexity. In some contexts, this can be done relatively directly because the subject's procedure can be inferred from natural behavior in the task. For instance, Banovetz & Oprea (2023), discussed in more detail in Section 4.2 below, study bandit tasks – search problems that feature an exploration-exploitation tradeoff – that are simple enough to allow the authors to structurally identify the procedures subjects use to identify and select a terminal choice in the task simply by observing how subjects go about conducting search. By formalizing these procedures as algorithms (finite state machines) and making use of direct measurement of the component costs of procedures from previous work (Oprea 2020), they are able to show that subjects use simpler procedures than optimal ones, providing some evidence that subjects' choices are distorted by efforts to economize on complexity.

In most settings, this kind of direct measurement of the subject's procedure is not naturally available and as a result the literature has developed a number of tools for eliciting procedures from subjects. One particularly important technique is *process tracing*, which involves measuring the order and timing with which subjects assess pieces of information in a decision problem when processing that information to inform a choice; see Schulte-Mecklenbeck et al. (2017) for a review. In a classic early example, Payne et al. (1988) use a computer protocol in which subjects uncover information by mousing over hidden pieces of information called *Mouselab* to identify the procedures subjects use to arrive at lottery choices (Payne et al. 1993). Similarly, Ba et al. (2024) use

Mouselab to study the procedures subjects use to economize on the representational complexity of Bayesian updating tasks. More recently, eye tracking technology has allowed researchers to do the same thing naturalistically by measuring how the subject's gaze moves between pieces of information. For instance Arieli et al. (2011) use eye tracking to provide evidence that subjects use heuristic procedures to evaluate lotteries, rather than optimal procedures that are consistent with the maximization of risk preferences. In some cases experimental protocols can be designed so that natural choice process data provides a similar window into the subject's procedure. For instance, Murawski & Bossaerts (2016) study knapsack problems in which decision makers must select a set of objects, each of which has a value and a weight: the task is to maximize the summed value of selected objects without exceeding a summed weight limit. By allowing subjects to assemble and revise their bundles in real-time using a special computer interface, Murawski & Bossaerts (2016) are able to partially observe the algorithms subjects use to solve the problem and show how they are simple relative to optimal algorithms. Bohren et al. (2024) elicit subjects' beliefs about the properties of lotteries and use these beliefs to identify characteristics of the procedures subjects use to form representations of them. By comparing these representations to the true characteristics of lotteries, they are able to show that subjects use simplified procedures that result in mistaken representations of informationally rich lotteries – evidence that behavior in these tasks is distorted by representational complexity.

A third method for measuring the subject's procedure is to use subjects' explicit descriptions of their own procedure, a method called "protocol analysis" (Ericsson & Simon 1980, 1993). One approach is the "think aloud protocol" in which the decision maker verbally describes her decision-making process while making decisions. For instance, Capra (2019) uses this method to uncover subjects' reasoning process in beauty contest game, and finds that verbal data accurately classifies subjects according to the sophistication of their reasoning. Another approach is to incentivize subjects to retrospectively verbally describe the rule they believe is optimal or that they used to make their own decisions. Kendall & Oprea (2024) study forecasting tasks in which subjects are shown a dataset of variables and must infer an accurate forecasting procedure from that data and use it to forecast future outcomes. Subjects are directly asked to describe the optimal procedure in words, and are incentivized to do this accurately: the procedure they describe will be given to a future participant to guide her forecasts, and the subject will be paid according to the accuracy of those forecasts. Kendall & Oprea (2024) transform these descriptions into formal models of algorithms (finite automata) and thereby are able to characterize what features of optimal procedures make them difficult to identify. Arrieta & Nielsen (2024) ask subjects to make a number of decisions between sets of charities or sets of lotteries, and incentivize subjects to characterize the procedure they themselves used to make those decisions. To incentivize these self-descriptions, the subject's descriptions are given to future participants who are incentivized to *match the subject's own choice*, thus giving the subject an incentive to accurately describe not the optimal procedure, but the procedure they themselves used. They use this data to show that procedures become more describable and replicable as the tasks become more computationally intensive.

**Choice Inconsistency**. One type of Procedural Measurement is particularly easy to collect: choice (in)consistency, or the difference between subjects' choices in repetitions of identical tasks. A long literature in neuroscience emphasizes that it is less cognitively expensive to perceive, represent and aggregate data imprecisely than precisely, making the use of imprecise (noisy) procedures a natural way of economizing on the complexity costs associated with, precise, optimal decision-making (see Woodford (2020) for an overview). Intuitively, casually or carelessly processing information is cognitively cheaper, ceteris paribus, than marshalling the sharp attention and cognitive control required to make precise decisions. Because of this, the influence of complexity over choice can sometimes be measured relatively easily without using the elaborate procedural measurement techniques just discussed.

Noise in decision-making can easily be measured simply by repeating the same choice task in an experiment, typically by distributing both repetitions in a series of other tasks to camoflauge the repetition. Choice inconsistency (the submission of different choices in identical tasks) is evidence that the subject's decision-making includes noise, which may be an indication that subjects are using a simpler-than-optimal noisy procedure, economizing on complexity. Of course the presence of noise may stem from other sources (e.g., random utilities or preferences for randomization). For this reason, most convincing evidence linking complexity to a behavior via choice inconsistency is focused on documenting a *systematic* relationship between noise and that behavior. For example, Khaw et al. (2021) show that small stakes risk aversion is correlated with choice inconsistency, as predicted by noisy coding models that root lottery anomalies in the use of imprecise decision procedures. Recent work has likewise linked noisy procedures to risk attitudes Frydman & Jin (2022) probability weighting, (Enke & Graeber 2023, Oprea 2024, Frydman & Jin 2023, Vieider 2024, Oprea & Vieider 2024), the description-experience gap (Oprea & Vieider 2024), inference errors (Enke & Graeber 2023, Ba et al. 2024) and hyperbolic discounting (Enke, Graeber & Oprea 2024, Gershman & Bhui 2019, Vieider 2021, Gabaix & Laibson 2022).[24]

**Mistakes**. The third effect of complexity (Effect 3 from Section 1.2) is that it induces people to make mistakes – to make choices that fail to maximize $\pi$. Because of this, clear evidence that a subject has made a mistake is often an important tool for detecting the influence of complexity. In some settings (e.g., simple optimization problems) this inference is direct. For instance, Murawski & Bossaerts (2016), discussed in more detail in Section 4.2.4, study knapsack problems, constrained optimization problems in which there is a well-defined optimum that deterministically determine's the subject's payment. Murawski & Bossaerts (2016) show that subjects fail to find this optimum at a high rate and are able to interpret this failure as a consequence of the computational complexity of these tasks. However, as we discuss in Section 3 below, identifying complexity from mistakes is sometimes not so straightforward because of the confounding influence of motivational, epistemic

[24]Evidence that a behavior is systematically correlated with noise can be interpreted as evidence that the behavior is a consequence of an imprecise procedure, and thus suggest the effects of complexity. The reverse inference does not hold: many heuristic responses to complexity are quite stable, producing predictable behavior. Thus evidence of stable behavior is not evidence *against* the effects of complexity on choice.

and competing cognitive drivers of the behavior. Because of this, mistakes are sometimes difficult to directly observe and therefore must be inferred using other methods reviewed in this Section.

### 2.1.3 Belief-Based Metrics

Finally, researchers can sometimes measure *task complexity* using elicitations of subjects' beliefs about the task. This is a third class of complexity metric, which is useful especially in experiments in which it is difficult to clearly identify what procedures subjects have used or to clearly establish whether the subject has made a mistake.

**Beliefs About Optimality**. A very direct way of determining whether complexity has caused a subject to use a simpler-than-optimal procedure is simply to ask them, directly. In particular, a researcher can ask a subject how likely it is that their own choices are optimal: the probability that $a^{**} = a^*$ and implicitly that $p^{**} = p^*$. This approach is taken by Enke & Graeber (2023), who (i) ask subjects to first perform a series of tasks and, in each case, (ii) ask them to submit an unincentivized estimate of the percentage likelihood that their decision in the task is optimal (i.e., maximized $\pi$). This allows them to gather a measure of what they call "cognitive uncertainty" (CU): the subject's level of doubt that they submitted action $a^*$ in the task. Enke & Graeber (2023) show that this measure is highly predictive of a number of behavioral anomalies including probability weighting, biases in Bayesian updating and compression of expectations; Enke, Graeber & Oprea (2024) similarly show that this measure is also highly correlated with hyperbolic discounting. Enke, Graeber, Oprea & Yang (2024) measures these beliefs for 30 separate experiments, covering a wide range of decision contexts, and shows that in nearly all of these settings cognitively uncertain subjects are markedly less sensitive to variation in payoff-relevant parameters. They use this evidence to argue that complexity is responsible for the *excessive insensitivities* that characterize many classic behavioral anomalies. To whatever degree we expect subjects to be *aware* that they have used simpler-than-optimal procedures to avoid complexity costs, belief measures like these will serve as an effective indicator that a subject found that task complex (i.e., costly or difficult to optimize in). In Section 3.1 we discuss the use of such measures to identify complexity in more detail.

**Subjective Ranking**. Another belief-based approach is to ask subjects to report how subjectively difficult they believe it would be to make an optimal choice in a task. One way to do this might be to ask subjects to provide a response on a Likert scale by asking, "on a scale of 1-10, how difficult do you think it would be to make an optimal choice in this task." However such measures are potentially difficult for subjects to provide and for the analyst to interpret, given subjectivity in the interpretation of the scale. An alternative is to instead ask subjects to ordinally rank tasks in terms of their relative difficulties. Gabaix & Graeber (2023) develops a methodology for implementing and interpreting this kind of measure and demonstrates it on intertemporal consumption choice decisions. Subjects are shown a pair of intertemporal choice tasks on their screen side by side and asked to identify in which of the two tasks it is more difficult to make choices that lie within some window of the optimum, calibrated so that around half of actual responses fall within that window. By doing this, each task receives a complexity ranking. By calibrating these rankings using "anchor tasks" involving mental algebra and applying psychometric techniques first proposed by Bradley & Terry (1952), Gabaix & Graeber (2023) are able to assign each task an overall, cross-subject complexity score that can be used to predict complexity judgements. They show that these judgements can be predicted based on objective features of tasks and that these measured judgements in turn predict attenuated responses to payoff-relevant variables (a key prediction of several models of complexity, including one they propose that builds from Gabaix (2014)).[25]

## 2.2 Treatment Instruments

In some contexts, complexity can be measured or inferred directly simply by examining metrics like those described in the previous subsection. However an alternative and often complementary approach is to study whether a studied behavior *changes* in response to treatment interventions that are linked to complexity. Complexity Effect 5 from Section 1.2 highlights that one of the key consequences of complexity is behavioral sensitivity to features of tasks and decision contexts that should have no effect on a decision maker who does not face procedural costs, i.e. variations that do not influence $\pi$ or $\pi(p^*)$ but do influence $c(p^*)$ or $R$. These sensitivities often form the basis for treatment manipulations meant to detect the influence of complexity. We will group these into three types of manipulations: manipulations of the experimental task, manipulations of the cognitive resources available for the task and manipulations of incentives in the task.

### 2.2.1 Manipulating Characteristics of the Problem

A first type of treatment manipulation, is to vary features of the task either by changing the choice problem or by changing superficial details of the choice environment that plausibly influence the costs of information processing (or the way imperfect procedures respond to idiosyncracies in the task), but that should not cause a perfectly rational decision maker to change her behavior. To the degree a behavior responds to such manipulations, we have some evidence that that behavior is shaped by complexity.

**Manipulating Task Features**. One common strategy is to attempt to modify the choice problem itself in such a way as to increase the costs of optimizing: to change the choice problem in such a way as to increase the cost $c(p^*)$ of the optimal procedure $p^*$ without changing the overt rewards to the problem $\pi(a^*)$ or the optimal action in the task $a^*$. To the degree such a manipulation intensifies or attenuates the behavior, we have suggestive evidence that complexity influenced the

[25]Notice that this method does not ask subjects to rank how difficult they found the task, but instead how difficult it would be to make an optimal decision in the task. The former would be an indirect measure of $c(p^{**})$ (the complexity of the procedure actually used) while the latter measures $c(p^*)$ (the complexity of the optimal procedure for the task, i.e., the complexity of the task itself).

behavior in the original task. To the degree it induces changes in the complexity effects discussed in Section 1.2 (i.e., Effects 1-4) such designs can produce very direct evidence that the manipulated feature generated complexity in the task. For instance, researchers studying the complexity of lotteries have manipulated features of lotteries thought to influence complexity without influencing factors usually thought to influence preferences for the outcomes of lotteries. Examples include, for instance, the number of outcomes (Huck & Weizsacker 1999, Bernheim & Sprenger 2020, Puri 2023, Bohren et al. 2024) or the compoundness of the lottery (Wilcox 1993). As we discuss in Section 4.1 below, this strategy not only identifies a role for complexity in the behavior, it is often also used as a method to identify features of tasks that systematically contribute to complexity.

**Manipulating Procedure and Description**. A subtle effect of complexity is that it causes seemingly superficial details of the choice environment or the presentation of information that do not alter $\pi$ to nonetheless alter behavior (Payne et al. 1992). One reason for this is that the choice environment can influence the menu of procedures the subject has available to her to tackle the problem (influencing the complexity of implementation) or what procedures are easiest to find in a process of search (influencing the complexity of identification). In some choice environments, $c(p^*)$ may be larger than in others, even if they induce the same optimal action. Moreover, in some choice environments, simpler-than-optimal procedures are more sensitive to details of the choice environment than optimal procedures are. For instance, a subject choosing rationally for the best alternative from a list of options will search all items in the list in order to identify the best option, meaning the order in which objects are shown to the subject will have no influence on the optimality of her choice. By contrast Simon's famous *satisficing* heuristic (Simon 1955), a simpler procedure (Bettman et al. 1990, Salant 2011, Sanjurjo 2024), involves searching the list until reaching the first one that is "good enough," making the optimality of choice highly dependent on the order in which items are presented, a superficial detail of the choice environment.

Because of this, it is sometimes possible to detect the influence of complexity by comparing tasks that vary in superficial respects that should not influence behavior if subjects are reliably using optimal procedures. If such manipulations of superficial aspects of the choice environment change behavior, we have evidence that subjects are using suboptimal procedures that respond to the environment-specific decision costs of the problem or that are "overfit" to superficial details of the task. For example, there is a large and growing literature on *procedural invariance*, showing that subjects' apparent preferences for risk are strongly influenced by the method used to elicit those preferences (Friedman et al. 2017, 2022, Beauchamp et al. 2020, Bauermeister et al. 2018, Oprea 2024, Freeman & Mayraz 2019, Freeman et al. 2019), and that measured tastes for risk are only very weakly correlated within-subject across these seemingly superficial changes to the choice environment. This suggests that subjects' choices are only weakly influenced by their personal objective function for risk ($\pi$), and are instead strongly influenced by elicitation-specific costs of implementing an optimal procedure that complies with those preferences. Gigerenzer & Hoffrage (1995) shows that subjects' ability to form Bayesian posteriors in inference tasks often depends on whether the primitives of the task are described in a frequentist or probabilistic format – subjects are often far more Bayesian in frequentist settings. Gigerenzer & Hoffrage (1995) argues that the number of computations required for Bayesian inference are lower and the obviousness of the optimal procedure is higher when information is frequentist, suggesting that the complexity of identifying and implementing the optimal procedure is lower in these settings. This in turn suggests that complexity is an important determinant of inferential failures in Bayesian updating.

**Removing Availability of Alternative Mechanisms**. An alternative to manipulating the complexity of the task, is to remove competing explanations for the behavior (i.e., explanatory forces other than complexity) from the task. To the degree the original behavior survives this removal of alternative motivations, we have evidence that that behavior was a consequence of the complexity of the task rather than of these removed explanatory forces. Oprea (2024) applies this idea to the key lottery anomalies motivating prospect theory (Kahneman & Tversky 1979, Tversky & Kahneman 1992), including probability weighting and loss aversion. By paying subjects the expected value of the lottery rather than stochastically based on lotteries' constituent probabilities, the experiment removes risk from the problem and with it scope for standard risk preference-based explanations for these behaviors. However doing this retains much of the information processing required to value lotteries, thus preserving the complexity of the task. Oprea (2024) shows that probability weighting and loss aversion continue to appear in these "deterministic mirrors" of lotteries, strongly suggesting that these anomalies are to a great extent a consequence of complexity rather than risk or risk preferences. Enke, Graeber & Oprea (2024) similarly study "atemporal mirrors" of intertemporal choice problems by removing time delay from these tasks and use this to show that hyperbolic discounting is primarily a response to complexity following similar logic. We will discuss this approach in more depth in Section 3.1 below.

### 2.2.2 Manipulating Availability of Cognitive Resources

Instead of manipulating the complexity of the task, in some settings it is possible to identify complexity by manipulating the cognitive resources available to the decision maker to cope with the complexity of the task. If the original behavior was *not* driven by complexity (i.e., by constraints in available cognitive resources), such manipulations should have limited effect. However, to the degree adding or subtracting the availability of cognitive resources intensifies or attenuates the original behavior, we have some evidence that that behavior was shaped by complexity.

**Removing Cognitive Resources**. Researchers can sometimes identify the influence of complexity by artifically *reducing* the availability of resources required to implement an optimal procedure. Suppose $c(p)$ is denominated in terms of an objective resource like time, working memory or attention. Then, following (2), by limiting the availability of resource $R$ , the experimenter can manipulate the subject's ability to identify or implement the optimal procedure in a task. To the degree such a manipulation intensifies a given behavior, we have evidence suggesting that that behavior is shaped by complexity.[26] The two most common resources the literature has used for

[26]Notice that this doesn't operate by manipulating the complexity of the task, but instead by manipulating the

this type of manipulation are (i) the amount of working memory the subject has available for the task and (ii) the amount of time the subject has available for the task. These two types of manipulations map neatly into the two types of complexity most often discussed in computer science: (i) *space complexity* (complexity denominated in terms of required working memory) and (ii) *time complexity* (complexity denominated in terms of required time-consuming operations).

Cognitive load manipulations attempt to limit the availability of working memory, and therefore test for a role for what computer scientists call space complexity. The most common method for inducing cognitive load is to task subjects with holding a long sequence of numbers, letters or symbols in memory while making a decision. This has been shown in a number of contexts to reduce performance in cognitive tasks, worsen biases and produce intensifications of behaviors often attributed to preferences, $\pi$, like risk aversion and impatience; see Deck & Jahedi (2015) for a recent review of this literature. Time pressure manipulations instead attempt to limit the availability of decision time, and therefore test for a role for what computer scientists call time complexity. These manipulations typically involve simply constraining and varying the number of seconds available to the subject to submit a decision. They have been shown to have similar effects to those of cognitive load manipulations, often worsening performance in optimization and inference tasks and intensifying behaviors often attributed to preferences; see Spiliopoulos & Ortmann (2018) for a comprehensive review of the literature. Deck et al. (2021) show that these two types of manipulations have strongly parallel effects in a range of tasks, suggesting that both space and time complexity have important influences on human behavior in a wide range of settings and, notably, that they induce the selection of similar simplified procedures. Similarly, Ba et al. (2024) use Mouselab protocols (discussed above) to reduce the *attentional resources* available to subjects, and show that doing so increases the uptake of procedures that economize on the representational complexity of updating.

A closely related approach is to study the correlation between measures of subjects' cognitive ability and their behavior. Instead of exogenously varying subjects' access to cognitive resources, this strategy uses natural variation in those resources across people, as measured by performance on cognitive batteries, to detect the influence of complexity. The assumption behind this exercise is that subjects with higher measured cognitive ability either suffer lower constraints (e.g., have higher working memory capacity and thus a higher $R$) or suffer smaller cognitive costs from implementing optimizing procedures (i.e., $c(p^*)$ is lower). To the degree either is true, evidence that a behavior is correlated with measured cognitive ability is evidence that that behavior was influenced by the complexity of the task. See Benjamin et al. (2013) for an example of this approach.[27]

resources the DM has to cope with a task's complexity.

[27]Stanovich & West (2008) show that while many classical biases are correlated with measures of cognitive ability, many others are not. The authors argue that biases that are uncorrelated with cognitive ability may be biases that result from limitations in the suite of learned procedures the subject have available to them for approaching the task (what the authors call "mindware"). In our terminology, this suggests that errors driven by the complexity of identification may be less related to cognitive ability than errors driven by the complexity of implementation. This may be because the cognitive resources required to, e.g., infer Bayes' rule from first principles far exceeds the

**Automating Information Processing**. While it is often straightforward to *remove* access to cognitive resources, $R$, it is often harder to directly add to them (it is, for instance, generally not possible to directly increase subjects' reservoir of working memory). However, one closely related alternative is to introduce automation tools that allow the subject to offload some cognitive/procedural costs to the experimenter or to a computer. This effectively lowers the costs of the optimal procedure for the subject, making it less likely that resource constraints bind (and lowering the hedonic costs of information processing as well). To the degree automating some of the information processing to a computer (thereby lowering procedural costs) erodes a behavior, we have evidence that that behavior was in part due to complexity.

For example, Banovetz & Oprea (2023), discussed above, measure the procedures subjects use to make choices in simple bandit problems. In a baseline condition they find that most subjects use simpler-than-optimal procedures – procedures that contain fewer than the optimal number of automaton states – leading to significant mistakes. In a second treatment, they have the computer track information on behalf of the subjects in a specific way designed to remove the cognitive costs associated with tracking states. When these costs are removed, subjects switch to using "complex," optimal procedures and therefore make optimal choices at a high rate. Because the treatment intervention does not change the rewards to the problem, $\pi$, this result therefore identifies complexity as the driver of the simplified procedures used in the original treatment. Similarly, Guan & Oprea (2024) automate the implementation of the procedures (repeated game strategies) subjects select in repeated prisoner's dilemmas games – thereby removing the complexity of implementation – and find that this leads to the adoption of much more complex strategies. In a second set of treatments they show that calculating the payoff of strategies on behalf of subjects (thereby removing the complexity of identification) causes subjects to make far fewer mistakes in selecting optimal strategies in games. These treatments therefore provide evidence that both the complexity of implementation and of identification are major determinants of cooperative behavior in repeated games. In a similar vein, Bohren et al. (2024) show that recording past information in sequential sampling tasks (reducing representational costs in the problem) leads to significant changes in risky choices, establishing that these choices were distorted by representational complexity.

### 2.2.3 Manipulating Rewards

**Variation in Rewards**. Finally, it is sometimes possible to detect the influence of complexity by varying the incentives in the experiment, either by scaling up $\pi$ or intensifying its slope or loss function relative to $a^*$. To the degree subjects rationally trade off rewards and complexity costs, increasing the financial rewards of the experiment and in particular the rewards from submitting an optimal action, $\pi(a^*)$, will reduce the effects of complexity on the optimality of behavior. For instance, if a behavior is driven by cost complexity a'la equation (1), increasing the rewards from optimizing may increase the likelihood the subject finds it worthwhile to make an optimal decision

resources available to the typical subject!

by making the benefits from optimizing more likely to exceed its complexity costs.

A recent example of this approach is Alaoui & Penta (2015) who study the role (cognitive) cost / (financial) benefit tradeoffs play in strategic games. The investigation is motivated by a model in which players suffer cognitive costs from iteratively reasoning about how others reason, and trade this cost off against the payoffs from reasoning more deeply. In their experiment they show that, consistent with this kind of model, when the rewards in the experiment, $\pi$, are altered in such a way as to reward deeper reasoning, subjects reason more deeply: they choose actions that indicate more iterative rounds of costly reasoning. Alaoui & Penta (2022) expand this type of theoretical analysis to reasoning procedures more generally, and show under what conditions a reasoning problem can be represented as a procedural cost-benefit tradeoff of this sort. Another recent example is Caplin et al. (2020) who vary the rewards to a simple attentional task – estimating the number of multi-sided polygons that have seven vs. nine sides – and show that subjects' likelihood of correctly performing this task is sensitive to the rewards offered from doing so. Caplin et al. (2020) show that, consistent with a cost complexity account, performance responds positively to incentives, reducing mistakes. By independently varying the difficulty of the task, and thus $c(p^*)$, and the rewards in the experiment $\pi(a^*)$, and using classic ideas from production theory in economics and psychometrics from psychology, the authors are able to use variation in rewards to estimate complexity costs from the task and show how those costs increase as task difficulty increases. Caplin et al. (2020) thus show how variation in rewards can be used to measure subjective optimization costs and thus what we have called cost complexity. Overall, incentives have been found to influence performance in a diverse range of cognitive tasks including Stroop tasks (Krebs et al. 2010), selective attention (Padmala & Pessoa 2011) and even intelligence tests (Duckworth et al. 2011), thus linking errors in these tasks to complexity.

Nonetheless, there are clear limitations to the use of variation in rewards to identify and measure complexity. First, as Alaoui & Penta (2022) emphasize, because we do not know the cost function ex ante, evidence that a decision is not sensitive to financial incentives is not immediate evidence against a role for complexity in the task: the costs of the optimal procedure may be so large that the rewards varied in the experiment are not sufficient to motivate its adoption. Second, as Lipman (1995) discusses, making boundedly rational decisions that sensitively trade off complexity costs and rewards is itself potentially a highly complex task. For this reason, complexity needn't entail fluent sensitivity to cost-benefit tradeoffs and in some contexts complexity responses therefore may be insensitive to rewards. Third, in some contexts complexity is better modeled as a constraint than a cost. Optimal procedures that tax decision time, working memory, cognitive control or attention in excess of the subject's endowment of these resources will be unavailable to subjects, regardless of incentives. Thus, in very complex tasks or tasks where optimization requires unavailable resources, we should expect optimal behavior to be *insensitive* to rewards. Fourth, in some contexts (see e.g., Camerer & Hogarth (1999) and Gneezy & Rustichini (2000)) rewards may produce costs of their own by generating anxiety or distractions, worsening decisions in complex problems rather than improving them; see Alaoui & Penta (2022) for a thoughtful discussion. Finally, financial rewards may have a stronger effect on errors rooted in the complexity of implementation than errors rooted in the complexity of identification: it may be that rewards are more effective at inducing the DM to implement an already obviously optimal procedure, than it is at inducing people to discover or identy an optimal procedure they don't already know or that isn't immediately obvious. Osborn Popp et al. (2023) use standard category learning tasks to show that *cognitive discovery* (related to the complexity of identification) is less sensitive to incentives than execution of known rules (related to the complexity of implementation). Enke, Gneezy, Hall, Martin, Nelidov, Offerman & van de Ven (2023) show that several canonical biases, including base-rate neglect, anchoring and failures of contingent thinking, are largely insensitive to very large variations in incentives (i.e., amounting to a month's worth of wages), perhaps highlighting the large role the complexity of identification plays in these tasks.

## 3 Differentiating Complexity From Other Things

One of the most important uses of the tools discussed in Section 2, is the simple *detection* of the influence of complexity on a behavior and the differentiation of complexity from other potential explanations for that behavior. Although this kind of differentiation is central to *any* exercise aimed at measuring complexity, it is also itself an important end use of these methods for at least three reasons. First, as we will argue below, identifying a behavior as growing out of complexity has immediate implications for how we interpret the behavior: as a normative indication of what the decision maker desires (that a utilitarian social planner might want to accommodate) vs. a potentially correctible mistake (that a utilitarian social planner might want to prevent). Second, to the degree a behavior is driven by complexity it is potentially connected to other behaviors that are similarly driven by complexity, and therefore potentially understandable using parsimonious models that reduce the number of explanations for behavior required to conduct social science.[28] Finally, patterns of errors that are a consequence of the complexity of a task can potentially be eliminated by making those tasks simpler for decision-makers. Understanding a behavior's roots in complexity therefore is important for the design of policy and institutions – a central task of the social sciences.

In this section we consider the task of differentiating *complexity* from three salient alternative potential drivers of behavior. In Section 3.1 we consider the basic task of separating cognitive *mistakes* from the rational expression of the *tastes and preferences* of the decision maker – an important recent use of these methods especially in behavioral economics. In Section 3.2, we briefly consider some principles for designing experiments that effectively separate mistakes due to complexity from mistakes due to insufficient access to information. Finally, in Section 3.3 we briefly

[28] For example, Enke, Graeber, Oprea & Yang (2024) show that nearly 30 highly diverse economic decision tasks are distorted by complexity in a similar, predictable way, and that therefore many anomalies from behavioral economics that have been previously attributed to a wide range of domain-specific mechanisms are instead likely all instances of the same, generic complexity-driven pattern.

discuss the relationship between complexity and other cognitive frictions as drivers of mistakes.

### 3.1 Separating complexity from tastes and preferences

One of the most direct pieces of evidence that a task is complex is that it induces decision makers to make a clear mistake (as discussed in Section 2.1 above).[29] Consequently, one of the key challenges to identifying complexity is that it is not always easy to directly identify a decision maker's behavior as a mistake. To identify a subject's action $a$ as a mistake, the researcher first must show that the action is not optimal, i.e., is not equal to $a^*$. But optimal action $a^*$ can only be defined relative to $\pi$, the reward function the subject is trying to optimize over. In many task environments this reward function is induced *directly* by deterministically paying the subject based on the quality of her behavior relative to an objective benchmark. In such tasks, the researcher only needs to make the mild assumption that the decision-maker prefers more money to less, ceteris paribus, in order to identify what $\pi$ motivates the subject's behavior, allowing her to clearly identify mistakes when they occur. But in many choice tasks, the induced payoffs of the task do not fully specify the reward function that motivates the subject's behavior, because the task allows scope for subjects to express her own idiosyncratic tastes and preferences. For instance, in the face of common task characteristics like risk, temporal delay and social interaction there is scope for idiosyncratic, subject-specific risk, time and social preferences that shape $\pi$ but that we generally do not directly observe, meaning we do not directly have a $\pi$ against which to identify $a^*$. The immediate consequence is that we cannot tell whether a decision $a$ is an optimal response to those unobserved preferences rather than a complexity-derived mistake.

This matters for interpreting behavior in a variety of behavioral contexts, but it matters most sharply for efforts by behavioral researchers to *measure* subjects' preferences, $\pi$, a major task in social science experiments especially in behavioral economics. Consider, for instance the task of valuing (attaching a dollar value to) a lottery that pays off \$25 with probability 0.9 and \$0 otherwise. Tasks like this are frequently used to measure subjects' preferences for risk, and results from these experiments have been used as the basis of a number of theories of non-standard risk preferences like prospect theory (Kahneman & Tversky 1979, Tversky & Kahneman 1992). Suppose a decision maker values this lottery at something other than its expected value of \$22.50, e.g., at \$20. One possibility is that the lottery really is worth its expected value to the DM, but he has made a mistake, undervaluing the lottery. However another possibility is that the subject has risk preferences (i.e., tastes for risk) that lead her to truly value the lottery at less than its expected value, and is rationally expressing these "risk averse" preferences in choice. Indeed, preferences of this sort have been the dominant explanation for widely observed deviations from expected value and expected utility theory in the literature on risky choice. But the alternative possibility that the behavior is a consequence of complexity-derived mistakes, nonetheless, stands as an alternative

[29] Though, as we will argue in Section 3.2 and 3.3 below there are potentially contexts in which mistakes occur for reasons other than complexity.

explanation.

This indeterminacy is widespread in attempts by behavioral scientists to measure preferences or rationalize behavior in terms of preferences, influencing settings that involve risk (via risk preferences), ambiguity (ambiguity preferences), intertemporal choice (time preferences) and social behavior (social preferences). In order to understand the role complexity plays in behavior in settings like these – settings in which the subject can "bring her own $\pi$ into the experiment," – behavioral researchers have developed a number of methods for separating preferences from mistakes, deploying many of the techniques discussed in Section 2 above. In illustrating these methods, we will focus on the problem of eliciting preferences for risk, which has been the primary proving ground for these methods in recent years.

**Choice Inconsistency Correlations.** A first method for detecting the influence of complexity in preference elicitations, is to examine the correlation between a measured preference and choice inconsistency (e.g., noise). As we've discussed, behavioral noise can often be interpreted as evidence of the use of a simpler-than-optimal procedure, since imprecise computations, perceptions and representations are less costly for the brain to process. Noise of course is not a smoking gun for complexity: many approaches that have little to do with complexity (e.g., random utility models, models of preferences for randomization) likewise predict imprecise decision-making. However such alternative explanations for noise typically do not predict a systematic *relationship* between a subject's apparent preferences and inconsistencies in these choices. Complexity-based explanations like noisy coding models (e.g. Woodford 2020, Khaw et al. 2021, Frydman & Jin 2022) by contrast do. If subjects use noisy procedures, but attempt to limit the downside effects of this noise by shading their valuations of objects like lotteries towards a Bayesian prior or other default (Enke & Graeber 2023) or simply take steps to prevent those values from taking on implausible values (Blavatskyy 2007) this can produce systematic behavior that looks like anomalous risk preferences. Khaw et al. (2021) show for instance that small stakes risk aversion is strongly correlated with choice inconsistency, measured by comparing subjects' behavior in repeated instances of the same tasks), strongly suggesting that apparent risk aversion in these experiments is influenced by the use of simplified (noisy) evaluation procedures. Enke & Graeber (2023) and Oprea (2024) show that the same is true of probability weighting, another classic lottery anomaly often attributed to preferences.

**Belief Correlations.** A second method is to elicit subjects' beliefs (e.g., cognitive uncertainty) about the optimality of their own choices. Enke & Graeber (2023) ask subjects to value lotteries typically used to measure probability weighting, a central component of prospect theory. After each task, they asked subjects to state their cognitive uncertainty: the percentage likelihood with which they believe their choices in the task failed to maximize their own preferences. Enke & Graeber (2023) show that the severity of probability weighting is strongly predicted by cognitive uncertainty: for subjects who are highly confident that they made an optimal choice, this anomaly virtually disappear. Because cognitive uncertainty measures subjects' self awareness that their valuation procedure was not optimal, this evidence therefore strongly suggests that a major component of prospect theory is a response to the complexity of valuing lotteries rather than of tastes for risk.[30]

**Elicited Representations.** A third approach is to elicit subjects' beliefs about specific properties of the task, allowing the researcher to examine whether the DM's choices are rooted in false, simplified representations of features of the problem – an indication that the subject has used a procedure that economizes on representational complexity (e.g., working memory or attentional costs). Bohren et al. (2024) studies subjects' choices between pairs of complicated, 11-state lotteries. By also eliciting subjects' beliefs about the payoff distributions of these lotteries, they are able to show that subjects in fact systematically mis-represent lotteries to themselves – a finding that explains the highly inconsistent behavior they find across theoretically isomorphic choice environments in their experiment. Because measured false representations *predict* subjects' patterns of choice in these tasks, Bohren et al. (2024) are therefore able to show that those choices are driven by systematic mistakes rather than special risk preferences. What's more, Bohren et al. (2024) show that these false representations (and the inconsistent behaviors they generate) disappear when the number of states in the lotteries are reduced, suggesting that the false representations measured in the experiment and the mistakes they generate were driven by the complexity of these tasks.[31]

**Removing Scope for Preferences.** A fourth way of separating preferences from complexity-derived mistakes is to attempt to remove scope for preferences from a decision-task, while retaining most of what makes the task potentially complex. Oprea (2024), for example, tests the hypothesis that classic lottery anomalies are driven by complexity rather than behavioral risk preferences. To do this, in addition to having subjects value lotteries, he has subjects value what he calls *deterministic mirrors* of the same lotteries. A deterministic mirror is described exactly as a lottery, but the subjects is paid the expected value of the lottery. Doing this removes scope for the rational expression of risk- or loss- preferences, meaning any deviations from expected value in subjects' valuations cannot be driven by preferences. In particular, this task removes scope for subjects to express their own risk preferences ($\pi$) by inducing risk neutral preferences using a deterministic reward function. He finds that standard lottery anomalies like the fourfold pattern, reference dependence, loss aversion and probability weighting appear with similar intensity in mirrors as in lotteries, and anomalies in the two cases are strongly correlated. Because these tasks hold the information processing in the task constant, Oprea (2024) is able to conclude that classic lottery

[30]Enke, Graeber, Oprea & Yang (2024) shows that the same complexity-driven insensitivity to parameter variation that characterizes probability weighting, also arises in nearly thirty other decision settings: cognitive uncertainty thus generically predicts insensitivity to payoff-relevant parameters in a vast range of decision-making contexts.

[31]Collecting this data also can allow researchers to understand details of the simplified procedures subjects use and how these procedures are adapted to the choice environment (a'la Complexity Effect 5). Bohren et al. (2024), for instance, are able to show that when subjects observe all of the states of a lottery simultaneously, they respond to complexity by using procedures that overweight salient states, guided by bottom-up attention (Bordalo et al. 2012, 2022). By contrast, when the properties of lotteries are learned sequentially, they use procedures that overweight rare events, due to resource constraints in working memory. Elicited beliefs reveal that both of these simplified procedures are abandoned by subjects in simpler lotteries, leading to a disappearance of the procedural variance between the two environments.

anomalies are a consequence of complexity-derived mistakes rather than risk preferences. Enke & Shubatt (2023), discussed in more depth in Section 4.1 below, use a related method by examining what characteristics of lottery pairs predict failures to correctly identify their relative expected values, and showing that these same factors systematically predict behavior in lottery choice.

**Comparing to More Direct Elicitation.** A fifth method for separating preferences from mistakes is to compare behavior in standard choice tasks used to infer preferences to preferences elicited more directly than the task itself does. Nielsen & Rehbeck (2022) elicit subjects' preferences for abstract rules for choosing between lotteries that can be interpreted as elicitations of subjects' preferences for the normative axioms underlying expected utility theory. They then have subjects choose between a number of lottery pairs in which subjects are known to often fail to behave consistently with expected utility theory. They find that (i) subjects tend to prefer the normative axioms underlying expected utility theory but (ii) often fail to make lottery choices that comply with these axioms. In a final step, they inform subjects when their normative axiom preferences conflict with their lottery choices and give them an option of changing either their axiom or their choice. They find that when the two conflict, subjects typically change their lottery choice to comply with their axiom preferences. The natural interpretation of these results is that subjects' initial lottery choices were distorted not by preferences (directly elicited at the axiom level) but by complexity-derived mistakes. Benjamin et al. (2023) in contemperaneous work follow a related strategy and find similar results.

**Manipulating Procedure and Description.** A sixth method is to study the robustness of elicited preferences for risk to Manipulation of Procedure and Description. As discussed in Section 2.2.1, a long literature has shown that measured risk preferences tend to be unstable across elicitation methods and are uncorrelated or even weakly correlated with one another at the subject level – a failure of *procedural invariance* predicted by standard preference-based models (Friedman et al. 2017, 2022, Beauchamp et al. 2020, Bauermeister et al. 2018, Oprea 2024, Freeman & Mayraz 2019, Freeman et al. 2019). This serves as evidence that complexity influences lottery choice, because it suggests that subjects are not using procedures across elicitations, that in each case, maximize $\pi$. Instead subjects use procedures that respond to details of the elicitation in ways that fail to reveal a stable risk preference. For example Beauchamp et al. (2020) show that apparent risk aversion changes in commonly used multiple price list elicitations when researchers simply change the bounds of the list, while Freeman & Mayraz (2019) and Freeman et al. (2019) show that apparent risk aversion is higher in binary choice than multiple price lists. These papers therefore show that subjects use superficial procedures that are sensitive to details of the elicitation rather than optimal procedures that consistently maximize a stable reward function. A natural interpretation of such results is that robustly optimal procedures for valuing lotteries are too costly to identify or implement, leading subjects to use simpler procedures that are sensitive to details of the elicitation procedure instead.

**Manipulating Task Features.** Finally, as discussed in more depth in Section 4.1 below, the role complexity plays in a behavior is sometimes identified by attempting to manipulate the complexity of the task itself. For instance, a role for complexity in lottery valuations has been identified by studying the effect of adding outcomes to the lottery (Huck & Weizsacker 1999, Bernheim & Sprenger 2020, Puri 2023, Bohren et al. 2024), turning a simple into a compound lottery (Wilcox 1993) or re-describing lottery primitives using complicated mathematical expressions (Enke & Graeber 2023). This method has the limitation that it is typically easier to increase the complexity of an object of a lottery than to decrease it, meaning this approach often can't firmly establish that the seemingly simplest lotteries (e.g., non-compound lotteries with only two outcomes) are influenced by complexity. But they can show that behaviors that appear in simpler lotteries intensify as the information processing required to value them increases, suggesting a link between complexity and the original behavior.

**Combining Methods.** In practice, these methods are particularly powerful when combined in one experimental design, producing mutually reinforcing evidence. Enke, Graeber & Oprea (2024) combines (i) choice inconsistency measurements, (ii) cognitive uncertainty measurements, (iii) attempts to manipulate complexity and (iv) removal of the availability of alternative mechanisms (i.e. task mirrors)[32] to study the role complexity plays in hyperbolic discounting, a key anomaly in intertemporal choice tasks. They show that hyperbolic discounting (i) is strongly correlated with choice inconsistency, (ii) highly correlated with cognitive uncertainty, (iii) intensifies significantly when efforts are made to make the task more complex and (iv) that very similar levels of hyperbolicity occur in atemporal mirrors (featuring no actual time delays) as in true intertemporal choice, and that the two instances of hyperbolicity are strongly correlated. All four methods strongly suggest that hyperbolicity is primarily a complexity-driven mistake, and quantitative decompositions based on these approaches generate extremely similar conclusions about the size of this effect across these methods.

### 3.2 Separating complexity from imperfect information

Even when a behavior can be identified as a mistake by ruling out preference-based explanations, it is nonetheless possible for the mistake to be driven by factors other than complexity. Most importantly it is possible for a mistake to be a consequence of bad or incomplete information. In particular if the decision-maker does not have sufficient information to understand the mapping between action and reward, $\pi$ – or if she is given mis-leading information about what $\pi$ is – she might choose an $a \neq a^*$ for reasons that have nothing to do with the costs of information processing. This means that in order to experimentally identify complexity, the experiment must be designed so that if the decision maker had no cognitive resource constraints (e.g., unlimited time, working memory, attention and tolerance for cognitive effort) she would be expected to choose $a = a^*$. This

[32] They do this using "atemporal mirrors" that remove true time delay from intertemporal choice problems, and instead induce exponential preferences in their subjects. In particular instead of asking subjects to value, e.g., $50 paid in 12 months, they ask subjects to value $50 paid immediately but shrunk 12 times, each time by 4%.

suggests two experimental design principles when attempting to measure complexity. First, the experiment cannot be imprecise or vague about the mapping between actions and experimental rewards, e.g., $\pi(\cdot)$. Otherwise, the subject is forced to fill in the mapping between actions and rewards herself, meaning the decision maker might in fact be optimizing according to this mental representation or model, making it difficult for the researcher to credibly attribute the mistake to complexity. Second, the experiment cannot rely on subjects' prior knowledge or beliefs about facts or relationships outside of the context of the experiment. For instance experiments that study subjects' beliefs about pieces of trivia or statistical relationships from the real-world produce behaviors that are difficult to crisply attribute to complexity. The reason, again, is that the subject's beliefs about the mapping between action and reward may be responsible for the mistake, rather than information processing costs.

### 3.3 Complexity and other cognitive frictions

What is the relationship between complexity and cognitive frictions that are often held responsible for mistakes in seemingly simple tasks? This is an important open question in the literature, and an intriguingly complicated one. Given this, I will make only two brief, contrasting points that I hope illustrate the richness of the question and the appeal of further research answering it. First it is important to emphasize that "complexity" as I've defined it can shape behavior and induce apparent mistakes even in seemingly simple tasks. In many settings even if optimal behavior is not terribly costly, it can nonetheless be costly enough to induce the use of simplified procedures.[33] Thus the apparent simplicity of a task is not always a reliable guide to the role complexity plays in the mistake: even small cognitive costs can motivate the use of less than optimal procedures.[34] Second, and conversely, even when complexity *is* a contributing factor to a mistake, it need not be the only or primary cognitive factor responsible and therefore may not be the most useful first-order lens under which to analyze the mistake. For example, many mistakes occur due to the appeal of compelling mental models (e.g. Gagnon-Bartsch et al. 2023) or intuitions (e.g. Frederick 2005) that lead the DM to misconceive the problem or the nature of an optimal solution *and not realize her misconception*, i.e. what Rabin (2013) calls "astray errors" in contrast to "bounds errors." Such errors may be related to complexity costs (e.g., they may not have occured if the DM had unlimited cognitive resources to interrogate their pre-conceptions and intuitions), but may

[33] For instance behaviors like default effects, anchoring effects, decoy effects, hysteresis or apparent randomization that often arise in seemingly simple tasks may sometimes simply be ways for the DM to avoid the costs of decision-making altogether (a'la complexity Effect 4, aversion) by relying on external cues instead. One of the reasons complexity was not identified as a major driver of risky and intertemporal choice anomalies (Section 3.1) is arguably that the tasks in which these anomalies are usually documented *seem* simple.

[34] One reason for this is that many seemingly severe mistakes are in fact not that costly in terms of task rewards: often large divergences from the optimum in *action space* are small divergences from the optimum in *payoff space* (i.e., $\pi(a^{**})$ can be very similar even when $a^{**}$ and $a^{*}$ are very different). See Harrison (1989) for an early discussion. Another is that, even when this is not true, decision-makers need not necessarily optimally trade off costs and benefits when deciding when to economize on cognitive costs for reasons discussed in Section 2.2.3.

nonetheless generate very different behavioral effects and sensitivities than those described in this essay because they are rooted in compelling, distorted beliefs about the nature of the optimum.[35,36]

# 4 Mapping Complexity

In the previous section we discussed the application of the methods developed in Section 2 to the task of *detecting* the influence of complexity on behavior. In this section we discuss the application of these same techniques to the richer problem of *mapping* complexity, by which I mean empirically establishing what characteristics of tasks and procedures produce or *predict* complexity costs. Sometimes this enterprise involves attempting to directly measure the resource costs that constitute complexity (e.g., subjective costs or time costs) as a function of task or procedural characteristics; other times it involves less directly examining the way the intensity of the behavioral effects of complexity we highlighted in Section 1.2 respond to those same characteristics. We will review both lines of inquiry in this section. To organize our discussion, we break this literature into two parts. In Section 4.1, we review a literature that attempts to map complexity according to descriptive features of *tasks*, such as the "size" of the task or approximate metrics of the amount of information processing required of the tasks. In Section 4.2 we review a complementary literature that attempts to deepen this kind of analysis by mapping complexity according to formal characteristics of the *procedures* or algorithms the decision-maker uses (or should optimally use) to tackle the problem.

## 4.1 Mapping at the Task Level

We begin with a literature that aims to predict and describe complexity at the *task level*, i.e., in terms of descriptive elements of the decision problem itself. This literature's goal is to uncover

[35]For instance, because misleading intuitions or miscalibrated mental models involve distortions of the DM's beliefs about the nature of the problem, they may fail to generate self-selection away from tasks (a'la Effect 4), resist measurement techniques rooted in beliefs (Section 2.1.3), and be less sensitive to reward manipulations (Section 2.2.3) than mistakes more directly rooted in complexity. Because of this one way to distinguish mistakes driven by misleading intuitions from those driven more directly by complexity is to elicit subjects' beliefs about (confidence in) the optimality of their own choices (i.e., cognitive uncertainty). To the degree a behavior is driven by a compelling intuition or mental model, we would expect this confidence to be uncorrelated with – or even inversely correlated with – the error. Enke, Graeber & Oprea (2023) measure and calculate this correlation for fifteen well-studied cognitive errors documented in the prior literature. Some tasks show a negative or zero correlation between confidence and decision-quality, suggesting that poor intuitions may play a role. However, for nearly 2/3 of the tasks they study, this correlation is positive, suggesting that mistakes in most of the tasks they study are likely driven directly by complexity rather than bad intuition.

[36]On the other hand, in some settings misleading intuitions or bad mental models may themselves sometimes be productively analyzed under the lens of complexity because the complexity of identifying an optimal approach to the problem (e.g., an optimal procedure) is sometimes the root cause of a mistake. Kendall & Oprea (2024) provides evidence from pattern-recognition forecasting problems suggesting that subjects are better at forming mental models in simpler settings than complex ones and that they prefer to adopt simpler mental models than complex ones when both can be justified.

descriptive characteristics of tasks that tend to predict complexity responses like those discussed in Section 1.2 and to gather clues about the nature of the information processing costs responsible for complexity and its behavioral effects. Identifying predictors at the task level has two basic values. First, it allows us to link behaviors to complexity and predict how changes to the task environment will change these behaviors. Second, it serves as a valuable source of clues as to what cognitive acts are difficult for subjects and therefore serves as a guide for the more fine grained style of analysis discussed in Section 4.2, below.

An obvious challenge with measuring complexity based on task characteristics is that, *ex ante*, the number of task metrics that might predict complexity in a choice problem is potentially very large. Famously Lloyd (2001) listed over 30 characteristics of problems or procedures that have been advanced as "complexity metrics" – a list that was nonexhaustive then and that is even less exhaustive now. It is therefore important to emphasize that this literature's goal is not to identify one "objective" theoretical metric of complexity and test its implication. Instead its goal is to identify metrics at the relatively coarse level of the task that tend to predict complexity responses.[37] In practice, attempts to identify task-level predictors of complexity in the experimental literature have rested on intuition (sometimes formalized in models) about what features of a task likely increase the volume of information processing required in the task, and thus the cost of the procedure required to optimally behave in that task. An exhaustive review of this type of research is beyond the scope of this methodological essay, but I will discuss some examples to give a flavor of this important line of research.[38]

[37]Many papers on complexity in the literature are framed as "tests" of the effects of "complexity" on behavior. In such papers, often an intuitive metric of complexity is proposed (e.g., the number of pieces of information to be processed) and treated as an objective marker of task complexity. Such papers use this to motivate designs that attempt to vary the complexity of the task by varying this marker, and test for a behavioral response. However, our definition of complexity suggests that we should be cautious in attempting to define objective task characteristics ex ante as "complexity," and "testing" whether complexity, so defined, induces a behavioral response. Until we articulate the way the putative complexity driver adds resource costs to the algorithm required to optimize in the task, we do not know whether or to what degree that task characteristic serves as an effective metric of complexity. For this, computational theories that articulate these costs explicitly by modeling the algorithms required to optimize and how the resource costs these algorithms generate change as task features change (such as computational complexity theory, discussed in Section 4.2.4) are invaluable.

For what we have called "cost complexity" (complexity described in terms of effort costs denominated in terms of the reward medium of the task) even this kind of careful mapping need not be sufficient, since cost complexity describes a subjective response to the task. In this case, we cannot, on the basis of purely ex ante classifications, "test" whether "complexity" changes a behavior, since complexity (in the case of cost complexity) is defined precisely in terms of its capacity to generate an aversive response in the subject. Cost complexity, in other words, is an empirically measurable subjective response to the algorithmic requirements of a task (a cost attached to optimizing), not an objectively articulable, ex ante mathematical property of a task. I read the "testing" framing of some empirical work in this area as a shorthand for what our discussions so far suggests such exercises are really doing: testing whether some task-level characteristic, hypothesized to be linked to complexity, in fact generate subjective costs sufficient to alter behavior.

[38]An early interdisciplinary literature (spanning management, accounting, education and several other applied fields) proposed a series of taxonomies of the complexity of tasks, with an eye towards categorizing practical work tasks. An early, influential contribution, Campbell (1988), proposed that tasks should be categorized according

**Task size effects.** By far the most common intuition about task is that the *size* of the task influences the complexity of the task (Payne et al. 1992). If, the intuition goes, two choice problems are similar (e.g., induce similar optimal action, $a^*$) but differ in either (i) the number of pieces of information in the description of the problem that the DM has to process to optimize or (ii) the number of choices available to the decision maker, the problem with higher (i) or (ii) is likely to be more complex. After all, it seems intuitive that mentally representing larger objects requires more attention and working memory (increasing representational complexity) and performing mental operations on richer representations requires more time and energy (increasing computational complexity). This intuition is often sound in a formal sense, since it is often the case that tasks that contain more information require more elaborate procedures to solve optimally, consuming more time, working memory and producing more subjective cost for the decision-maker. Indeed, as we discuss below in Section 4.2.4, the size of a task is one of the key objects of study in computational complexity theory since instance size (the size of the task) typically increases the time complexity of tasks, and sometimes other algorithmic metrics of complexity like space complexity as well. Similarly, as we will discuss in Section 4.2.2, the number of elementary information processes (another formal scheme for measuring characteristics of procedures that is known to predict complexity costs) required for optimal procedures increases in the number of alternatives and attributes in multi-attribute choice tasks. Thus the standard intuition that increasing task size increases complexity is broadly supported by finer grained, procedural measurements of complexity discussed in the next subsection.

For example, there is a growing literature on the complexity of lotteries that identifies the size of the lottery – the number of distinct outcomes it can produce – as one major determinant of its complexity. Experiments documenting this typically compare how people evaluate two lotteries with similar payoff-relevant characteristics (for instance similar expected value, spread, variance etc.) but whose supports include different numbers of outcomes (for instance a two outcome lottery vs. a four outcome lottery). A number of studies show that subjects tend to undervalue lotteries with more potential outcomes, choosing them less often in binary choice tasks and assigning lower dollar values to them in elicitations of certainty equivalents (Huck & Weizsacker 1999, Bernheim & Sprenger 2020, Puri 2023). Puri (2023) axiomitizes these effects, describing them as deriving from complexity costs suffered when choosing a lottery with more outcomes, and gathers systematic evidence consistent with the theory that subjects have a distaste for lotteries that are complex in this way, i.e., display complexity aversion (Effect 4).[39] Enke & Shubatt (2023), discussed in more depth below, provide evidence that the number of outcomes in a lottery induces not systematic aversion, but instead the use of decision procedures that attenuate responsiveness to a lottery's

to whether they provide multiple paths to desired outcomes, multiple desired outcomes, conflicts between desired outcomes and uncertainty, producing 16 categories of complexity. See Liu & Li (2012) for a more recent overview of this highly applied strand of the literature.

[39]Interestingly, this effect only occurs if the outcomes from lotteries are actually distinct in the sense that they produce different payoffs. Adding outcomes that are descriptively separated but produce the same payment consequences does not seem to produce this kind of aversion.

expected value, producing apparent aversion to lottery outcomes in some lotteries and apparent *preference for* more lottery outcomes in others. A different way of increasing the size of a lottery is to transform a simple lottery into an isomorphic compound lottery – a size manipulation that produces complexity responses including overt mistakes, probably because it increases the amount of information in the description of the lottery that must be attended to and remembered to accurately represent it, and the number of calculations required to value it. See Wilcox (1993) for an early paper using this manipulation to increase complexity.

Task size manipulations have been shown to generate complexity responses in a number of other contexts as well. Ba et al. (2024) show that increasing the number of states in Bayesian updating tasks produces greater over-reaction to signals via a complexity effect. Guan et al. (2024) show that increasing the number of signals or distinct posteriors induced by an information structure predicts valuation errors that are consistent with a complexity effect. A number of papers show that errors in processing information increase with the volume of information – a result that is predicted by rational inattention models of complexity costs (e.g. Caplin et al. 2020). Abeler & Jäger (2015) show that decision makers become less sensitive to tax incentives the more individual components a tax rule has – a result that is consistent with a number of simpler-than-optimal procedures. Search problems (Caplin et al. 2011) and multi-attribute search problems (Payne et al. 1993, Gabaix et al. 2006) produce more heuristic behavior, more mistakes and longer response times when they contain more options or when available options contain more characteristics to process. Relatedly, a large body of research, reviewed and meta-analyzed in Chernev et al. (2015), shows that in some contexts increasing the number of choice options (i.e., the action set $A$) produces reductions in decision quality (via "choice overload") that are consistent with such tasks being more complex.[40]

**Processing effects.** Of course, task size only scratches the surface of task characteristics that likely predict procedural costs and thus complexity. This is because it is not only the *volume* of information that must be processed (proxied by task size), but also the *way* information must be processed to arrive at an optimal choice that shape the costs of an optimal procedure. For instance tasks that require multiple decisions that are asymmetric tend to be more difficult than those that require decisions that are symmetric (Benartzi & Thaler 2001), and tasks that require non-linear responses to a parameter tend to be more difficult than those that require linear responses (Rees-Jones & Taubinsky 2020). Subjects, when faced with such tasks, tend to simplify by substituting to the use of symmetric or linear procedures rather than optimal asymmetric or non-linear ones, seemingly in order to simplify their behavioral response. In both cases, it is intuitive that the optimal procedure for making choices requires more sensitive and differentiated responses to information when the optimum is non-linear or asymmetric, and therefore requires more intensive computation, making them more costly to implement.

[40] Ortoleva (2013) axiomatizes a model of "thinking aversion" that describes how distortions in behavior of these sorts can arise from costs decision makers attach to the size of the option set.

Another subtle characteristic of tasks that seems to robustly predict complexity responses is what is sometimes called "tradeoff complexity": the degree to which the DM has to trade-off competing considerations when making decisions. The idea is that when one choice option is dominant (preferred) in all respects, it is easy to make a choice, but when different options are dominant on different dimensions the DM has to consider the tradeoffs between the relative importance of these competing considerations, making the task complex. For instance if one lottery first order stochastically dominates a second, it is preferred in all states of the world making it is relatively easy to choose the first lottery. But if one lottery pays more in one state and less in another, the decision maker faces the difficult task of assessing the relative importance of each state when choosing between lotteries, increasing the computation burden on the DM. Shubatt & Yang (2024) show that options that are similar on more dimensions are less tradeoff complex, and verifies that several empirical measures of complexity discussed in Section 2 intensify as lotteries, time-dated payments and multiattribute goods become more tradeoff complex in this sense. Using this insight, they show that altering tradeoffs can reverse classic anomalies like probability weighting and hyperbolic discounting (see also Rubinstein (1988) and He & Natenzon (2023) for related insights). Enke & Shubatt (2023) calculate an "excess dissimilarity" metric (the state-wise dissimilarity between lotteries in excess of the difference in expected value) in order to measure tradeoff complexity and shows that it is the single largest predictor of complexity responses in lotteries (see the discussion below). Enke, Graeber, Oprea & Yang (2024) document evidence of tradeoff complexity effects in dozens of distinct experimental settings, in contexts ranging from investment and savings decisions to taxes and policy choices, and from labor supply and product demand decisions to strategy formulation in games. They show that even within-task, complexity responses intensify when the requirement to make tradeoffs becomes more severe – a pattern that Enke, Graeber, Oprea & Yang (2024) show is responsible for the classic anomaly of diminishing sensitivity. See Shugan (1980) for an early discussion of similar issues and an explicit discussion of how tradeoffs produce (potentially measurable) procedural costs.

**Empirical horse races.** Because of the many possible predictors of complexity available in the typical choice task, researchers have turned to experiments designed to compare the performance of multiple predictors – a "horse race" research strategy that seems especially promising. I will discuss two recent and very stylistically different examples. First is Kendall & Oprea (2024) who focus on what we have called the complexity of identification (of an optimal procedure) in simple forecasting tasks. Subjects are shown values for a pair of variables, $x$ and $y$ and are asked to extract the causal algorithm (the rule) linking those variables so that they can predict future values of $y$ from values of $x$. Subjects are then asked to (i) describe the forecasting rule they've extracted from the data verbally and (ii) procedurally implement that rule in forecasting future values of $y$ given information on $x$. This allows Kendall & Oprea (2024) to characterize the forecasting procedure the subject actually identified from the data and compare it to the true algorithm generating the data. Kendall & Oprea (2024) find that subjects fail to identify the correct forecasting procedure much of the time, suggesting the task is complex, but that subjects are no worse at implementing the procedure than in verbally describing it. This suggests that the complexity of the task is rooted primarily in the costs of identifying the correct procedure, not in the costs of implementing it. Kendall & Oprea (2024) find a great deal of variation in complexity (revealed via mistakes) across different tasks, and the heart of the paper is an effort to understand what formalizable features of the data generating process (the correct forecasting algorithm connecting $x$ and $y$) predicts that complexity. To do this Kendall & Oprea (2024) compare the predictive power of a large number of formal metrics of complexity, drawn from computer science, information theory and economics including characteristics of finite automata descriptions of the procedure (states and transitions), machine-based metrics based on input gating, entropy, mutual information, Kolmogorov complexity etc. The paper finds that two closely related metrics of the data processing required to use the procedure to forecast – *partition complexity* (Lipman 1995) and *sparsity complexity* (Gabaix 2014) – strongly predict mistakes in identifying the optimal forecasting procedure in the data.[41]

Enke & Shubatt (2023) compare dozens of features and metrics of lotteries to identify which ones predict complexity responses – in particular, which features of tasks predict a failure on the part of the decision maker to choose their most-preferred lottery when given a choice between two. In doing this, Enke & Shubatt (2023) run against a fundamental problem, discussed in some depth in Section 3.1 above: because we typically cannot observe subjects' own preferences for risk (i.e., $\pi$) we cannot easily identify mistakes of this sort. To get around this Enke & Shubatt (2023) use a variation of a technique ("Removing Availability of Alternative Mechanisms") discussed in Section 2.2.1 by asking subjects not to choose between lotteries, but instead to identify which of two lotteries has the higher expected value – a question with an objectively correct answer. The idea is that the information processing required to identify which of two lotteries pays more on average is closely related to the processing required to identify which of two lotteries is more preferred.[42]

In the first step of their experiment, Enke & Shubatt (2023) have subjects identify which lottery has higher expected value for over 2,200 of pairs of lotteries and deliberately vary a large number of formal characteristics of these pairs. This allows them to examine the effects of a large set of characteristics including not only the number of distinct outcomes in individual lotteries (the focus of much of the prior literature), but also the presence of compounding, dominance relationships, the scale of payoffs, the presence of mixtures of gains and losses, similarities in expected values, and the outcome-wise similarity between the two lotteries. They then use a LASSO procedure to identify what characteristics predict both objective mistakes and subjects' beliefs about the quality of their own choices, building a predictive complexity index. In the second step, they use this index to predict true lottery choices using a large dataset consisting of over a million decisions and over 10,000 lottery choice problems.

[41]Interestingly, Kendall & Oprea (2024) also find that in tasks in which more than one forecasting procedures is equally good, subjects are *biased* towards identifying procedures that are simple in this same sense. This strongly reinforces the conclusion that such procedures are indeed less complex to identify.

[42]The idea that these two tasks – choosing between lotteries and identifying which lottery has a higher expected value – have similar complexity seems supported by the estimates of EIPs from Payne et al. (1988), discussed in Section 4.2.2 below.

The paper finds that a number of intuitive metrics of complexity predict mistakes (e.g., number of outcomes, compoundness) in the first stage of the experiment and, crucially, they can quantify the relative strength of these metrics in producing complexity responses. The most predictive measure is a metric of "excess dissimilarity" between the lotteries: the distance in the CDFs between the lotteries being compared in excess of the difference in expected value. Enke & Shubatt (2023) use this evidence to suggest that tradeoff complexity (discussed above) is the single most important driver of complexity-derived error. Finally, the paper uses the resulting index to interpret true lottery choice. They show that as the complexity index (formed, recall, based on evidence of unambiguous mistakes when comparing the expected values of lotteries) rises, subjects' ability to discern the relative desirability of lotteries falls: rates of selection of the higher and lower expected value lottery attenuate, converging to 50/50 as the complexity index grows large. As in many other settings, a primary effect of complexity is to produce behavioral attenuation – underresponse to features of the choice task that matter for optimal response (Enke, Graeber, Oprea & Yang 2024). They also show that choice inconsistency (one possible procedural response to complexity highlighted in Section 2.1) rises as the index rises.[43]

## 4.2 Mapping at the Procedural Level

Finally, we discuss a complementary research agenda aimed at "mapping the procedural cost function" – i.e., at understanding what makes one algorithm for processing information into a behavior more costly than another for humans. This is in some sense a more fundamental use for the methods we've discussed, because it involves reasearch aimed at understanding the primitives underlying any complexity-based explanation for a behavior: the cost of using or finding optimal procedures, and the economies that come from substituting to simpler-than-optimal procedures instead. While there is clear value in understanding what coarse descriptive features of *tasks* tend to be associated with greater complexity (as just discussed is Section 4.1), there is a clear intellectual and practical value in understanding complexity at the more fundamental level of the *procedure*. First, to the degree we can understand what makes algorithms complex (costly) for humans, we may be able to build predictive models of complexity that are domain-general, applying to behavioral questions in a wide range of choice settings using a single descriptive model of complexity costs. Second, understanding what makes procedures complex, may allow us to predict not only what tasks subjects will fail to optimize at, but also to understand what simpler-than-optimal procedures subjects are most likely to revert to instead – and crucially how this substitution is shaped by features of the choice environment that behavioral scientists often ignore. Thus, a real behavioral science of complexity depends to a great extent on understanding how complexity shapes behavior at this deeper level of description.

[43] As mentioned above, these results suggest that standard lottery complexity effects are not driven by a systematic aversion to complex lotteries, but rather to the distortionary effects of the insensitive procedures subjects increasingly use as lotteries become more complex.

### 4.2.1 Taxonomies of Algorithms

What is most important to understand about this line of research is that any effort at procedural measurement is unavoidably conducted relative to some choice about how to describe and differentiate procedures/algorithms from one another. Indeed, any given procedure/algorithm can be taxonomized in any of a large number of ways, using any of a large number of possible descriptive schemes. Computer science, for instance, includes a number of formal "languages" for describing or modeling algorithms, any of which can be used to model or describe any given information processing procedure. These descriptive languages differ in how they group similar computational acts, how finely they distinguish between these acts and how they summarize representations and computations in terms of formal descriptive components. When we attempt to empirically understand what makes one procedure more complex than another, we necessarily conduct measurement relative to one of these descriptive languages.

To illustrate, consider the famous satisficing procedure – a simple procedure for choosing from a list of options that Simon (1955) hypothesized people use to avoid the complexity costs associated with the rational alternative of exhustively evaluating the options in the list. While rational choice requires the DM to evaluate every item on the list one-by-one (and compare them to one another), satisficing requires the decision-maker only to search through the list of items until she reaches the first one that is is "good enough" (i.e. meets some threshold of value). The literature has procedurally modelled the satisficing procedure and contrasted it to the exhaustive rational choice rule in at least three distinct ways, using three different formal languages for describing and differentiating algorithms:

- Payne et al. (1993) describe the satisficing rule in terms of the number of "**elementary information processes**" or EIPs (individual computational steps) it requires, and showed that satisficing can be described using fewer EIPs than rational choice.
- Salant (2011) models the same rules instead as **finite state machines** (simple models of algorithms from computer science), and shows that satisficing requires fewer "states" and "transitions" (the two key mathematical components of finite state machines) to describe than the rational choice rule does.
- Sanjurjo (2024) describes satisficing and rational choice as **Turing machines**, the richest class of algorithms from computer science, and shows that satisficing requires less working memory space (a key mathematical property of Turing machines) than rational choice does.

Under the lens of these formal characterizations, we can describe in what procedural sense satisficing might be less complex than rational choice under each of these descriptive schemes. If it turns out empirically that (i) procedures requiring more EIP's tend to be more costly for humans, (ii) procedures requiring more states and transitions to describe tend to be more costly for humans or (iii) procedures that are more space complex when described as Turing machines tend to be more costly for humans, then we can *predict* that the satisficing procedure will tend to be less complex than the rational choice procedure, just as Simon (1955) hypothesized. And indeed, as we will see below, empirical work attempting to map complexity under each of these three taxonomic lenses (EIPs, finite automata, Turing machines) has produced direct evidence supporting each of these regularities. In particular, measurement exercises in the literature reviewed below have shown variously that procedures that require more EIPs tend to be more costly for humans (Section 4.2.2), procedures describable by finite automata with more states and transitions tend to be more costly for humans (Section 4.2.3), and that procedures describable by Turing machines that require more memory space are more costly for humans (Section 4.2.4). We can therefore *predict on a procedural basis* that behaviors like satisficing should be less complex (costly) than rational choice, using general measurements of the predictors of procedural costs that were conducted without any direct reference to the satisficing procedure.

In this section we will discuss each of these three efforts to map procedural complexity in turn. Again, what differentiates these strands of research is not their motivating question – all are aimed at measuring or detecting procedural costs at an algorithmic level – but rather the formal language used to describe and taxonomize algorithms. Because of this, these examples should not be taken as an exhaustive list of possible ways to taxonomize procedures. Theoretical computer and cognitive science are treasure troves of formalizations for describing algorithms, and empirically exploring them and their relative suitability for describing, organizing, measuring and predicting procedural costs is a major task for this literature and one that has barely begun. An important, open meta-question for this literature is what descriptive schemes (of the many available) are most useful for organizing and predicting complexity responses at the procedural level.

### 4.2.2 Elementary Information Processing

The oldest effort to taxonomize and measure the costs of procedures that I am aware of is a literature in cognitive science that attempts to break procedures down into a collection of primitive computational acts – called "elementary information processes" or EIPs (Newell & Simon 1972, Johnson & Payne 1985) – and use the number (and kinds) of EIPs required by a procedure as a way of taxonomizing procedures (Payne et al. 1993).[44] The literature applies this idea especially to a canonical class of problems in which the DM's task is to choose one of $M$ options ("alternatives") from a list, each of which has $N$ numerical values ("attributes") that collectively describe that alternative's value to the DM. Typically, the DM's reward, $\pi$ for choosing an alternative is the

[44]Unlike the other taxonomizing descriptive languages considered below, the EIP notion is not directly motivated by models of algorithms from computer science. However it has some relationship to a language for describing algorithms in computer science called Halstead complexity (Halstead 1977). Colliard & Georg (2023) used Halstead measures to classify the complexity of regulations, and report experiments in which subjects were asked to make decisions that comply with actual financial regulations. They found that systematic components of Halstead complexity predict compliance mistakes and complexity costs as measured by response time.

weighted sum of its attributes. Thus, the optimal procedure in these tasks, $p^*$ is to compute the weighted sum of each alternative's attributes, and pick the alternative with the highest sum – a procedure the literature calls WADD for "weighted additive."[45]

The literature taxonomizes procedures like the optimal WADD rule by pointing out that they can be broken down into and described as a sequence of "elementary information processes" or EIPs that summarize the individual cognitive acts required by the procedure, and hypothesizes that the complexity of the procedure is given by some weighted sum of the number of EIPs required to describe it. A typical set of EIPs might include: READ (place a value in short term memory), MOVE (move attention to consider another value), COMPARE (compare two values), DIFFERENCE (subtract one value from another), ADD (add two values stored in memory), PRODUCT (multiply two values together), ELIMINATE (remove a choice option from consideration) and finally CHOOSE (make a choice).[46] For instance, a DM choosing between a pair of two-outcome lotteries who wants to use the WADD procedure to choose the lottery with a higher expected value must: (i) read the probability and payoff from the first outcome (requiring two READ EIPs), (ii) multiply them together (requiring the PRODUCT EIP), (ii) do the same for the second outcome and add the result to the first (the ADD EIP), (iii) repeat (i)-(iii) for the second lottery and (iv) compare the two weighted sums (COMPARE) and finally (v) choose whichever lottery's weighted sum is higher (CHOOSE). Altogether then the WADD procedure might require eight READ EIPs, four PRODUCTs, two ADDs a COMPARE and a CHOOSE, requiring in total 16 EIPs (though in practice this count might depend on the exact way the WADD rule is modeled and which EIPs are used to model it).

A key observation of the literature is that to the extent the number of EIPs do indeed summarize cognitive costs, the optimal WADD procedure is particularly complex – it requires a particularly large number of EIPs to implement because implementing it requires the DM to observe and aggregate together every attribute of each alternative. By contrast, many famous, suboptimal heuristic procedures discussed in the literature require considerably fewer (and often very different) EIPs than WADD does including, e.g., the lexicographic choice rule (LEX, choosing based only on the most important attribute), elimination by aspects (EBA, comparing and eliminating alternatives first based on the most highly weighted attribute, then the next etc.) and satisficing (SAT, consider each alternative one-at-a-time and select the first that has attributes that satisfy some cutoff rule).

In order to measure the complexity of procedures and map that complexity to EIP counts, Bettman et al. (1990) uses what I call an "assigned procedure" design (like those used by Oprea (2020) and Sanjurjo (2015) below). As we discuss in Section 2.1.1, the difficulty in measuring complexity costs directly using conventional experiments is that in conventional designs subjects *choose*

[45] A lottery is an example of a problem of this class. Calculating the expected utility of a lottery involves calculating the sum of payoffs weighted by probabilities and some preference-specific weight on each payment.

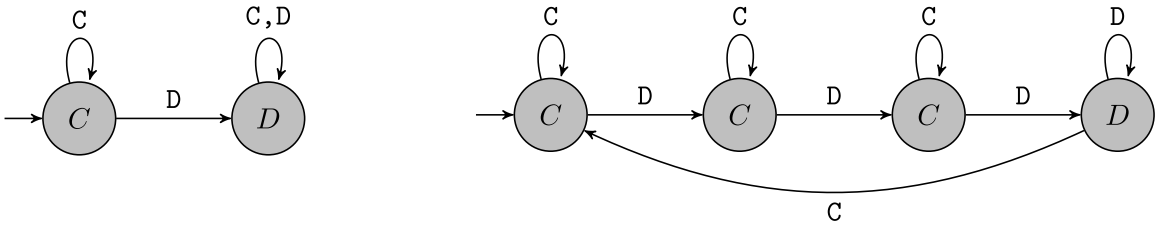
[46] Recent work (e.g. Fechner et al. 2018) has extended this approach by rooting EIPs in terms of yet more primitive cognitive resources like working memory and perceptual resources, and by structuring them in terms of a popular model of cognitive architecture (ACT-R).

their own procedures, leading to serious observability and endogeneity problems that counfound direct measures of procedural cost. To get around this, Bettman et al. (1990) assigns subjects a series of multi-attribute search problems and asks subjects not to optimize in the task by choosing their procedure, but instead to implement one of six *assigned*, classic procedures including, e.g., WADD, EBA and SAT. To verify that subjects are properly following the assigned procedure, Bettman et al. (1990) used a Mouselab design (see Section 2.1.2) that tracked the order in which subjects looked up information, moved between pieces of information etc. The basic finding of the experiment is that, as assigned procedures and tasks change, procedural complexity (as measured both by response time and self-reported effort) change in structured, predictable ways that are extremely well-described by EIPs: ignoring high level descriptive features of the task (like the specific procedure assigned or the number of attributes or alternatives in the choice problem), and focusing exclusively on EIP counts allows the researchers to strongly predict complexity responses. What's more, the estimated complexity generated by individual EIPs do not vary across assigned procedures or tasks, suggesting that EIPs do a reasonably complete job of summarizing what makes one procedure/task more complex than another. Based on these estimates, the researchers are able to show that optimal WADD procedures are, indeed, substantially more complex than heuristic procedures like EBA and SAT, meaning these measurements rationalize the use of these alternative procedures as attempts by subjects to economize on complexity costs.

To illustrate the value of this approach for understanding behavior, Payne et al. (1988) apply this methodology to lottery choice (i.e., alternatives are lotteries and attributes are probability-weighted outcomes), and use the measurement of the costs of EIPs from Bettman et al. (1990) to form hypotheses. Using simulations involving thousands of lotteries, the paper shows that *which* simpler (lower EIP) heuristic alternatives to the optimal WADD rule are most adaptive (come at the lowest cost to $\pi$) depends in sensitive ways to features of the problem like the number of distinct lottery outcomes or the dispersion in lottery payoffs. The best simpler-than-optimal procedure to use thus depends on subtle features of the task. Using a lottery choice experiment that uses Mouselab to measure the procedures subjects' use, Payne et al. (1988) show, remarkably, that subjects are sensitive to these adaptive consideration, to some degree tailoring which suboptimal procedures they use to the way both the costs (measured by estimated EIP costs) and relative benefits (in terms of $\pi$) of these procedures change as features of lotteries change. Thus, by measuring complexity at the procedural level using EIPs, the researchers are able to predict not only when people will fail to optimize, but which suboptimal rules they will adaptive substitute to instead.

### 4.2.3 Finite Automata

A second approach (with a similarly long history) instead taxonomizes procedures using formal descriptions of algorithms developed by computer scientists. In the 1980s, economic theorists began using finite state machines or finite automata – the simplest language for describing algorithms in




(a) Grim

(b) Tit-for-3-Tat

Figure 1: Automata representations of two rules (strategies in the repeated prisoner's dilemma).

computer science – to describe the complexity of strategies in repeated games like the repeated prisoner's dilemma, and study the implications for important social outcomes like cooperation (e.g. Rubinstein 1986, Abreu & Rubinstein 1988, Kalai & Stanford 1988).

Figure 1 shows sample automata descriptions of two classic strategies available to decision-makers in repeated prisoner's dilemma, to illustrate how automata taxonomize and describe procedures. Informally, a finite automata is defined by a set of states (shown in the diagrams as circles), instructions for behavior in each state (shown inside the circles), transitions between states (shown as arcs), and events that trigger these transitions (shown as letters next to the arcs). For instance, Figure 1a pictures the automata representation of the famous grim trigger strategy (Grim), which instructs the DM to begin by cooperating, choosing action "C" until the other player defects by choosing action "D", and afterwards defect for the remainder of the game. The freestanding arrow to the left instructs the player to begin in the left-most state, which instructs her to cooperate by choosing "C". The letters next to the arcs represent the player's opponent's most recent choice in the game. If the opponent chooses C/cooperate, the player is instructed by the loop above the state to respond by continuing to cooperate; if she chooses D/defects the player is instead instructed to transition to the right-most state and defect forever, regardless of what her opponent chooses afterwards. The automaton in Figure 1b represents another, more elaborate repeated prisoner's strategy called Tit-for-3-Tat, which requires more states and more transitions to describe.

The automaton literature considered the implications of various hypothesis about what algorithmic characteristics produce complexity costs – e.g., the number of states (Rubinstein 1986) or the number of transitions (Banks & Sundaram 1990). To the extent such features reliably map to (produce or predict) complexity, we can *predict* that a rule like Tit-for-3-Tat should be more complex than a rule like Grim. The theoretical literature has showed that, depending on which such assumptions are made about the algorithmic drivers of complexity costs, complexity can lead to fundamental changes in the way players make decisions in repeated games, since it shapes what strategies DMs choose to use. A further theoretical literature in economics has similarly used automata descriptions to model the complexity of procedures in order to explain behaviors ranging from status quo bias (Salant 2011), failures of Bayesian inference (Chauvin 2020), choice overload and satisficing (Salant 2011), randomization (Kalai & Solan 2003) and failures of backwards induction (Neyman 1985). However, this literature based these models on intuition about the drivers of procedural complexity, rather than direct empirical evidence.

Oprea (2020) reports an experiment that attempts to empirically map procedural costs using finite automata descriptions. Like Bettman et al. (1990) in the previous subsection and Sanjurjo (2024) in the next, he does this using an "assigned procedure" design in which subjects are directly assigned to implement algorithms describable as finite automata, and paid to do so correctly. Subjects in the experiment were shown a sequence of twenty letters (e.g., 'a' or 'b') one at a time and after observing each were required to type a letter in response (e.g., 'x' or 'y'). In each task, the subject was first given a rule for how to choose responses as a function of the letters she has been shown so far.[47] If the subjects correctly typed the 20 letters that complied with the rule, given the letters shown on their screen, they were paid a reward. The crucial component of the experiment was that after subjects had implemented a collection of such rules (which were selected to systematically vary the number of states and transitions in their automaton descriptions), they were asked to state the amount of money they would need to be paid to be willing to be assigned each of the same rules again in the future, using an incentive compatible elicitation method. Thus, subjects were asked to state how much money they were willing to "leave on the table" to avoid implementing the rule, thereby revealing the subjective cost they expected to experience from implementing it. The experiment thus thus directly measures the *cost complexity* for a range of rules, using methods discussed in Section 2.1.1.

The basic finding of the experiment is that subjects are willing to pay significant money simply to avoid implementing these procedures, revealing significant cost complexity, and this complexity has significant, measurable *structure* describable in terms of their automaton description. For instance, adding one automaton state to a rule increases complexity costs on average by about 20% while adding one transition increases these costs by half as much.[48,49] Finally, the experiment uses a special set of diagnostic rules to test whether finite automata are a nuanced enough descriptive language for taxonomizing and differentiating the complexity of procedures. The results are in the negative: by using a slightly richer class of algorithms to describe procedures called "pushdown automata", complexity costs can be better accounted for and predicted. (Pushdown automata

[47]For instance a rule used in one of the tasks reads "Choose x until you see a, after which switch to y for the rest of the round" while another read "Choose x until you see a, after which switch to y. Then, after you see b switch to x. Then, after you see another b switch to y. Then start over after you see a." The former is describable as an automaton containing two states and one transition (in excess of the number of states) while the second requires four states and four transitions.

[48]The experiment also shows how these costs are influenced by the context of the task: measured cost complexity rises significantly when the subject is asked to follow a rule all at once mentally rather than by taking actions once at a time, and measured costs fall significantly when the subjects is given opportunities to practice implementing the rule many times first.

[49]Oprea (2020) also collects response time data, and therefore is able to show that most of the main results measured for subjective cost also hold for response time. This suggests a linkage between drivers of subjects' distaste for procedures and drivers of subjects' use of one of the key resources expended in information processing, i.e., a linkage between cost complexity and time complexity.

are elaborations of finite state machines in which memory of past events are recorded as a working memory string, rather than represented by separate states, allowing them to make important distinctions between algorithms requiring different cognitive acts that finite state machine descriptions can't.) The experiment therefore suggests that using a richer language than finite automata to describe procedures will improve the predictive power of models of complexity. Given the many possible ways of taxonomizing and mapping procedures this kind of evidence is particularly valuable for learning how to map complexity at the procedural level effectively.

To give a sense for how these types of results might be useful in the social sciences, consider Banovetz & Oprea (2023) (discussed earlier) who study how procedural complexity influences the way subjects play simple "bandit" tasks – tasks in which subjects are tasked with repeatedly choosing between two arms to discover which is more valuable. Banovetz & Oprea (2023) first fomally show that we should *expect* optimal procedures in these tasks to be relatively complex since their automaton descriptions require four automata states, suggesting that the optimal procedure is relatively high-cost given the measurements from Oprea (2020). By contrast, a number of simpler suboptimal procedures (that involve under-exploring the arms or randomizing choices) are available that can be described with fewer states. Banovetz & Oprea (2023) show that in a baseline task subjects fail to optimize at a high rate in these tasks, and they show using structural estimates that subjects instead divert to simpler-than-optimal procedures, as predicted. In order to show that it was the information-processing costs described by automata states that *caused* this suboptimal behavior, Banovetz & Oprea (2023) remove these information-processing requirements altogether (using methods discussed in Section 2.2.2) in a follow-up treatment by having the computer process information on behalf of the subject (i.e., the computer reminds the subject of what she has done in the past and what the outcomes were in an immediately interpretable way). Removing costs associated with states in this way causes subjects to radically change their behavior: most subjects switch to using maximally "complex" optimal procedures, leading to very high rates of optimality in the task. Thus, the experiment shows that the complexity costs previously and more directly measured in Oprea (2020) were the cause of subjects' use of non-optimal procedures in the baseline task.[50]

[50]In a related paper, Guan & Oprea (2024) show the usefulness of automaton representations for understanding behavior in repeated prisoner's dilemmas, a key early motivation for this literature. Subjects are asked to directly choose from a menu of strategies to play against another subject. Guan & Oprea (2024) show that that varying whether subjects have to actually implement these strategies (instead of the computer doing it for them) themselves and whether they have to compute the payoff implications of these strategies (instead of having the computer compute them) both have significant effects on the automaton-complexity of the strategies they select. This suggests that both what we have called the complexity of implementation and the complexity of identification significantly influence behavior in repeated games, and that these costs are predictable in terms of finite automata descriptions.

### 4.2.4 Turing Machines and Computational Complexity Theory

One of the empirical conclusions of Oprea (2020) is that finite automata (among the simplest languages for describing algorithms in computer science) are too coarse to fully describe how complexity costs vary across procedures: in some contexts there are serious predictive gains to be had from using richer representations of algorithmic structure to measure and characterize complexity costs. Given this, a natural alternative for this task is Turing machines, which lie at the opposite end of the spectrum: they are highly nuanced formal ways of describing algorithms which make especially fine distinctions between the computational acts required across procedures. Roughly speaking, in their simplest form Turing machines describe algorithms or computations using a state/transition map (similar to finite automata) but add to this an infinite, multi-celled "memory tape" that the algorithm passes over and on which it can read and write information as it proceeds. One advantage of describing procedures using Turing machines (or equivalent descriptions) is that "time complexity," and "space complexity," the two most important complexity metrics discussed in computer science, can both be directly described in terms of their formal properties.[51] These metrics are appealing for two reasons. First, they have relatively transparent psychological interpretations – as effort and working memory load, respectively – and map roughly into the cognitive science-inspired distinction between computational complexity and representational complexity (Section 1.3), respectively. Second, these two metrics are at the center of computational complexity theory, the centerpiece theory of complexity in computer science (discussed in more depth below).

Sanjurjo (2024) uses Turing machines to study the space complexity (the working memory requirement) of procedures used to make choices in the same kinds of multi-attribute choice problems studied in the EIP literature and discussed in Section 4.2.2 above (e.g. Payne et al. 1993, Gabaix et al. 2006). Sanjurjo (2024) describes procedures in these task as search procedures that individually evaluate attribute values in some sequence. By describing these procedures as Turing machines, Sanjurjo (2024) shows that different types of procedures differ widely in their space complexity: alternative-intensive search (searching all of the attributes of each alternative fully before moving onto the next alternative) is minimally space complex, while attribute-intensive search (evaluating each attribute, one at a time) is highly space complex. Because of this, the procedure used to evaluate a set of options has a significant impact on how many pieces of information must be held in working memory, and thus the space requirements of the procedure. Importantly, Sanjurjo (2024) theoretically shows that when (and only when) the DM uses a space-inefficient procedure, the space complexity of the task also increases with the size of the problem (the number of attributes and alternatives), meaning DMs who use efficient procedures can insulate themselves from the space complexity that would otherwise result as the task grows large.[52]

[51] In a Turing machine description of a procedure, time complexity is simply the number of times the tape moves, and space complexity the number of places on the tape the algorithm must read to perform the procedure.

[52] One interesting subtlety of this analysis is that it complicates normal intuition that larger problems are more complex. Sanjurjo shows that this need not be true: in multi-attribute search problems, if subjects use a maximally efficient procedure, space complexity remains constant regardless of the size of the problem. The paper thus shows in

In order to study the mapping between complexity responses and this metric of Turing machines, Sanjurjo (2024) uses data from Sanjurjo (2015) which uses an "assigned procedure" design like Bettman et al. (1990) and Oprea (2020) discussed above. Instead of letting subjects choose the procedure for searching through attributes in the task (as in a conventional design), this experiment forces them to follow a pre-specified search path through attributes (using a Mouselab-like interface that reveals attribute/alternative pairs one at a time). In the experiment, subjects are sometimes forced to use a space-efficient alternative-intensive and sometimes an inefficient attribute-intensive search procedure to search through attributes prior to choosing an alternative. Using a simple model in which the DM's chances of becoming "overloaded" (and thereby being forced to choose randomly) rises in the theoretical space complexity of the problem, Sanjurjo (2024) predicts that mistakes rates should increase in space complexity. Varying the size of the problem (the numbers of attributes and alternatives) between 2 and 4, Sanjurjo (2024) finds that indeed variation in space complexity across tasks and procedures strongly predicts mistakes rates. The paper therefore identifies a first-order driver of complexity, measurable using a Turing description of procedures.[53]

Finally, Sanjurjo (2024) shows the potential value of mapping procedural costs in this way, by applying this analysis to two of the most commonly observed heuristics observed in the prior behavioral literature: satisficing (Simon 1955) and elimination-by-aspects (Tversky 1972). Sanjurjo (2024) shows that both of these heuristics allow decision-makers in search problems to economize on space complexity, though each in a different way, each describable in terms of primitives of a Turing machine. Returning to his analysis of the Sanjurjo (2015) data, he performs a calibration that shows that, given the estimated effect of space complexity on decision makers' chances of becoming overloaded, the use of these heuristics can be expected to dramatically reduce subjects' mistake rates relative to using an "optimal" procedure (which can be expected to produce high rates of costly mistakes). Once again, this exercise demonstrates the promise of measuring complexity at the level of the procedure: it not only allows the researcher to understand the proximal source of complexity-derived mistakes, it also allows the researcher to understand and anticipate procedures subjects are likely to use to economize on complexity. It also illustrates an important fact about complexity: simpler-than-optimal procedures not only economize on the direct costs (i.e., hedonic suffering or the opportunity cost of time) of making a decision, but also on the costs of mistakes that might result from using optimal procedures without sufficient cognitive resources to perform them properly.

**Computational Complexity Theory.** Sanjurjo (2024) provides an example of how Turing machines might be used to describe predictors of complexity at the procedural level, using an approach similar to that taken by Bettman et al. (1990) (for EIPs) and Oprea (2020) (for finite

a particularly nice way the kind of insight into complexity afforded by examining tasks at the procedural level rather than merely in terms of primitives of the task.

[53]Notice that our three leading examples of studies that attempt to directly map procedural complexity – Bettman et al. (1990), Oprea (2020) and Sanjurjo (2024) – use three distinct metrics of complexity from Section 2.1 for this task: response time, elicited subjective costs and mistakes rates, respectively.

automata). However, an alternative way Turing machines might be used to structure the search for predictors of complexity responses is though *computational complexity theory*, the primary literature for assessing the complexity of tasks used in computer science. While Sanjurjo (2024) focuses on *space complexity*, computational complexity theory often structures its classifications based on metrics of *time complexity*, the other key theoretical complexity metric produced by Turing machines.

In *computational complexity theory* computer scientists typically classify tasks based, roughly, on the *worst case* asymptotic properties of the simplest algorithm capable of correctly solving a problem. Roughly speaking, a problem is usually classified as computationally intractable and thus "complex" if, in the worst case, its time complexity – the number of operations it requires – increases more than polynomially as the problem grows large (for instance, as the number of items to be compared in an optimization task grows large). There are important questions about whether these asymptotic worst case classifications map effectively onto human behavior (i.e., predict human complexity responses like those discussed in Section 2), given that individual instances of *the same* type of task can vary widely in complexity metrics like time and space complexity. Because of this it is also valuable to pair these worst case classifications with measures of the *instance complexity* of a tasks – measures of the complexity of specific instance of the decision problem – and to study how metrics like time and space complexity of efficient solution algorithms predict complexity effects. Bossaerts & Murawski (2017) provide an excellent introduction to computational complexity theory for behavioral scientists.

A rich theoretical literature spanning several disciplines, explores the implications of computational complexity theory for human behavior, and attempts to uncover adaptations of that theory that make it maximally useful to the behavioral sciences. For instance, an important literature in cognitive science uses computational complexity theory to operationalize the "tractable cognition thesis": the idea that the range of available of human cognitive algorithms or mechanisms for explaining behavior can be narrowed down by focusing on those mechanisms that are computationally tractable (van Rooij 2008, van Rooij et al. 2019). van Rooij (2008), for instance, argues that a recent branch of computational complexity theory called "parameterized complexity" is particularly useful for describing human cognition, allowing for a relaxation called "fixed-parameter tractability" that is more conservative and therefore more useful for categorizing and differentiating the difficulty of algorithms. Likewise, a growing literature in economics highlights that computational complexity theory implies serious limits on the types of rationality typically assumed in economic models. For instance, economists have theoretically explored how computational complexity constrains consumer choice theory (Gilboa et al. 2021, Echenique et al. 2011), strategy choice in cooperative interactions (Gilboa 1988) and the formation of beliefs (Aragones et al. 2005). Camara (2024) shows that taking tractability constraints from computational complexity theory seriously immediately implies that humans *must* heuristically engage in choice bracketing, a widely observed behavioral anomaly (e.g. Tversky & Kahneman 1981, Rabin & Weizsäcker 2009) in which the DM breaks choice problems into sub-problems, optimizing them one at a time rather than altogether as optimality typically requires.

A small experimental literature has developed in the last decade examining how well computational complexity theory organizes decision-making when *humans* are faced with some of the workhorse computational tasks studied in computer science. This literature has so far focused especially on knapsack problems and closely related combinatorial optimization problems. Murawski & Bossaerts (2016), for instance, study 0-1 knapsack problems in which the subject is shown a number of objects, each with a *value* and a *weight* and is tasked with selecting a subset that maximizes the sum of values, subject to a constraint on the sum of weights. These tasks are NP-hard, meaning they are generally thought to be worst-case asymptotically computationally intractable. Murawski & Bossaerts (2016) study how humans solve these problems, and uses a number of the tools for measurement described in Section 2 to do so. Subjects in the experiment were asked to select subsets of objects on their screen and shown the results, and were allowed to repeat this as many times as they liked before submitting a final choice. This design feature allowed the researchers to directly observe key aspects of the procedure/algorithm subjects used to attempt to solve this problem. By examining both clock time and the number of changes subjects made to their selections, they were able to get two measures of the time complexity of the procedure subjects employed. Crucially, Murawski & Bossaerts (2016) also varied the task in such a way as to vary a standard metric of the theoretical time complexity of each instance of the problem, under a Turing description of the optimal procedure. Thus they varied the instance size (the number of items available), a standard way of attempting to manipulate complexity (see 4.1 above), but which here has a crisp algorithmic interpretation: doing so increases the time complexity of the instance. Moreover, they also varied the tasks in such a way as to vary Sahni-k, a metric of instance complexity that is proportional to the time and space complexity required to solve the problem using sophisticated variations on the greedy algorithm.

Murawski & Bossaerts (2016) found that these computationally complex problems are also complex for humans: subjects fail to find the optimum in the majority of cases studied. However, they also found that these mistakes increased in the time/space complexity of the instance (as measured by Sahni-k): as the time and working memory burden of the task increased, subjects' ability to find the optimum decreased. Murawski & Bossaerts (2016) argue that this reflects working memory limitations in humans relative to machines, which constrains humans from performing optimal procedures. Detailed analysis of subjects' patterns of search yield insight into how subjects respond to this instance complexity heuristically in order to economize on computational costs. Metrics of effort (clock time and the number of steps in the algorithm used) also increased with the difficult of the problem, suggesting that subjects use satisficing (Simon 1955) rather than optimizing algorithms. There is also evidence from detailed analysis of subjects' procedures that episodic memory constraints cause subjects to use path-dependent algorithms that fail to reconsider choices in earlier steps of the search. Finally, Murawski & Bossaerts (2016) show that subjects' choice of procedures is adaptive (Payne et al. 1993), changing in measurable ways with characteristics of the problem instance.[54] Franco et al. (2018) expands this type of analysis to other classic problems from computer science (e.g., the travelling salesman problem) and other approaches to estimating the computational difficulty of instances of the task. In another extension, Bossaerts et al. (2024) provide evidence that formal notions of the computational complexity of *approximating* an optimal solution ("approximation complexity") also predict humans' ability to approximate optimal choice.

# 5 Conclusion

The empirical study of the effects complexity has on human behavior has become an increasingly active area of research in the behavioral sciences. Although the idea has deep roots, stretching back decades in many fields (particularly in cognitive science) it has made significant inroads into other behavioral disciplines especially in recent years. For instance, the idea of a cognitive cost has gained new currency in neuroscience and cognitive psychology and understanding the physiological roots of these costs has become an active topic of interest in these fields. Likewise, research on the role complexity plays in long-documented deviations from standard rational choice predictions has seen a sharp uptick in behavioral economics in recent years. Converging, cross-disciplinary interest in complexity makes it a natural point of consilience between behavioral disciplines.

This development is potentially important for three reasons. First, complexity is one of a handful of theoretical ideas with the potential to organize a great number of apparent anomalies in human behavior under a unified framework. The behavioral sciences are in need of fewer individual theories, not more, and the simple and venerable idea that humans have constraints in their ability and willingness to expend cognitive resources is one with particularly broad explanatory potential. Second, framing the problem of cognitive costs (as complexity research does) in terms of the procedures or algorithms that generate behavior means that the study of complexity can be a constructive enterprise that does more than rationalize behavior on a case by case basis by posulating the existence of unobserved costs and constraints. Careful, empirical study of how costs are generated by the characteristics of the algorithms or procedures that generate behavior (and their fit to the choice environment) may allows us to tightly connect very different behaviors in very different settings to one another, in a unified descriptive (and predictive) language. Most importantly, mapping the roots of these costs at the algorithmic level means we may one day be able to predict and explain not only what optimal behaviors humans are unlikely to display, but also what alternative behaviors they will tend to display instead. Finally, because complexity connects individual behavior to the properties of rules, the same ideas and measurements we use to explain anomalies in choice may allow us to better understand the structure of the social world, connecting the behavioral sciences to the social sciences in a fundamental way. Social institutions (and even human cultures) are composed to a great extent of constellations of rules, and a major task of a social scientist is to understand where these rules come from and when and why they

[54] Hong & Stauffer (2023) show that very similar overall conclusions arise in knapsack tasks performed by monkeys.

often fail to support human flourishing. Understanding the costs and limitation people face when enacting *internally generated* rules to guide their own behavior may help us to understand their limitations in reasoning about and complying with the rules erected *externally* by the organizations and political entities that scaffold the social world.

The success of a cross-disciplinary science of complexity will depend especially on careful measurement of complexity and its effects. In this essay I have reviewed some of the most promising methods for doing this kind of work and illustrated using some frontier areas of research. Some of these approaches are decades old, but many are new, emerging only in the last few years (and others are being applied in recent years for new purposes). This is a rapidly developing area of research, with many opportunities for innovation in methods. There are therefore reasons to hope that this essay will become quickly outdated in the coming years.

## References


Abeler, J. & Jäger, S. (2015), 'Complex Tax Incentives', *American Economic Journal: Economic Policy* **7**(3), 1–28.

Abreu, D. & Rubinstein, A. (1988), 'The Structure of Nash Equilibrium in Repeated Games with Finite Automata', *Econometrica* **56**(6), 1259–1281.

Alaoui, L. & Penta, A. (2015), 'Endogenous Depth of Reasoning', *The Review of Economic Studies* **83**(4), 1297–1333.

Alaoui, L. & Penta, A. (2022), 'Cost-benefit analysis in reasoning', *Journal of Political Economy* **130**, 825–1145.

Alós-Ferrer, C., Granić, D., Kern, J. & Wanger, A. (2016), 'Preference reversals: Time and again', *Journal of Risk and Uncertainty* **52**, 65–97.

Aragones, E., Gilboa, I., Postlewaite, A. & Schmeidler, D. (2005), 'Fact-Free Learning', *American Economic Review* **95**(5), 1355–1368.

Arieli, A., Ben-Ami, Y. & Rubinstein, A. (2011), 'Tracking decision makers under uncertainty', *American Economic Journal: Microeconomics* **3**(4), 68–76.

Arora, S. & Barak, B. (2009), *Computational Complexity: A Modern Approach*, Cambridge University Press.

Arrieta, G. R. & Nielsen, K. (2024), 'Procedural decision-making in the face of complexity'.

Ba, C., Bohren, J. A. & Imas, A. (2024), 'Over- and underreaction to information'.

Banks, J. S. & Sundaram, R. K. (1990), 'Repeated Games, Finite Automata, and Complexity', *Games and Economic Behavior* **2**(2), 97–117.

Banovetz, J. & Oprea, R. (2023), 'Complexity and procedural choice', *American Economic Journal: Microeconomics* **15**(2), 384–413.

Bauermeister, G.-F., Hermann, D. & Musshof, O. (2018), 'Consistency of determined risk attitudes and probability weightings across different elictation methods', *Theory and Decision* **84**, 627–644.

Beauchamp, J., Benjamin, D., Laibson, D. & Chabris, C. (2020), 'Measuring and controlling for the compromise effect when estimating risk preference parameters', *Experimental Economics* **23**, 1069–1099.

Becker, G., Degroot, M. & Marschak, J. (1964), 'Measuring Utility by a Single-Response Sequential Method', *Behavioral Science* **9**(3), 226–232.

Benartzi, S. & Thaler, R. H. (2001), 'Naive diversification strategies in defined contribution saving plans', *American Economic Review* **91**(1), 79–98.

Benjamin, D., Brown, S. & Shapiro, J. (2013), 'Who is 'behavioral'? cognitive ability and anomalous preferences.', *Journal of the European Economic Association* **11**(6), 1231–1255.

Benjamin, D. J., Fontana, M. A. & Kimball, M. S. (2023), 'Reconsidering risk aversion', *Unpublished manuscript* .

Bernheim, B. D. & Sprenger, C. (2020), 'On the empirical validity of cumulative prospect theory: experimental evidence of rank-independent probability weighting', *Econometrica* **88**(4), 1363–1409.

Bettman, J. R., Johnson, E. J. & Payne, J. W. (1990), 'A componential analysis of cognitive effort in choice', *Organizational Behavior and Human Decision Processes* **45**, 111–139.

Blavatskyy, P. (2007), 'Stochastic expected utility theory', *Journal of Risk and Uncertainty* **34**, 259–286.

Bohren, J. A., Hascher, J., Imas, A., Ungeheuer, M. & Weber, M. (2024), A cognitive foundation for perceiving uncertainty. Unpublished manuscript.

Bordalo, P., Gennaioli, N. & Shleifer, A. (2012), 'Salience theory of choice under risk', *Quarterly Journal of Economics* **127**(3), 1243–1285.

Bordalo, P., Gennaioli, N. & Shleifer, A. (2022), 'Salience', *Annual Review of Economics* .

Bossaerts, P., Franco, J. P., Hsu, A., Murawski, C. & Yadav, N. (2024), 'How well can humans approximate optimality in computationally hard problems?'.

Bossaerts, P. & Murawski, C. (2017), 'Computational Complexity and Human Decision-Making', *Trends in Cognitive Sciences* **21**(12), 917–929.

Bradley, R. A. & Terry, M. E. (1952), 'Rank analysis of incomplete block designs: The method of paired comparisons', *Biometrika* **39**(3/4), 324–345.

Camara, M. (2024), 'Computationally tractable choice', *Unpublished Manuscript* .

Camerer, C. F. & Hogarth, R. M. (1999), 'The effects of financial incentives in experiments: A review and capital-labor-production framework', *Journal of Risk and Uncertainty* **19**, 7–42.

Campbell, D. J. (1988), 'Task Complexity: A Review and Analysis', *The Academy of Management Review* **13**, 40–52.

Caplin, A., Csaba, D., Leahy, J. & Nov, O. (2020), 'Rational inattention, competitive supply, and psychometrics', *The Quarterly Journal of Economics* **135**, 1681–1724.

Caplin, A., Dean, M. & Martin, D. (2011), 'Search and Satisficing', *American Economic Review* **101**(7), 2899–2922.

Capra, C. M. (2019), 'Understanding decision processes in guessing games: a protocol analysis approach', *Journal of Economic Science Association* **5**, 123–135.

Chauvin, K. P. (2020), 'Euclidean Properties of Bayesian Updating'.

Chernev, A., Böckenholt, U. & Goodman, J. (2015), 'Choice overload: A conceptual review and meta-analysis', *Journal of Consumer Psychology* **25**(2), 333–358.

Colliard, J.-E. & Georg, C.-P. (2023), 'Measuring regulatory complexity'.

Deck, C. & Jahedi, S. (2015), 'The effect of cognitive load on economic decision making: A survey and new experiments', *European Economic Review* **78**, 97–119.

Deck, C., Jahedi, S. & Sheremeta, R. (2021), 'On the consistency of cognitive load', *European Economic Review* **134**, 5.

DeClippel, G., Moscariello, P., Ortoleva, P. & Rozen, K. (2024), Caution in the face of complexity. Unpublished manuscript.

Duckworth, A. L., Quinn, P. D., Lynam, D. R., Loeber, R. & Stouthamer-Loeber, M. (2011), 'Role of test motivation in intelligence testing', *Proceedings of the National Academy of Sciences* **108**, 7716–7720.

Echenique, F., Golovin, D. & Wierman, A. (2011), 'A Revealed Preference Approach to Computational Complexity in Economics', *Proceedings of the 12th ACM Conference on Electronic Commerce* .

Enke, B. (2024), The cognitive turn in behavioral economics. Unpublished manuscript.

Enke, B., Gneezy, U., Hall, B., Martin, D., Nelidov, V., Offerman, T. & van de Ven, J. (2023), 'Cognitive biases: Mistakes or missing stakes?', *The Review of Economics and Statistics* **105**(4), 818–832.

Enke, B. & Graeber, T. (2023), 'Cognitive uncertainty', *The Quarterly Journal of Economics* **138**, 2021–2067.

Enke, B., Graeber, T. & Oprea, R. (2023), 'Confidence, self-selection and bias in the aggregate', *American Economic Review* **113**(7), 1933–1966.

Enke, B., Graeber, T. & Oprea, R. (2024), 'Complexity and time'.

Enke, B., Graeber, T., Oprea, R. & Yang, J. (2024), 'Behavioral attenuation'.

Enke, B. & Shubatt, C. (2023), 'Quantifying lottery choice complexity', *Unpublished manuscript* .

Ericsson, K. A. & Simon, H. A. (1980), 'Verbal reports as data', *Psychological Review* **87**(3), 215–251.

Ericsson, K. A. & Simon, H. A. (1993), *Protocol Analysis: Verbal Reports as Data*, MIT Press, Cambridge, MA.

Fechner, H. B., Schooler, L. J. & Pachur, T. (2018), 'Cognitive costs of decision-making strategies: A resource demand decomposition analysis with a cognitive architecture', *Cognition* **170**, 102–122.

Franco, J. P., Yadav, N., Bossaerts, P. & Murawski, C. (2018), 'Where the Really Hard Choices Are: A General Framework to Qquantify Decision Difficulty', *BioRxiv* p. 405449.

Frederick, S. (2005), 'Cognitive reflection and decision making', *Journal of Economic Perspectives* **19**(4), 25–42.

Freeman, D. J., Halevy, Y. & Kneeland, T. (2019), 'Eliciting risk preferences using choice lists', *Quantitative Economics* **10**(1), 163–206.
**URL:** *https://doi.org/10.3982/QE760*

Freeman, D. J. & Mayraz, G. (2019), 'Why choice lists increase risk taking', *Experimental Economics* **22**, 131–154.
**URL:** *https://doi.org/10.1007/s10683-018-09596-7*

Friedman, D., Habib, S., James, D. & Williams, B. (2022), 'Varieties of risk preference elicitation', *Games and Economic Behavior* **forthcoming**.

Friedman, D., Isaac, R. M., James, D. & Sunder, S. (2017), *Risky curves: on the empirical failures of expected utility*, Routledge.

Frydman, C. & Jin, L. J. (2022), 'Efficient coding and risky choice', *The Quarterly Journal of Economics* **137**, 161–213.

Frydman, C. & Jin, L. J. (2023), 'On the source and instability of probability weighting', *Unpublished Manuscript* .

Fudenberg, D., Strack, P. & Strzalecki, T. (2018), 'Speed, Accuracy, and the Optimal Timing of Choices', *American Economic Review* **108**(12), 3651–3684.

Gabaix, X. (2014), 'A Sparsity-Based Model of Bounded Rationality', *The Quarterly Journal of Economics* **129**(4), 1661–1710.

Gabaix, X. & Graeber, T. (2023), 'The complexity of economic decisions', *Unpublished manuscript* .

Gabaix, X. & Laibson, D. (2022), Myopia and discounting, Technical report, National bureau of economic research.

Gabaix, X., Laibson, D., Moloche, G. & Weinberg, S. (2006), 'Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model', *American Economic Review* **96**(4), 1043–1068.

Gagnon-Bartsch, T., Rabin, M. & Schwartzstein, J. (2023), Channeled attention and stable errors. Unpublished manuscript.

Gershman, S. J. & Bhui, R. (2019), 'Rationally inattentive intertemporal choice', *bioRxiv* p. 680652.

Gigerenzer, G., Hertwig, R. & Pachur, T. (2015), *Heuristics: The Foundation of Adaptive Behavior*, Oxford University Press.

Gigerenzer, G. & Hoffrage, U. (1995), 'How to improve bayesian reasoning without instruction: Frequency formats', *Psychological Review* **102**, 684–704.

Gigerenzer, G. & Selten, R. (2001), *Bounded Rationality: The Adaptive Toolbox*, MIT Press, Cambridge, Mass.

Gilboa, I. (1988), 'The Complexity of Computing Best-Response Automata in Repeated Games', *Journal of Economic Theory* **45**(2), 342–352.

Gilboa, I., Postlewaite, A. & Schmeidler, D. (2021), 'The complexity of the consumer problem', *Research in Economics* **75**, 96–103.

Gneezy, U. & Rustichini, A. (2000), 'Pay enough or don't pay at all', *The Quarterly Journal of Economics* **115**(3), 791–810.

Gonçalves, D. (2024), Speed, accuracy, and complexity. Unpublished.

Guan, M. & Oprea, R. (2024), 'Three faces of complexity in strategic choice'.

Guan, M., Oprea, R. & Yuksel, S. (2024), 'Too much information'.

Halstead, M. H. (1977), *Elements of Software Science*, Elsevier.

Harrison, G. W. (1989), 'Theory and misbehavior of first-price auctions', *The American Economic Review* **79**(4), 749–762.

He, J. & Natenzon, P. (2023), 'Random choice and learning', *American Economic Review: Insights* .

Hong, T. & Stauffer, W. R. (2023), 'Computational complexity drives sustained deliberation', *Nature Neuroscience* **26**, 850–857.

Huck, S. & Weizsacker, G. (1999), 'Risk, complexity and deviations from extpected-value maximizagion: Results of a lottery choice experiment', *Journal of Economic Psychology* **149**(9), 1644–1683.

Johnson, E. J. & Payne, J. W. (1985), 'Effort and accuracy in choice', *Management Science* **31**, 394–414.

Kahneman, D. (1973), *Attention and Effort*, Prentice-Hall, Englewood Cliffs, NJ.

Kahneman, D. & Tversky, A. (1979), 'Prospect theory: An analysis of decision under risk', *Econometrica* **47**(2), 263–291.

Kalai, E. & Solan, E. (2003), 'Randomization and Simplification in Dynamic Decision-Making', *Journal of Economic Theory* **111**(2), 251–264.

Kalai, E. & Stanford, W. (1988), 'Finite Rationality and Interpersonal Complexity in Repeated Games', *Econometrica* **56**(2), 397–410.

Kendall, C. & Oprea, R. (2024), 'On the complexity of forming mental models', *Quantitative Economics* **15**, 175–211.

Khaw, M. W., Li, Z. & Woodford, M. (2021), 'Cognitive imprecision and small-stakes risk aversion', *Review of Economic Studies* **88**(4), 1979–2013.

Kool, W. & Botvinick, M. (2018), 'Mental Labour', *Nature Human Behavior* **2**, 899–908.

Kotseruba, I. & Tsotsos, J. K. (2020), '40 years of cognitive architectures: core cognitive abilities and practical applications', *AI Review* **53**, 17–94.

Krajbich, I., Oud, B. & Fehr, E. (2014), 'Benefits of neuroeconomic modeling: New policy interventions and predictors of preference', *American Economic Review* **104**(5), 501–506.

Krebs, R. M., Boehler, C. N. & Woldorff, M. G. (2010), 'The influence of reward associations on conflict processing in the stroop task', *Cognition* **117**, 341–347.

Kurzban, R., Duckworth, A., Kable, J. W. & Myers, J. (2013), 'An opportunity cost model of subjective effort and task performance', *Behavioral and Brain Sciences* **36**(6), 661–679.

Lempert, K. M. & Phelps, E. A. (2013), Neuroeconomics of emotion and decision making, *in* P. W. Glimcher, ed., 'Neuroeconomics: Decision Making and the Brain: Second Edition', Elsevier Inc., pp. 219–236.

Lieder, F. & Griffiths, T. (2020), 'Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources', *Behavioral and Brain Sciences* **43**.

Lieder, F., Plunkett, D., Hamrick, J. B., Russell, S. J., Hay, N. J. & Griffiths, T. L. (2014), Algorithm selection by rational metareasoning as a model of human strategy selection, *in* 'Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2', NIPS'14, MIT Press, p. 2870–2878.

Lipman, B. L. (1995), 'Information Processing and Bounded Rationality: A Survey', *Canadian Journal of Economics* **28**(1), 42–67.

Liu, P. & Li, Z. (2012), 'Task Complexity: A Review and Conceptualization Framework', *International Journal of Industrial Ergonomics* **42**(6), 553–568.

Lloyd, S. (2001), 'Measures of complexity: a nonexhaustive list', *IEEE Control Systems Magazine* **21**, 78.

Loewenstein, G. & Wojtowicz, Z. (2024), 'The economics of attention'.

Marr, D. (1982), *Vision: A Computational Approach*, Freeman & Co., San Francisco.

Maćkowiak, B., Matějka, F. & Wiederholt, M. (2023), 'Rational inattention: A review', *Journal of Economic Literature* **61**(1), 226–273.

Murawski, C. & Bossaerts, P. (2016), 'How Humans Solve Complex Problems: The Case of the Knapsack Problem', *Scientific reports* **6**, 34851.

Newell, A. & Simon, H. A. (1972), *Human Problem Solving*, Prentice Hall, Englewood Cliffs.

Neyman, A. (1985), 'Bounded Complexity Justifies Cooperation in the Finitely Repeated Prisoners' Dilemma', *Economics Letters* **19**(3), 227–229.

Nielsen, K. & Rehbeck, J. (2022), 'When choices are mistakes', *American Economic Review* **112**(7), 2237–2268.

Oprea, R. D. (2020), 'What Makes a Rule Complex', *American Economic Review* **110**(12), 3913–3951.

Oprea, R. D. (2024), 'Decisions under risk are decisions under complexity'. Forthcoming, American Economic Review.

Oprea, R. & Vieider, F. M. (2024), Minding the gap: On the origins of probability weighting and the description-experience gap. Manuscript.

Ortoleva, P. (2013), 'The Price of Flexibility: Towards A Theory of Thinking Aversion', *Journal of Economic Theory* **148**(3), 903–934.

Osborn Popp, P. J., Newell, B., Bartels, D. & Gureckis, T. (2023), Can cognitive discovery be incentivized with money? Unpublished.

Padmala, S. & Pessoa, L. (2011), 'Reward reduces conflict by enhancing attentional control and biasing visual cortical processing', *Journal of Cognitive Neuroscience* **23**, 3419–3432.

Payne, J. M., Bettman, J. R. & Johnson, E. J. (1992), 'Behavioral decision research: A constructive processing perspective', *Annual Review of Psychology* **43**, 87–131.

Payne, J. W., Bettman, J. R. & Johnson, E. J. (1988), 'Adaptive strategy selection in decision making', *Journal of Experimental Psychology: Learning, Memory, and Cognition* **14**, 534–552.

Payne, J. W., Bettman, J. R. & Johnson, E. J. (1993), *The Adaptive Decision Maker*, Cambridge University Press.

Posner, M. I. & Snyder, C. R. R. (1975), Attention and cognitive control, *in* R. L. Solso, ed., 'Information Processing and Cognition: The Loyola Symposium', Erlbaum Assoc., Hillsdale, NJ.

Puri, I. (2023), 'Simplicity and risk', *Unpublished Manuscript* .

Rabin, M. (2013), 'Incorporating limited rationality into economics', *Journal of Economic Literature* **51**(2), 528–543.

Rabin, M. & Weizsäcker, G. (2009), 'Narrow bracketing and dominated choices', *American Economic Review* **99**(4), 1508–43.

Rees-Jones, A. & Taubinsky, D. (2020), 'Measuring "schmeduling"', *The Review of Economic Studies* **87**, 2399–2438.

Rieskamp, J. & Otto, P. E. (2006), 'Ssl: A theory of how people learn to select strategies', *Journal of Experimental Psychology: General* **135**, 207–236.

Rubinstein, A. (1986), 'Finite Automata Play the Repeated Prisoner's Dilemma', *Journal of Economic Theory* **39**(1), 83–96.

Rubinstein, A. (1988), 'Similarity and decision-making under risk (is there a utility theory resolution to the allais paradox?)', *Journal of Economic Theory* **46**, 145–153.

Rubinstein, A. (2013), 'Response time and decision making: An experimental study', *Judgment and Decision Making* **8**(5), 540–551.

Rubinstein, A. (2016), ‘A typology of players: Between instinctive and contemplative’, *The Quarterly Journal of Economics* **131**, 859–890.

Salant, Y. (2011), ‘Procedural Analysis of Choice Rules with Applications to Bounded Rationality’, *American Economic Review* **101**(2), 724–748.

Sanjurjo, A. (2015), ‘Search, memory, and choice error: an experiment’, *PLoS ONE* **10**(6), e0126508.

Sanjurjo, A. (2024), ‘Complexity and choice’.

Schulte-Mecklenbeck, M., Johnson, J. G., Böckenholt, U., Goldstein, D. G., Russo, J. E., Sullivan, N. J. & Willemsen, M. C. (2017), ‘Process-tracing methods in decision making: On growing up in the 70s’, *Current Directions in Psychological Science* **26**(5), 442–450.

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D. & Botvinick, M. M. (2017), ‘Toward a Rational and Mechanistic Account of Mental Effort’, *Annual Review of Neuroscience* **40**, 99–124.

Shiffrin, R. M. & Schneider, W. (1977), ‘Controlled and automatic information processing: Ii. perceptual learning, automatic attending, and a general theory’, *Psychological Review* **84**, 127–190.

Shubatt, C. & Yang, J. (2024), ‘Tradeoffs and comparison complexity’.

Shugan, S. M. (1980), ‘The cost of thinking’, *Journal of Consumer Research* **7**(2), 99–111.

Simon, H. A. (1955), ‘A Behavioral Model of Rational Choice’, *The Quarterly Journal of Economics* **69**(1), 99–118.

Simon, H. A. (1990), ‘Invariants of human behavior’, *Annual Review of Psychology* **41**, 1–19.

Sims, C. A. (2003), ‘Implications of Rational Inattention’, *Journal of Monetary Economics* **50**(3), 665–690.

Spiliopoulos, L. & Ortmann, A. (2018), ‘The bcd of response time analysis in experimental economics’, *Experimental Economics* **21**, 383–433.

Stanovich, K. E. & West, R. F. (2008), ‘On the relative independence of thinking biases and cognitive ability’, *Journal of Personality and Social Psychology* **94**(4), 672–695.

Stewart, N., Chater, N. & Brown, G. D. (2006), ‘Decision by sampling’, *Cognitive Psychology* **53**, 1–26.

Thagard, P. (2005), *Mind: Introduction to Cognitive Science*, The MIT Press.

Todd, P. M. & Gigerenzer, G. (2012), *Ecological Rationality: Intelligence in the World*, Oxford University Press.

Tversky, A. (1972), 'Elimination by aspects: A theory of choice', *Psychological Review* **79**(4), 281.

Tversky, A. & Kahneman, D. (1981), 'The framing of decisions and the psychology of choice', *Science* **211**(4481), 453–458.

Tversky, A. & Kahneman, D. (1992), 'Advances in prospect theory: Cumulative representation of uncertainty', *Journal of Risk and Uncertainty* **5**(4), 297–323.

van der Wel, P. & van Steenbergen, H. (2018), 'Pupil dilation as an index of effort in cognitive control tasks: A review', *Psychonomic Bulletin and Review* **25**, 2005–2015.

van Rooij, I. (2008), 'The tractable cognition thesis', *Cognitive Science* **32**, 939–984.

van Rooij, I., Blokpoel, M., Kwisthout, J. & Wareham, T. (2019), *Cognition and Intractability: A Guide to Classical and Parameterized Complexity Analysis*, Cambridge University Press.

Vieider, F. (2024), 'Decisions under uncertainty as bayesian inference on choice options', *Management Science* .

Vieider, F. M. (2021), Noisy coding of time and reward discounting, Technical report, Ghent University, Faculty of Economics and Business Administration.

Webb, R. (2018), 'The (neural) dynamics of stochastic choice', *Management Science* **65**(1), 230–255.

Wilcox, N. (1993), 'Lottery Choice: Incentives, Complexity and Decision Time', *The Economic Journal* **103**(421), 1397–1417.

Woodford, M. (2020), 'Modeling imprecision perception, valuation and choice', *Annual Review of Economics* **12**.