

Getting the Picture

Robert Akerlof, Richard Holden, and Hongyi Li*

September 20, 2026

Abstract

Standard economic theory assumes that people have perfect information-processing ability: they can derive *all* logical consequences of information *immediately*. In practice people struggle to do so. A person deciding whether to refinance a mortgage or when to realize a capital gain may hold every relevant fact and still fail to see what follows. We offer a theory of the reasoning process—of how people connect the dots. Our key assumption is that agents have limited *processing capacity*: a constraint on the number of pieces of information—dots—they can think about and connect *at once*. The constraint determines both what makes a choice problem difficult and the mistakes people make. Faced with problems too hard to solve by deduction, or that take too long, agents *extrapolate*: attending selectively to part of the problem, they settle on conclusions that fit the facts well enough. The leaps they make are systematically biased, in a direction set by what they attend to and what they leave out. The theory yields a range of predictions. Agents satisfice, and fall back on defaults when decisions are complex. More complex decisions take longer to make and lead to more mistakes. More options can mean fewer purchases. A product with one large advantage may outsell one whose smaller advantages are larger in total. Agents can hold inconsistent beliefs—producing, for instance, intransitive choice. And in persuasion, it pays to make one’s case first.

JEL Classification: D01, D80, D90.

Keywords: Cognition, reasoning, perception, narratives.

*Akerlof: UNSW Business School, r.akerlof@unsw.edu.au. Holden: UNSW Business School, richard.holden@unsw.edu.au. Li: UNSW Business School, hongyi@unsw.edu.au. We are grateful to Drew Fudenberg, Rachel Kranton, Raghav Malhotra, Stephen Morris, Sharun Mukand, Sendhil Mullainathan, Frank Schilbach, Andrei Shleifer, Richard Thaler, and DJ Thornton for helpful comments, as well as seminar participants at Harvard, MIT, Chicago, Brown, Monash, and UNSW. All authors acknowledge financial support from the Manos Institute for Cognitive Economics at UNSW. Li additionally acknowledges financial support from the Australian Research Council Discovery Project DP240103257.

1 Introduction

Standard economic theory assumes that people have perfect information processing capabilities. Given a set of information, people can immediately derive all logical consequences from it. Two words are worth emphasizing here: *immediately* and *all*.

Reality is not so tidy. There is by now a vast body of evidence that people struggle to process information, and many contributions have offered rationales for why behavior deviates from full rationality. With some important exceptions that we discuss below, these either posit a behavioral tendency as a primitive or subject a Bayesian agent to a cognitive constraint. What goes unexamined, in either case, is the reasoning process itself: how an agent who has the facts works out what they imply.

Our paper focuses on this reasoning process: how people, endowed with all the information needed to draw a conclusion, go about “connecting the dots” to determine what is optimal. Our central assumption is that agents have limited *processing capacity*—a constraint on the number of pieces of information, or “dots,” an agent can think about and connect *at once*. As a result, the agent’s ability to draw conclusions is bounded, and the speed at which they reach those conclusions is slowed—computation becomes part of the agent’s problem. We show that this mechanism sheds new light on economic problems ranging from the simplest choice to political persuasion, and that limited processing capacity not only constrains agents’ ability to connect the dots but leads to systematic, predictable biases.

The idea of a bounded processing capacity has deep roots in psychology. People can hold only a handful of items in mind at once—seven by Miller (1956)’s classic estimate, four by Cowan (2001)’s—and attention has long been modeled as a limited capacity (Broadbent, 1958; Kahneman, 1973). We use the term *processing capacity* broadly, for the limited resource that can be brought to bear on combining information at any one moment, and remain agnostic about whether the processing is conscious or unconscious.¹

¹The two may nonetheless be related: there is some evidence that the processing which rises to conscious-

Ordinary financial decisions show how such limits bind. Deciding whether to refinance a mortgage, how much to save for retirement, or how to time the sale of an asset requires holding several facts in mind at once and combining them. And a person with every relevant fact in front of them may still fail to see what the facts jointly imply. Each fact on its own is plain. The difficulty lies in putting them together.²

In the model, an agent faces a choice problem where the optimal action depends on the state of the world—which we refer to as a “picture” made up of “pixels.” A “feature” is an event in picture-space: something the agent might know about the picture. To choose, the agent must work out a *decision-relevant* feature, one that reveals which action is best. The agent’s “knowledge” is the set of features they have worked out so far. It may contain all the information needed to determine the decision-relevant feature without containing that feature: the agent must reason their way to it, drawing on facts they know to reach new ones. Two limits constrain this process. Limited processing capacity restricts how many facts the agent can draw on at once—in computing terms, the agent has scarce RAM but a vast hard drive.³ And the agent has a limited set of *concepts*, and can draw only on facts they can conceptualize.⁴ Without intermediate concepts to build on, even ample processing capacity may not suffice to assemble the raw facts into a decision.

We consider two variants of the model. In the first, the agent is purely deductive. This variant isolates how the agent reasons, but the agent is never wrong: they either reach the right answer or reach no answer at all. In the second, the agent *extrapolates*, accepting a conclusion when it “fits the data” sufficiently well—when relatively few alternative explanations remain. Extrapolation lets the agent go beyond what they can deduce, and it lets them reach a conclusion in fewer steps, saving time and effort. In both cases it

ness is precisely that which integrates information across distant, otherwise-separate regions of the brain (see Dehaene, 2014).

²In Section 2 we work through an explicit example, in which Mr. Green must decide whether it is better to realize a capital gain this year or next.

³We treat the store of knowledge as unbounded. It may be bounded in reality, but treating it as unbounded isolates the role of limited processing capacity.

⁴Cognitive scientists and neuroscientists have shown that the brain encodes complex information in terms of concepts to reduce cognitive load (see, for example, Baddeley and Hitch (1974)).

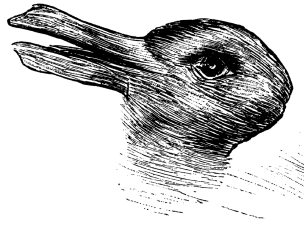


Figure 1: The Rabbit-Duck Illusion

means attending selectively to aspects of the problem, and in both it opens the door to error.

Even the deductive model makes predictions. The agent reasons in steps, combining a few facts at a time, so agents who can hold fewer facts in mind reach fewer conclusions, and reach them more slowly. Experts outperform novices by chunking—compressing familiar configurations into single units, as chess masters do with board positions. And we develop a measure of problem difficulty—the number of steps a problem requires—that predicts how long problems take to solve.

The extrapolation model—which is the heart of the paper—makes richer predictions which are more psychological in nature. More complex decisions take longer to make and lead to more mistakes. Consumers who face more options may make fewer purchases, and improving an inferior option can make the choice harder rather than easier. Agents may *satisfice* (Simon, 1955), attend selectively to attributes, and fall back on defaults. They may hold beliefs that cannot all be true at once without noticing—as households do in carrying high-interest credit-card debt alongside low-yielding savings, or investors in judging the market overpriced while still expecting it to rise. And the mistakes extrapolation produces are systematically biased, in a direction set by the structure of the problem: for instance, a product with one large advantage may outsell one whose smaller advantages are larger in total.

Finally, we consider salience. Adding the assumption that “powerful” features are particularly salient makes thinking *path dependent*. Where reasoning begins shapes where

it ends, as when, having seen the rabbit in the rabbit–duck illusion (Figure 1), one finds it hard to see the duck. The framework thus predicts *primacy effects*—early information counts for more—and it pays a persuader to make their case first.

Related Literature

The study of bounded rationality in economics begins with Simon (1955), who pointed to “limits on computational capacity” as a constraint on “actual human choice” and made the reasoning process itself an object of study. A theory of choice should explain not just which decision an agent reaches but the procedure by which they reach it (Simon, 1976), modeled as a search in which the agent moves toward a conclusion one step at a time (Newell and Simon, 1972; see also Anderson, 1993). Economics largely set this procedural approach aside⁵ in favor of two others. The psychological tradition of heuristics and biases, initiated by Kahneman and Tversky, holds that cognitive limitations lead people to mental shortcuts that can produce systematic errors. It generally posits a behavioral tendency as a primitive—present bias, say, or loss aversion—and traces its consequences (Thaler, 1980; Tversky and Kahneman, 1974). A second tradition keeps a Bayesian agent but subjects them to a cognitive constraint: a limit on the information to which they attend, as in rational inattention (Sims, 2003),⁶ a bound on memory (Mullainathan, 2002; Wilson, 2014), or a cost of deliberate thinking (Ilut and Valchev, 2023).⁷

Our approach differs from both traditions. Where the psychological tradition posits behavioral biases as primitives, we take up Simon’s idea of thinking in steps to derive them, reconciling the two: procedure produces what psychology assumes. And our agent is not Bayesian. Indeed they have no priors to guide them.

The work closest to ours organizes around three themes: which facts come to mind; which interpretation of the facts an agent adopts; and how complexity distorts choice.

⁵Aragones et al. (2005) also begin from the observation that an agent may hold all the relevant facts yet not know what follows from them, and show that extracting regularities from data is computationally hard. Theirs is a complexity result about reasoning’s endpoint; we model the reasoning process itself.

⁶Relatedly, Gabaix (2014) models an agent who copes with such limits by attending *sparsely*, optimizing over a deliberately simplified model.

⁷A related strand studies constraints on communication rather than cognition—for instance, Cremer et al. (2007) on the optimal design of an organizational code under limited communication.

The first theme—what comes to mind—is the focus of the body of work by Bordalo, Gennaioli, Shleifer, and coauthors (Bordalo et al., 2023, 2012, 2013a,b, 2016, 2020, 2026; Gennaioli and Shleifer, 2010; Mullainathan et al., 2008a), hereafter BGS. They focus on two faculties: attention and memory. On attention, they posit that people overweight features that are *salient*—prominent, contrasting, or surprising—so that the most salient attributes of a problem loom largest in choice.⁸ On memory, they model how cues shape what an agent recalls, and how those recollections in turn color judgment.

Our framework is complementary, but starts from a different primitive. BGS ask which features an agent attends to or recalls. We ask how many they can bring to bear *at once*. Because we make this combination step explicit, our framework also speaks to phenomena beyond attention and memory—such as why an agent may reach no conclusion at all, and so delay or fall back on a default.⁹

A second theme is which interpretation of the facts an agent adopts. Schwartzstein and Sunderam (2021) model persuasion as the proposing of models (see also Schwartzstein and Sunderam, 2024). A persuader offers a way of organizing past data, and receivers find a model compelling when it explains the data well—as in our framework, where the agent accepts conclusions that have high fit with the facts. In their framework, models arrive whole. In ours they must be built step by step. A persuader can therefore supply only part of an interpretation—a conjecture and a few supporting facts—leaving the agent’s own reasoning to complete it. And because construction is sequential, interpretation is path dependent. An early suggestion steers the conclusions the agent goes on to draw.¹⁰

⁸Closely related are the focusing model of Kőszegi and Szeidl (2013) and the relative-thinking model of Bushong et al. (2021), in which an attribute’s weight depends on how it stands out within the choice set; Section 3.4 contrasts our account of bias with this weighting family. Attention can also entrench a model: Schwartzstein (2014) shows that an agent who does not attend to the right data may never learn their model is wrong (see also Gagnon-Bartsch et al., 2023).

⁹Limited *depth* of reasoning, as in level- k and cognitive-hierarchy models (Costa-Gomes and Crawford, 2006; Stahl and Wilson, 1994), is a related limit, developed for strategic settings, on how far an agent can carry a chain of reasoning. Alaoui and Penta (2016) endogenize depth through a cost-benefit tradeoff, which parallels the possibility, discussed in Section 3, that an agent lowers their extrapolation threshold until a conclusion becomes reachable.

¹⁰A broader literature studies narratives—the vehicles of such persuasion—as causal structures (Eliaz and

A third theme is how complexity distorts choice. A framework that bears closely on ours here is the cognitive-uncertainty account of Enke and Graeber (2023), in which the noise in an agent’s judgments rises with a problem’s complexity, compressing behavior toward a mental default. The accounts differ in what gives rise to the distortion. In the cognitive-uncertainty account the *direction* of any systematic bias is set by the exogenous default. In our framework it arises endogenously from the structure of the problem—by whether an option’s advantages are concentrated on a few facts the agent can hold at once or dispersed across many they cannot (see Section 3.4). The two accounts are complementary, and both fit a growing body of experimental work showing that much of what looks like nonstandard preference is the imprint of computational constraints.¹¹

We also draw on cognitive psychology and computer science: on the step-by-step process view of Newell and Simon (1972), which we embed in an economic model of choice; on the literature on *chunking* (Chase and Simon, 1973; Miller, 1956); and on the intuition of the *minimum description length* literature that people favor explanations which encode the data simply (Chater, 1996; Chater and Vitányi, 2003), since processing capacity puts a premium on economical representations.¹²

Roadmap

The paper proceeds as follows. Section 2 presents the purely deductive version of the model, shows how limited processing capacity and the agent’s stock of concepts bound what they can conclude and how quickly, and distills the analysis into a portable measure of a problem’s difficulty. This agent thinks in steps but makes no mistakes, so the heuristics and biases of the psychological tradition cannot yet appear. Section 3 lets the agent extrapolate—accept conclusions that fit the facts well enough—and shows how this gives rise to choice overload, systematic bias, and unnoticed inconsistency. Extrapolation

Spiegler, 2020), as categorizations (Gibbons et al., 2021; Mullainathan et al., 2008b), or as models reevaluated as data arrives (Wojtowicz, 2024); Bénabou et al. (2018) model their spread.

¹¹See Enke and Graeber (2023), Oprea (2024), and Ambuehl et al. (2022), and for an overview Gabaix and Graeber (2024). Such constraints are what Oprea (2020) measures directly, finding that the cost of implementing a rule rises with the number of elements it requires tracking at once.

¹²From computer science, we draw on database theory (Abiteboul et al., 1995), whose distinction between storage and retrieval parallels ours between knowledge and processing.

thus delivers Kahneman and Tversky's heuristics and biases, microfounded by the procedural model of Section 2. Section 4 shows how the model gives rise to path-dependent thinking and draws out the implications for persuasion. Section 5 sketches two extensions: the selection of the agent's concepts, and their generalization to an internal code. Section 6 concludes. Appendix A lets the agent delete beliefs to resolve inconsistencies, which succeeds only in part and can make thinking cyclic. Proofs are in Appendix B.

2 Deduction

We begin with the deductive version of the model. Throughout the section we use a single running example, which we introduce first.

A Running Example: Mr. Green

Mr. Green holds an asset on which he has accrued a \$400,000 capital gain and must decide whether to realize it this year or next. His ordinary income is \$100,000 this year and \$200,000 next year, and the gain falls entirely in whichever year he sells. Under the ordinary tax, income is taxed at 30% and the gain at a preferential 15%. Under the alternative minimum tax, the preferential rate is disallowed—income and gain are combined into a single base, and the amount above a \$100,000 exemption is taxed at 25%. In each year he owes whichever of the two taxes is larger, and he seeks the year of sale that minimizes his total tax across the two years. Five facts describe his situation: the figures (\$400,000, \$200,000, 30%, etc.); that the gain falls in whichever year he chooses; that the ordinary tax is the tax on income plus the tax on the gain; that the minimum tax combines income and gain into one base and taxes the amount above the exemption; and that in each year he pays the larger of the two taxes.

These five facts are everything. Nothing is hidden, and from them the answer follows with certainty. Mr. Green should wait—realizing the gain next year costs \$155,000 in total tax, against \$160,000 if he realizes it this year.¹³ Yet almost no one can infer the answer

¹³Realizing the gain this year, he pays \$100,000 in year 1 (the minimum tax, which exceeds the \$90,000

immediately. Each fact is plain on its own. The difficulty lies in putting them together. The model that follows captures this difficulty, and we return to Mr. Green throughout to illustrate.

2.1 Model

An agent faces a choice problem in which the optimal choice depends on the state of the world. The state of the world is represented by a set of N vectors,

$$P = \begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_N \end{bmatrix},$$

where the x th vector is $P_x = (p_{x1}, p_{x2}, \dots, p_{xM_x})$. When the vectors share a common length ($M_1 = \dots = M_N = M$), P is an $N \times M$ matrix:

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots & p_{1M} \\ p_{21} & p_{22} & p_{23} & \cdots & p_{2M} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ p_{N1} & p_{N2} & p_{N3} & \cdots & p_{NM} \end{bmatrix}.$$

Each vector element p_{xy} takes values in a finite domain Ω_{xy} . We call P a *picture* and the elements p_{xy} *pixels*. Let $\mathcal{P}$ denote the set of all pictures whose pixels lie in their respective domains—equivalently, the set of all possible states of the world. We write Q for a generic element of $\mathcal{P}$, with pixels q_{xy} .

Picture P might, for instance, be a literal black-and-white image, with each pixel tak-

ordinary tax) and \$60,000 in year 2; realizing it next year, \$30,000 in year 1 and \$125,000 in year 2 (again the minimum tax). Because the minimum tax taxes income at 25% where the ordinary tax takes 30%, its excess over the ordinary tax shrinks as income rises—\$10,000 in the low-income year against \$5,000 in the high-income year.

ing the value *black* or *white* ($\Omega_{xy} = \{\text{black}, \text{white}\}$). For the financial problem faced by Mr. Green, think of the picture as consisting of three vectors. The first represents the problem's *givens*—the facts that hold regardless of when the capital gain is realized: Mr. Green's ordinary income in each of the two years, the capital gain amount, the ordinary tax rates on income and on the gain, and the minimum-tax rate and exemption, which we denote $I_1, I_2, G, \tau_I^{\text{ord}}, \tau_K^{\text{ord}}, \tau^{\text{amt}}$, and E^{amt} . The next two represent *hypotheticals*—the outcomes that depend on Mr. Green's choice—with hypothetical i giving the outcomes that follow from realizing the gain in year i ($i = 1$ or $i = 2$). Each hypothetical vector records the capital gain, ordinary tax, minimum tax, and actual tax realized in each year j , which we denote $G_j(i), T_j^{\text{ord}}(i), T_j^{\text{amt}}(i)$, and $T_j(i)$.

$$P = \begin{bmatrix} \text{givens} = (I_1, I_2, G, \tau_I^{\text{ord}}, \tau_K^{\text{ord}}, \tau^{\text{amt}}, E^{\text{amt}}) \\ \text{hypothetical 1} = (G_j(1), T_j^{\text{ord}}(1), T_j^{\text{amt}}(1), T_j(1))_{j=1,2} \\ \text{hypothetical 2} = (G_j(2), T_j^{\text{ord}}(2), T_j^{\text{amt}}(2), T_j(2))_{j=1,2} \end{bmatrix}$$

Features

We refer to any proper, non-empty subset f of $\mathcal{P}$ as a *feature*, and we say that f *applies* to P if $P \in f$. Let $\mathcal{F}$ denote the set of all features. Each feature is an event in picture space and represents something the agent might know about the picture.

A feature may fix a single pixel, as in $\{Q \in \mathcal{P} : q_{11} = 10\}$, which represents knowing that “ $p_{11} = 10$.” Or it may express a relationship among pixels, as in $\{Q \in \mathcal{P} : T_j(i) = \max(T_j^{\text{ord}}(i), T_j^{\text{amt}}(i)) \text{ for all } i, j\}$, which represents knowing that “in both years, the tax paid equals the larger of the ordinary tax and the minimum tax.”

To formalize the first kind, we call a feature that specifies the value of a single pixel and nothing else a *pixel value*: a feature of the form $\{Q \in \mathcal{P} : q_{xy} = \omega\}$ for some $\omega \in \Omega_{xy}$. We write F_{pixels} for the set of all pixel values and $F_{\text{pixels}}(P)$ for those that apply to P ; in a slight abuse of terminology, we usually refer to $F_{\text{pixels}}(P)$ simply as the *pixels* of P .

Decision-relevant Features

We think of the agent as facing a choice problem in which they must choose an action a from a set A . They have a utility function $u : A \times \mathcal{P} \rightarrow \mathbb{R}$ that depends on their action and the picture they face. The *decision-relevant features* are those that tell the agent which action is optimal. For instance,

$$f_a = \{Q \in \mathcal{P} : u(a, Q) \geq u(a', Q) \text{ for all } a' \in A\}$$

represents the event “ a is an optimal action.”¹⁴ In Mr. Green’s case, the decision-relevant feature is “the total tax is higher if the gain is realized in year 1 than in year 2” ($T_1(1) + T_2(1) \geq T_1(2) + T_2(2)$). Learning this feature tells Mr. Green that it is optimal to wait and realize the gain in year 2.

We say the agent is ready to *choose* action a once they deduce f_a . Our focus will be on determining which features the agent can learn and how long it takes to learn them—and, in particular, on whether they reach the decision-relevant features that resolve their choice.

Knowledge and Concepts

The agent’s *knowledge* is the set of features they think apply to P . We denote the agent’s knowledge at time t by K_t . The agent may add to knowledge over time, in discrete steps.

An important assumption we make is that not all features are knowable. We denote by $C \subset \mathcal{F}$ the set of knowable features. We refer to features $f \in C$ as *concepts* and set C as the agent’s *conceptualization*.

That people find it difficult to conceptualize certain features of data accords with a large body of research in cognitive psychology. For instance, it has been shown that people have trouble identifying faces when they are shown upside-down—a phenomenon known as the face inversion effect (Farah et al., 1998; Yin, 1969). In financial settings such as Mr. Green’s, investors find certain ideas tricky—compound interest, say—and this may prevent them from incorporating those ideas into their thinking.

¹⁴We assume the choice problem is non-degenerate: each action a is optimal in some but not all pictures. This makes each f_a a proper, nonempty subset of $\mathcal{P}$, and hence a feature in the sense above.

We will see that the agent’s conceptualization plays an important role in their reasoning—and in which conclusions they can and cannot reach. We treat it as exogenous for now, and discuss later how it might be endogenized (see Section 5). We assume the agent can conceptualize every pixel value ($F_{\text{pixels}} \subseteq C$)—a pixel value being among the most elementary facts one can state about the picture—and every decision-relevant feature ($f_a \in C$ for all $a \in A$).

The Agent’s Initial Knowledge

At the outset, the agent holds some initial knowledge K_0 about the picture, which we assume is conceptualized ($K_0 \subseteq C$).¹⁵ For a literal image, the agent might know every pixel value, $K_0 = F_{\text{pixels}}(P)$, as one would from simply viewing the image. Mr. Green’s initial knowledge consists of the five facts describing his problem, which we now state formally as F1–F5. F1 collects the givens—seven pixel values, one per pixel of the givens vector, each a separate feature—while F2–F5 are each a single feature relating the pixels:

F1. The givens: $I_1, I_2, G, \tau_I^{\text{ord}}, \tau_K^{\text{ord}}, \tau^{\text{amt}}, E^{\text{amt}}$.

F2. “ $G_i(i) = G$, and $G_j(i) = 0$ for $j \neq i$.”

F3. “ $T_j^{\text{ord}}(i) = \tau_I^{\text{ord}} I_j + \tau_K^{\text{ord}} G_j(i)$.”

F4. “ $T_j^{\text{amt}}(i) = \tau^{\text{amt}} \max(I_j + G_j(i) - E^{\text{amt}}, 0)$.”

F5. “ $T_j(i) = \max(T_j^{\text{ord}}(i), T_j^{\text{amt}}(i))$.”

Notice that Mr. Green initially knows none of the hypotheticals—yet the givens together with F2–F5 pin the picture down completely, determining those hypotheticals as well: intersecting every feature in K_0 leaves a single picture, $\bigcap_{f \in K_0} f = \{P\}$. Everything about the problem is therefore deducible from what he knows at the outset, including the conclusion that matters—which year leaves the smaller total tax. What Mr. Green lacks

¹⁵Since the agent adds to their knowledge only features they can conceptualize, their knowledge remains conceptualized at every stage: $K_t \subseteq C$ for all t .

initially is not information but a particular *feature* of the picture—the one on which his decision turns.

The visual case differs only in where the gap falls. There, an agent may know every pixel of the picture—in that sense holding complete information—yet still not know an important feature such as “the picture depicts a rabbit.” In both cases the agent’s initial knowledge already determines the picture, yet leaves further features to be drawn out.

Learning

The agent learns about the picture through a sequence of deductions. We denote the agent’s thought at time t by θ_t . A thought is composed of two elements: $\theta_t = (\phi_t, D_t)$. We refer to the first element $\phi_t \in C$ as the agent’s *conjecture*. The second element $D_t \subseteq K_t$ consists of the facts the agent draws on in evaluating the conjecture—facts already known to apply to the picture. The agent adds the conjecture to knowledge if it is deducible from these facts.

To give an example, suppose Mr. Green is considering the case in which he realizes the capital gain this year, and draws on three facts he already knows: “the year-1 ordinary tax is \$90,000,” “the year-1 minimum tax is \$100,000,” and “the tax paid in each year is the larger of the ordinary tax and the minimum tax.” He would add “the year-1 tax is \$100,000” to knowledge if this were his conjecture.

Formally, ϕ_t is deducible if $\bigcap_{f \in D_t} f \subseteq \phi_t$. The agent’s knowledge at time $t + 1$ is $K_{t+1} = K_t \cup \{\phi_t\}$ if ϕ_t is deducible, and $K_{t+1} = K_t$ otherwise.

Limited Processing Capacity

Critically, the agent has limited *processing capacity*: they can draw on at most L facts at once: $|D_t| \leq L$. As a result, the agent can use only a limited share of their knowledge at any one time.

2.2 Which conclusions are reachable?

We now ask which conclusions the agent can reach, and how many steps it takes to reach them. These are two senses in which a problem may be difficult: a conclusion can lie beyond the agent's reach altogether, or within reach but only after many steps. Two forces shape both: the agent's processing capacity L and their conceptualization C . We first give two definitions, then present two propositions that isolate the role of each.

Throughout, we remain agnostic about which thought occurs to the agent at time t : which conjecture they entertain, and which facts they bring to bear. Our results thus concern which conclusions are within the agent's reach, and the fewest steps they require—not which thoughts the agent lights on along the way.¹⁶

Definition 1. Let $C(P)$ denote the concepts that apply to P and suppose $f \in C(P)$.

1. f is reachable in τ steps ($f \in R_\tau$) if there exists a sequence of thoughts $\theta_0, \dots, \theta_{\tau-1}$ such that $f \in K_\tau$.
2. f is reachable ($f \in R$) if it is reachable in τ steps for some finite τ .
3. f is derivative if it follows from a single feature in K_0 : $f \supseteq g$ for some $g \in K_0$.

Definition 2. Knowledge K specifies P if it fully pins down the picture: $\bigcap_{f \in K} f = \{P\}$. Its essential size $s(K)$ is then the size of the smallest subset that still specifies P :

$$s(K) = \min \left\{ |F| : F \subseteq K, \bigcap_{f \in F} f = \{P\} \right\},$$

and $s(K) = \infty$ if K does not specify P .

Two examples are worth noting. Mr. Green's initial knowledge K_0 comprises eleven features (the seven givens of F1, together with F2–F5). These features specify P and have

¹⁶Section 4.2 imposes one restriction on the agent's thoughts—that powerful features command a place in them—and we return to the broader question of what governs thoughts in the conclusion.

essential size eleven: dropping any one leaves something about P unknown. Likewise, if the agent's initial knowledge is the value of each pixel, K_0 fully specifies P and its essential size equals the number of pixels, since dropping any pixel value leaves part of the picture undetermined.

Processing Capacity

The following proposition isolates the role of processing capacity.

Proposition 1.

1. *Suppose K_0 specifies P . Then:*

(i) *if $L \geq s(K_0)$, the agent can reach every concept that applies to P in a single step ($R_1 = C(P)$).*

(ii) *if $L = 1$, only derivative concepts are reachable.*

2. *In the intermediate range $1 < L < s(K_0)$, reachability may be partial and may require multiple steps: for some picture P , knowledge K_0 specifying it, and conceptualization C ,*

(i) *some, but not all, concepts that apply to P are reachable ($K_0 \subset R \subset C(P)$).*

(ii) *some concepts are reachable only in multiple steps ($R_1 \subset R_2$).*

Part 1 covers the boundary cases: with capacity to hold a specifying subset of K_0 at once, everything is reachable in a single step, while with capacity for one fact at a time, nothing can be built beyond what each fact already implies. Part 2 is the substantive intermediate case, in which some concepts are out of reach and others reachable only in several steps.

An Illustration of Multi-Step Reasoning

To understand Part 2 of the proposition, consider how Mr. Green might reach the conclusion that " $T_1(1) = \$100,000$ "—that he owes \$100,000 this year if he realizes the gain this year.

Reaching this conclusion requires combining ten facts: the six givens that bear on this year's tax—every given in F1 except next year's income, " $I_2 = \$200,000$ "—together with F2–F5. If $L \geq 10$, Mr. Green can hold all ten in mind at once and deduce " $T_1(1) = \$100,000$ " in a single step.

Suppose instead that $L = 5$. He can no longer reach the conclusion in one step, since that would require drawing on all ten facts simultaneously. But he can still reach it in four steps, drawing on the individual givens of F1 and the rules F2–F5, at each stage computing an intermediate quantity and carrying it forward in place of the facts that produced it:

1. Mr. Green draws on " $G = \$400,000$ " and F2 (two facts) to conclude that the gain falling in year 1 is " $G_1(1) = \$400,000$."
2. He draws on " $I_1 = \$100,000$," " $\tau_I^{\text{ord}} = 30\%$," " $\tau_K^{\text{ord}} = 15\%$," " $G_1(1) = \$400,000$," and F3 (five facts) to conclude that the ordinary tax is " $T_1^{\text{ord}}(1) = \$90,000$."
3. He draws on " $I_1 = \$100,000$," " $\tau^{\text{amt}} = 25\%$," " $E^{\text{amt}} = \$100,000$," " $G_1(1) = \$400,000$," and F4 (five facts) to conclude that the minimum tax is " $T_1^{\text{amt}}(1) = \$100,000$."
4. He draws on " $T_1^{\text{ord}}(1) = \$90,000$," " $T_1^{\text{amt}}(1) = \$100,000$," and F5 (three facts) to conclude that " $T_1(1) = \$100,000$."

Multi-step reasoning works here because each step replaces several facts with a single conclusion that stands in for them: once Mr. Green has the ordinary tax from Step 2, he carries " $T_1^{\text{ord}}(1) = \$90,000$ " forward and no longer needs the income, the rates, and the gain that produced it. He can thus operate within a processing capacity far smaller than the ten facts the conclusion nominally depends on, by reformulating the problem in terms of intermediate concepts—the gain in a given year, the ordinary tax, the minimum tax—that are more economical than the raw facts for the task at hand.

Conceptualization Ability

Proposition 1 focused on the role of processing capacity, taking the agent’s conceptualization as given. The following proposition highlights the role of the agent’s conceptualization. Let $R(C)$ denote the set of reachable concepts under conceptualization C , and let $k(f, C)$ denote the minimal number of steps required to reach f given C . We obtain the following proposition.

Proposition 2. *Fix a picture P and initial knowledge K_0 specifying P .*

1. *Expanding the set of concepts reduces the number of steps required to reach any previously reachable concept. That is, for any C and $C' \supset C$ and any $f \in R(C)$, $k(f, C') \leq k(f, C)$. The inequality is strict for some picture P , K_0 , $C \subset C'$, and $f \in R(C)$.*
2. *Expanding the set of concepts enlarges what is reachable among the original concepts. That is, for any C and $C' \supset C$, $R(C') \cap C \supseteq R(C)$. The inclusion is strict for some picture P , K_0 , and $C \subset C'$.*
3. *If $L > 1$ and the set of concepts contains every feature that applies to P , then every such feature is reachable: $R(C) = C(P)$.*

Part 1 formalizes the idea that richer conceptualizations speed deduction. Even when a concept is already reachable under C , a richer conceptualization $C' \supset C$ can shorten the path to it. For example, an intermediate concept such as “the ordinary tax on Mr. Green’s year-1 income is \$30,000,” once computed, can be stored and drawn on wherever it is needed, sparing him from reassembling it from the underlying facts each time.

Part 2 says that expanding the conceptualization makes new deductions possible. A concept can be unreachable simply because the agent lacks the intermediate concepts needed to bridge the gap between their initial facts and the target. A concrete illustration is given below.

Part 3 describes an extreme case: with processing capacity of two or more and a conceptualization rich enough to contain every feature that applies to P , every such feature is reachable—the agent can fold known features together two at a time until P is pinned

down. Some features may nonetheless be reachable only after a great many steps, and in this sense be *close* to unreachable; how many steps a conclusion requires, and how difficult a problem therefore is, is taken up in Section 2.4.

An Illustration of the Conceptualization's Role

To illustrate Part 2—that a richer conceptualization can put otherwise-unreachable conclusions within reach—we return to Mr. Green.

In the previous illustration, Mr. Green reached the conclusion “ $T_1(1) = \$100,000$ ” with a processing capacity of $L = 5$. If his only concepts are the pixel values and the rules F2–F5, then $L = 5$ is in fact the smallest capacity at which that conclusion is reachable. Additional concepts relax this requirement. With three of them, for instance, $L = 3$ suffices:

1. “ $\tau_I^{\text{ord}} I_1 = \$30,000$ ”—the ordinary tax on his year-1 income;
2. “ $\tau_K^{\text{ord}} G_1(1) = \$60,000$ ”—the ordinary tax on the gain, were it realized in year 1;
3. “ $I_1 + G_1(1) - E^{\text{amt}} = \$400,000$ ”—the amount subject to the minimum tax in year 1, were the gain realized then.

Each concept packages a single arithmetic operation. Mr. Green can reach the conclusion in the following steps:

1. He draws on “ $G = \$400,000$ ” and F2 (two facts) to conclude that “ $G_1(1) = \$400,000$.”
2. He draws on “ $I_1 = \$100,000$ ” and “ $\tau_I^{\text{ord}} = 30\%$ ” (two facts) to conclude that “ $\tau_I^{\text{ord}} I_1 = \$30,000$.”
3. He draws on “ $\tau_K^{\text{ord}} = 15\%$ ” and “ $G_1(1) = \$400,000$ ” (two facts) to conclude that “ $\tau_K^{\text{ord}} G_1(1) = \$60,000$.”
4. He draws on “ $\tau_I^{\text{ord}} I_1 = \$30,000$,” “ $\tau_K^{\text{ord}} G_1(1) = \$60,000$,” and F3 (three facts) to conclude that “ $T_1^{\text{ord}}(1) = \$90,000$.”

5. He draws on " $I_1 = \$100,000$," " $G_1(1) = \$400,000$," and " $E^{\text{amt}} = \$100,000$ " (three facts) to conclude that " $I_1 + G_1(1) - E^{\text{amt}} = \$400,000$."
6. He draws on " $\tau^{\text{amt}} = 25\%$," " $I_1 + G_1(1) - E^{\text{amt}} = \$400,000$," and F4 (three facts) to conclude that " $T_1^{\text{amt}}(1) = \$100,000$."
7. He draws on " $T_1^{\text{ord}}(1) = \$90,000$," " $T_1^{\text{amt}}(1) = \$100,000$," and F5 (three facts) to conclude that " $T_1(1) = \$100,000$."

Notice that while this thought process reaches the conclusion using less processing capacity, it now takes seven steps rather than four, reflecting a tradeoff familiar from computer science: between the time a computation takes and the space it requires.¹⁷

Taken together, Propositions 1 and 2 show that bounded rationality arises from two distinct sources: limited processing capacity and limited conceptualization. They also reveal a sense in which the two act as substitutes. With enough processing capacity, the agent can deduce any applicable concept directly from their initial facts, with no intermediate concepts at all. With a rich enough conceptualization, even very limited processing capacity can suffice. Intermediate concepts let the agent accomplish over several steps what would otherwise require holding many facts at once.

2.3 Chunking

Short-term memory, cognitive psychologists find, is limited less by the volume of information than by how it is organized. Miller (1956) introduced the term *chunking* for the binding of individual pieces of information into a smaller number of meaningful units that can be held more readily in mind. A canonical example is the telephone number: the same digits are far easier to remember when organized into chunks (e.g. 512-3482) than when presented as an unstructured sequence (e.g. 5-1-2-3-4-8-2).

¹⁷For example, a Turing machine can often compute a result using fewer steps if it is allowed more tape to work on, and more steps if its tape is restricted. Processing capacity plays the role of space and the number of steps the role of time.

A more striking illustration comes from chess. Chase and Simon (1973) showed expert and novice players a board position for five seconds and asked them to reconstruct it from memory.¹⁸ On positions from actual games, experts recalled about 16 of 20 pieces, novices about 4—the standard interpretation being that experts have learned to chunk familiar configurations.

These findings map cleanly onto our model. Processing capacity limits the number of facts an agent can draw on at once, not how much information each fact carries, and chunking is the folding of several facts into a single concept. Mr. Green’s two income figures, “ $I_1 = \$100,000$ ” and “ $I_2 = \$200,000$,” are separate facts, so drawing on both at once costs two units of capacity; an agent whose conceptualization includes the single concept “Mr. Green’s income is \$100,000 this year and \$200,000 next year” folds them into one, drawing on the same information at a cost of one unit.

Figure 2 casts the experiment in these terms. A 5×10 black-and-white picture stands in for the board (Panel a). An expert whose conceptualization includes “the left-hand side is an H” and “the right-hand side is an I” holds the entire board in two facts (Panel b); a novice who conceptualizes only pixel values needs fifty, far beyond any plausible L . The expert’s edge is a difference in conceptualization, not processing capacity.

Chase and Simon also found that on random boards the expert’s advantage disappeared, both groups recalling only about three pieces on average. Our model explains why: the random position matches no configuration the expert’s concepts encode, so he too must fall back on the individual pixels, no better off than the novice.



Figure 2: Chunking a Game Board

¹⁸Chase and Simon (1973) built on earlier work on chunking in chess by De Groot (1965).

2.4 A portable measure of problem difficulty

How difficult is one decision problem compared with another? In our model, the answer depends both on the nature of the problem and on the concepts the agent brings to it (C). The same chess problem may be easy for an expert and hopelessly difficult for a novice, with the difference lying in the concepts available to them. While concepts play an important part, we can isolate the role of the problem itself by assuming the agent has every concept at hand ($C = \mathcal{F}$). Doing so yields a highly portable version of the model—one easily applied to virtually any setting of interest.

With every concept available, every feature that follows from the agent’s initial knowledge is reachable provided $L \geq 2$ (Proposition 2), so reachability no longer separates one problem from another. What separates them is *time*: a reachable conclusion may lie one step away or a great many—and in the latter case it is, in practical terms, close to unreachable. We therefore measure a problem’s difficulty by the number of steps it takes to solve.

The following definitions make this precise. A *problem* is a pair $x = (K_0, \{f_a\}_{a \in A})$, where K_0 is what the decision-maker knows at the outset and $\{f_a\}_{a \in A}$ are the decision-relevant features—the features that settle which action is best. The problem is *well-posed* if some decision-relevant feature follows from the initial knowledge ($\bigcap_{f \in K_0} f \subseteq f_a$ for some a), and *nontrivial* if no such feature is already among the agent’s facts ($f_a \not\subseteq K_0$ for all a). We assume both throughout.¹⁹ With all concepts available to the agent, we define the problem’s *difficulty* $D(x)$ as the fewest steps in which the agent can reach one of the decision-relevant features: that is, the fewest steps required to make a choice.

Difficulty turns on a single quantity: the number of initial facts on which the conclusion rests. For a feature f , let

$$s(K_0, f) = \min \left\{ |F| : F \subseteq K_0, \bigcap_{g \in F} g \subseteq f \right\}$$

¹⁹If K_0 specifies P , the problem is automatically well-posed: some action is optimal at P , and $\bigcap_{f \in K_0} f = \{P\} \subseteq f_a$ for that action.

denote the *essential premises* of f : the size of the smallest set of initial facts from which f can be deduced.²⁰

Proposition 3. Suppose $L \geq 2$ and $C = \mathcal{F}$. For any feature $f \notin K_0$ that applies to P and follows from the agent's initial knowledge, the fewest steps in which f is reachable is $\max\{1, \lceil (s(K_0, f) - 1) / (L - 1) \rceil\}$. Consequently, the difficulty of any well-posed, nontrivial problem $x = (K_0, \{f_a\}_{a \in A})$ is

$$D(x) = \max\left\{1, \left\lceil \frac{s^* - 1}{L - 1} \right\rceil\right\}, \quad \text{where } s^* = \min_{a \in A} s(K_0, f_a).$$

In short: difficulty is essential premises, compressed at the rate capacity permits. The logic is that of chunking. Each thought folds up to L known items into one conclusion that stands in for them—a net compression of $L - 1$ per step—and reaching a conclusion that rests on $s(K_0, f)$ premises means compressing them into one. The proposition says the folding count is exact. With every feature conceptualizable, the agent can always form the intermediate conclusions the folding requires. It also makes the time–capacity tradeoff of Section 2.3 precise. Mr. Green's decision-relevant conclusion—that the total tax is lower if the gain is realized in year 2—rests on all eleven of his initial facts. Both hypotheticals bear on the comparison, and no fact can be spared. At $L = 11$ the conclusion takes one step, at $L = 5$ three, at $L = 2$ ten.²¹

To see the measure at work, suppose $L = 2$ and let the picture be a single vector of four binary pixels—four checks on a loan application, each reading *pass* or *fail*: adequate income, adequate collateral, clean credit history, stable employment. The file is good: all four checks pass, and the lender's initial knowledge is the four pixel values. Now compare two lending standards. Under a *conjunctive* standard, approval is optimal just when all four checks pass: the decision-relevant feature is the event that every pixel reads *pass*, its essential premises are all four facts, and readiness takes $\lceil 3/1 \rceil = 3$ steps—the lender

²⁰ $s(K_0, f) = \infty$ if no subset suffices. The notion extends essential size: when K_0 specifies P , every specifying subset implies every feature that applies, so $s(K_0, f) \leq s(K_0)$ for all such f .

²¹By contrast, the earlier illustrations needed four steps at $L = 5$, and seven at $L = 3$, to reach the intermediate conclusion " $T_1(1) = \$100,000$," which rests on ten facts (all but next year's income). Their extra steps reflect Mr. Green's limited conceptualization there, whereas the proposition assumes every feature is available as a stepping stone.

must fold the checks together, two at a time, before the file as a whole is approved. Under a *sufficient-condition* standard—approval is optimal whenever the first check passes, as when a loan carries a government guarantee—the decision-relevant feature follows from a single fact, and one step suffices. Same file, same facts. The difficulty resides not in the data but in the relation between the initial knowledge and the feature on which the decision turns.

This comparison scales, and it generates some interesting economics. A conjunctive standard with m conditions has difficulty $\lceil (m - 1)/(L - 1) \rceil$, growing with the length of the checklist. A sufficient-condition standard has difficulty one regardless of m . The two also treat good and bad news asymmetrically. Under the conjunctive standard a single failed check settles the matter, so establishing compliance is slow while spotting a violation is quick. Verifying that a borrower meets every underwriting criterion, or a firm every covenant, is laborious because the conclusion rests on many premises at once, while flagging a disqualifying condition rests on one. And institutions respond by reducing the premise count down. Safe harbors, bright-line rules, and categorical eligibility replace a many-premise standard with a one-premise substitute. A taxpayer need only check that payments matched last year's liability, an applicant that they already receive a qualifying benefit. In our terms, such rules lower s^* , and with it the number of steps to readiness.

This is what makes the measure portable. In any setting, one need only write down what the decision-maker knows at the outset, identify the feature on which the decision turns, and count the fewest initial facts from which it follows. Processing capacity then sets the rate at which those premises fold together.

3 Extrapolation

Deduction may fail the agent in three ways. Sometimes the conclusion follows from their initial facts but lies beyond their reach. Mr. Green, lacking the right intermediate concepts or the capacity to hold enough facts at once, may never deduce which year to sell. Other

times the initial facts fail to pin down the picture fully—and what they leave unknown, no deduction can establish. And even when a conclusion is within reach, deducing it may take many steps—seven, in one version of Mr. Green’s problem—each costing time and effort.

With this in mind, we consider the possibility that the agent supplements deduction with *extrapolation*. Under deduction, the agent accepts a conjecture ϕ_t only when it is logically implied by the facts they draw on, D_t . Under extrapolation, they accept it when it is merely *suggested* by those facts. Extrapolation addresses all three limitations of deduction. It lets the agent reach conclusions that follow from their knowledge but lie beyond what limited processing capacity can deduce. It lets them *predict*. When their initial knowledge does not fully specify the picture ($\bigcap_{f \in K_0} f \supset \{P\}$), extrapolation supports an educated guess at what is left open—filling in blanks the facts alone leave empty. And it economizes on thought. A conclusion that deduction reaches only after many steps can be reached in fewer, so the agent decides faster and with less effort. Each of these extends the agent’s reach, but they come at a price. To extrapolate is to attend selectively—to draw on some features of the problem and set the rest aside—and the same license to go beyond the facts opens the door to error. As we will see, the errors are not merely random noise. Extrapolation introduces biases that are systematic and predictable, and it can leave the agent’s beliefs inconsistent in ways they may not detect.

3.1 Model

To capture extrapolation, we relax the requirement that a conjecture ϕ_t be *deducible* from the facts D_t . Instead, the agent adopts ϕ_t whenever it accords well enough with those facts: they add ϕ_t to knowledge if its *fit* with D_t , denoted $\Psi(\phi_t; D_t)$, weakly exceeds a threshold $\alpha \in (0, 1]$.

We are not wedded to a particular functional form for $\Psi(\cdot)$, but we adopt the following. Let $\delta(S) \subseteq \mathcal{P}$ be the set of pictures consistent with every feature in S ($\delta(S) = \bigcap_{f \in S} f$) and let $\psi(S) = \log |\mathcal{P}| - \log |\delta(S)|$ measure how strongly S restricts the feasible pictures.

The fit of ϕ_t with facts D_t is then

$$\Psi(\phi_t; D_t) = 1 - \frac{\psi(D_t \cup \{\phi_t\}) - \psi(D_t)}{\psi(\{\phi_t\})}.$$

Intuitively, $\psi(D_t \cup \{\phi_t\}) - \psi(D_t)$ is the *incremental restriction* induced by adding ϕ_t on top of D_t , while $\psi(\{\phi_t\})$ is the restriction induced by ϕ_t on its own. Thus $\Psi(\phi_t; D_t)$ is a normalized measure of how “redundant” ϕ_t is given what is already known.²²

This functional form has several attractive properties:

1. Fit is higher when ϕ_t holds in a larger fraction of pictures consistent with D_t , i.e. when $\psi(D_t \cup \{\phi_t\}) - \psi(D_t)$ is small.
2. Fit attains its upper bound of one ($\Psi(\phi_t; D_t) = 1$) if and only if ϕ_t is *deducible* from D_t (equivalently, $\delta(D_t) \subseteq \delta(\{\phi_t\})$), because then $\delta(D_t \cup \{\phi_t\}) = \delta(D_t)$.
3. Fit equals $-\infty$ if D_t and ϕ_t are incompatible ($\delta(D_t \cup \{\phi_t\}) = \emptyset$).
4. Fit equals zero when D_t is uninformative about ϕ_t (for instance, if $D_t = \emptyset$).

Finally, note that when $\alpha = 1$, the agent adds ϕ_t if and only if it is deducible from D_t . Thus the extrapolation model nests the deduction-only model as the case $\alpha = 1$.

Assessing fit might seem too demanding a calculation for a boundedly rational agent—taken literally, Ψ counts the pictures consistent with a set of features. But that is not how we think of it. Fit is better understood as a judgment delivered to the agent than as a calculation they carry out: the felt sense that a conjecture accords with the facts in mind.²³

²²More precisely, under the uniform distribution over pictures, $\log |\delta(S)|$ is the Boltzmann entropy of the posterior support given S , and $\psi(S) = \log |\mathcal{P}| - \log |\delta(S)|$ is the information gained by imposing S . Then $\psi(D_t \cup \{\phi_t\}) - \psi(D_t) = \log |\delta(D_t)| - \log |\delta(D_t \cup \{\phi_t\})|$ is the information gain from adding ϕ_t conditional on D_t , and $\Psi(\phi_t; D_t)$ normalizes this conditional gain by the unconditional informativeness of ϕ_t .

²³The distinction here—between performing an operation and gathering the pieces it works on—is a familiar one in computer science and cognitive science (Anderson, 1993; Anderson et al., 2004; Backus, 1978; Newell, 1990). A processor runs even an elaborate instruction in a single step, once its inputs are gathered; what is scarce is how many it can take at once. Computer scientists call this the *von Neumann bottleneck*—the channel feeding the processor, not the processor, is the limit (Backus, 1978). Cognitive scientists model minds the same way: in architectures such as ACT-R, built to simulate human reasoning, the binding constraint is the few items held in working memory at once; operations on them—retrieving a match, checking a pattern—run automatically.

3.2 Product Choice

To explore how extrapolation works, we turn to a choice setting of particular interest: the choice among products. A consumer must decide which of several products to buy, each described by its ratings on a common set of attributes. There are n products and m attributes. The picture is an $n \times m$ matrix

$$P = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1m} \\ v_{21} & v_{22} & \cdots & v_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ v_{n1} & v_{n2} & \cdots & v_{nm} \end{bmatrix}, \quad v_{xy} \in \{1, \dots, V\},$$

where v_{xy} is product x 's rating on attribute y , and a higher rating is always preferable. The consumer values a product by the total of its ratings, $u(x, Q) = \sum_{y=1}^m v_{xy}$, and to choose must reach one of the decision-relevant features

$$f_i = \{Q : u(i, Q) \geq u(x, Q) \text{ for all } x\},$$

which holds when product i is weakly preferred to every other.

The consumer's initial knowledge consists of every rating ($K_0 = F_{\text{pixels}}(P)$), but they must still work out which product is best. We assume a processing capacity of $L < n \times m$ and, for simplicity, no intermediate concepts. Without intermediate concepts to build on, deducing the best product generally requires holding all $n \times m$ ratings at once, which exceeds the consumer's capacity.²⁴ To choose, then, the consumer must extrapolate.

Mr. Stitch

For illustration, let us begin with the simplest version of the problem. Mr. Stitch is deciding which of two suits to buy, each described by its price and quality. His picture is the

²⁴This is so except for extreme profiles—for instance, a product rated V on every attribute is weakly best whatever the other ratings, and can be certified without them. Setting such cases aside, deduction requires the full picture.

2×2 matrix

$$P = \begin{array}{cc} & \begin{array}{cc} \text{Price Rating} & \text{Quality Rating} \end{array} \\ \begin{array}{c} \text{Suit 1} \\ \text{Suit 2} \end{array} & \begin{bmatrix} 10 & 3 \\ 2 & 8 \end{bmatrix} \end{array}$$

Figure 3: Mr. Stitch's Problem

Ratings run from 1 to 10 ($V = 10$), and Mr. Stitch has processing capacity two ($L = 2$).

How might Mr. Stitch use extrapolation to reach a choice between suits? Lacking intermediate concepts, the only extrapolations open to him run directly from pixels to a conclusion about which suit is best. He might base the extrapolation on a single *attribute*, drawing on both suits' ratings for it, or on a single *suit*, drawing on its ratings across both attributes. We examine each in turn.²⁵

Mr. Stitch: Attribute route

Suppose Mr. Stitch focuses on a single attribute. Selective attention of this kind figures centrally in the work of BGS, though here it arises from limited processing capacity rather than from the salience of particular attributes. Say he attends to price, drawing on the ratings "10" and "2." What can he conclude, if anything?

Consider first the conjecture that "Suit 1 is weakly preferred" ($\phi_t = f_1$). Its fit with the two price ratings is $\Psi(f_1; D_t) \approx 0.98$.²⁶ The fit is near one because, given Suit 1's large price advantage, there is just one case in which Suit 2 comes out ahead. Provided his threshold satisfies $\alpha \leq 0.98$, Mr. Stitch can reach the conclusion that "Suit 1 is weakly

²⁵Mr. Stitch could also extrapolate from a single rating, or from a pair taking one rating from each suit on *different* attributes—one suit's price against the other's quality, say.

²⁶Explicitly,

$$\Psi(f_1; D_t) = 1 - \frac{\psi(D_t \cup \{f_1\}) - \psi(D_t)}{\psi(\{f_1\})} = 1 - \frac{(\log 10^4 - \log 99) - (\log 10^4 - \log 10^2)}{\log 10^4 - \log \frac{1}{2}(10^4 + 670)} \approx 0.98,$$

where $|\mathcal{P}| = 10^4$, as each of the four pixels takes ten values; $|\delta(D_t)| = 10^2$, since the two price ratings pin down two of them; $|\delta(D_t \cup \{f_1\})| = 99$, since of the 10^2 pictures consistent with D_t exactly one has Suit 2 strictly preferred; and $|\delta(\{f_1\})| = \frac{1}{2}(10^4) + \frac{1}{2}(670) = 5335$, the number of pictures in which Suit 1 is weakly preferred.

preferred” and choose Suit 1.²⁷

Can he instead reach the conclusion “Suit 2 is weakly preferred” (feature f_2)? Evaluated on the same two price ratings, 10 and 2, the fit of f_2 is -4.58 —negative, because those ratings are evidence *against* Suit 2, not for it. That extrapolation is unavailable.

More generally, the fit of “Suit i is weakly preferred,” evaluated on the two suits’ ratings for an attribute, depends on how far Suit i leads on it (Table 1): the smaller its advantage, the lower the fit. For a given threshold α , then, Suit i ’s advantage on the attribute must be large enough for the extrapolation to go through.²⁸

Suit i ’s rating advantage	Fit of “ i preferred”	Fit of “ j preferred”
9	1	-6.33
8	0.98	-4.58
7	0.95	-3.48
6	0.90	-2.66
5	0.83	-2.02
4	0.74	-1.48
3	0.62	-1.03
2	0.48	-0.62
1	0.29	-0.27
0	0.05	0.05

Table 1: Fit as a function of Suit i ’s rating advantage on an attribute

Table 1 shows that had Mr. Stitch attended instead to quality—where Suit 2 leads by five—he could have reached “Suit 2 is weakly preferred” (f_2), provided $\alpha \leq 0.83$.

In sum, with a very high threshold ($\alpha > 0.98$) Mr. Stitch cannot extrapolate to either conclusion; with a moderate one ($0.83 < \alpha \leq 0.98$) he can reach only “Suit 1 is weakly preferred,” by attending to price; and with a low one ($\alpha \leq 0.83$) he can reach either, depending on the attribute to which he attends. His choice is then to some degree indeterminate, which accords with experimental evidence that choices follow attention: subjects choose one gamble yet price the other higher, since pricing draws attention to payoffs

²⁷Here Mr. Stitch selects an option on the basis of an attribute. A richer version of the problem would also let him *reject* an option on the basis of an attribute: this only requires that he has intermediate concepts of the form “the suits in S are strictly worse than those not in S .” Tversky (1972) argues that such sequential elimination is common in complex choice settings.

²⁸When the two ratings are equal, both f_1 and f_2 have positive fit (0.05): because each f_i expresses a *weak* preference, the two features overlap rather than partition, so their intersection is non-empty.

while choosing draws it to probabilities (Lichtenstein and Slovic, 1971; Grether and Plott, 1979; Tversky et al., 1988), and, even with the task held fixed, subjects often answer the same question differently across repetitions (Agranov and Ortoleva, 2017, 2022).

Mr. Stitch: Product route

Rather than focus on one attribute, Mr. Stitch might focus on one suit—say Suit 1, drawing on its two ratings, “10” and “3.” The fit of the conjecture “Suit 1 is weakly preferred” is then

$$\Psi(f_1; D_t) = 1 - \frac{\psi(D_t \cup \{f_1\}) - \psi(D_t)}{\psi(\{f_1\})} = 1 - \frac{(\log 10^4 - \log 72) - (\log 10^4 - \log 10^2)}{\log 10^4 - \log \frac{1}{2}(10^4 + 670)} \approx 0.48.$$

The calculation is as before, except that 72 replaces 99: the number of Suit 2 rating combinations at which Suit 1 remains weakly preferred.

More generally, the fit of “Suit i is weakly preferred” rises with Suit i ’s value (Table 2). So for a given threshold α , Suit i ’s value must clear a certain level before the extrapolation “Suit i is weakly preferred” becomes feasible.

Value of i	Fit of “ i preferred”	Fit of “ j preferred”	Value of i	Fit of “ i preferred”	Fit of “ j preferred”
20	1	-6.33	10	-0.27	0.29
19	0.98	-4.58	9	-0.62	0.48
18	0.95	-3.48	8	-1.03	0.62
17	0.90	-2.66	7	-1.48	0.74
16	0.83	-2.02	6	-2.02	0.83
15	0.74	-1.48	5	-2.66	0.90
14	0.62	-1.03	4	-3.48	0.95
13	0.48	-0.62	3	-4.58	0.98
12	0.29	-0.27	2	-6.33	1
11	0.05	0.05			

Table 2: Fit as a function of Suit i ’s value

This kind of extrapolation amounts to *satisficing* in the sense of Simon (1955): Mr. Stitch is willing to choose Suit i once its value is high enough to be “good enough.” Table 2 shows the mirror case as well—as Suit i ’s value falls, the fit of “Suit j is weakly preferred” rises—so if Suit i ’s value drops below a threshold, Mr. Stitch can instead extrapolate to “Suit j is weakly preferred,” and choose the other suit.

A Generalization

The patterns we see in the Mr. Stitch example carry over to any number of products and attributes, as the following proposition demonstrates.

Proposition 4. *Consider the choice among n products described by m attributes, with ratings in $\{1, \dots, V\}$, and a consumer with no intermediate concepts.*

1. (Attribute route) *Suppose the consumer has processing capacity $L = n$ and attends to a single attribute y , drawing on the n ratings $v_{1y}, \dots, v_{ny}$. Let $\Delta_{ij}^y = v_{iy} - v_{jy}$ be product i 's advantage over product j on attribute y . The fit of “product i is weakly preferred” depends on these ratings only through the advantages $\{\Delta_{ij}^y\}_{j \neq i}$, and is strictly increasing in each. Hence, for a given threshold α , attending to attribute y yields the conclusion “product i is weakly preferred” if and only if product i 's advantages $\{\Delta_{ij}^y\}_{j \neq i}$ are large enough.*
2. (Product route) *Suppose the consumer has processing capacity $L = m$ and attends to a single product, drawing on its m ratings. Then:*
 - (i) *Attending to product i , the fit of “product i is weakly preferred” depends on these ratings only through its value $u_i = \sum_y v_{iy}$, and is strictly increasing in u_i . Hence, for a given threshold α , attending to product i yields “product i is weakly preferred” if and only if u_i is high enough.*
 - (ii) *Attending instead to a rival product j , the fit of “product i is weakly preferred” depends on j 's ratings only through its value u_j , and is strictly decreasing in u_j . Hence, for a given threshold α , the consumer can extrapolate to “product i is weakly preferred” and choose i only if u_j is low enough.*

3.3 Struggling to Choose

To make a decision, a consumer must work from their starting facts to a conclusion about which product is best. When they cannot—when neither deduction nor extrapolation

delivers such a conclusion—they struggle to choose.²⁹

Several factors govern whether a consumer can reach such a conclusion. Two we have already met. We know from Proposition 1 that the consumer struggles more when their processing capacity L is lower, and from Proposition 2 that they struggle more when they lack intermediate concepts. We now turn to two further factors: option similarity and choice complexity.

Option Similarity

An empirical literature finds that people struggle to choose when options are similar. When alternatives are close in attractiveness, people engage in costly deferral even though deferring is irrational (e.g. Dhar, 1997; Tversky and Shafir, 1992).³⁰

Our framework predicts this directly. Consider a consumer choosing between two products, and suppose neither product’s value is high enough for the product route to be feasible. If one product far outrates the other on some attribute, the consumer can still choose by the attribute route, extrapolating to “that product is weakly preferred” from its lead there. Now raise the inferior product’s ratings, keeping it inferior. Its rival’s lead narrows, the fit of the attribute-route extrapolation falls, and the route may close—leaving the consumer unable to choose. The options have improved, yet the choice is harder.

To illustrate, consider the following versions of Mr. Stitch’s problem. Panel (a) of Figure 4 shows one; panel (b) shows a second in which Suit 1 is unchanged but Suit 2 is more attractive—and so more similar to Suit 1.

Suppose Mr. Stitch’s extrapolation threshold is $\alpha = 3/4$. The product route is closed in both cases: to choose a suit outright its value would need to be 16 or more, and to reject

²⁹That decision-making falters when an agent cannot reach a conclusion about which option is best is supposed across disciplines. In Shafir et al. (1993)’s “reason-based choice,” people struggle when they cannot find a unifying rationale for one option over another; in cognitive-load theory (Sweller, 1988), the failure to integrate disparate information brings cognitive burden and “analysis paralysis” (Baumeister et al., 1998); Chang (2017) examines the difficulty of choosing without clear evaluative criteria; and the literature on conviction narratives (Johnson et al., 2023) argues that people decide confidently only once they can form an integrated understanding of their situation.

³⁰A rational agent might defer more readily as options’ overall attractiveness declines, but not as they grow closer in relative attractiveness. Tversky and Shafir (1992) find that relative attractiveness matters nonetheless.

	Price Rating	Quality Rating		Price Rating	Quality Rating
Suit 1	7	5	Suit 1	7	5
Suit 2	2	5	Suit 2	6	5

(a) (b)

Figure 4: Two choice problems for Mr. Stitch

one in favor of its rival its value would need to be 6 or less—but every suit here has a value in between. A choice can come only by the attribute route. In panel (a), Suit 1 leads on attribute 1 by five; attending to it, Mr. Stitch reaches “Suit 1 is weakly preferred” (fit 0.83) and chooses Suit 1. In panel (b), Suit 2’s improvement narrows that lead to one, the fit falls to 0.29, and the attribute route closes as well. Suit 2 has gotten better, yet Mr. Stitch can no longer choose.

Choice Complexity

People also find it harder to choose as a problem grows more complex. A classic illustration is the “jam study” of Iyengar and Lepper (2000), in which shoppers were more likely to buy from an assortment of six than from one of twenty-four—more options, fewer purchases.³¹

Complexity, in our framework, means more products or more attributes—a larger n or m . Either makes extrapolation harder, as the following proposition shows.

Proposition 5. *Consider the product-choice setting of Proposition 4. Sufficient complexity makes a choice impossible by the attribute route or the product route: adding attributes shuts down the attribute route, and adding products the product route. Specifically:*

³¹Subsequent meta-analyses find choice overload to be conditional rather than universal (Chernev et al., 2015; Scheibehenne et al., 2010), and its conditions are the model’s: overload is strongest when options are numerous and attributes many (fit falls as n and m grow, as we show below); when no option stands out (no route clears the threshold, as under “Option Similarity”); and when preferences are uncertain (a consumer who must first infer how attributes map to utility has further to travel; see Section 3.6). Conversely, overload should recede with expertise, which supplies the intermediate concepts that ease integration.

1. (More attributes.) *Let product i lead on the attended attribute y , and fix its advantages $\{\Delta_{ij}^y\}_{j \neq i} > 0$. Attending to attribute y , the fit of “product i is weakly preferred” tends to zero as the number of attributes m increases, and is strictly decreasing in m once m is large enough.*
2. (More products.) *Holding product i ’s value $u_i < mV$ fixed and attending to product i , the fit of “product i is weakly preferred” is strictly decreasing in the number of products n and tends to $-\infty$. Holding a rival product’s value u_j fixed and attending to this rival, the fit of “product i is weakly preferred” tends to zero as n increases.*

The intuition concerns dilution. A single attribute speaks to a smaller part of the choice as more attributes crowd in around it. Knowing that one product leads on price says less about which is best overall when there are ten attributes than when there are two. The same holds along the product route—knowing that one product is strong settles less of the comparison when there are many rivals it must beat than when there is only one. In each case a fixed piece of evidence is asked to support a conclusion about an ever-larger whole, and its fit erodes accordingly.

Mr. Stitch’s original problem illustrates both (Table 3). Along the attribute route, attending to price—on which Suit 1 leads by eight—the fit of “Suit 1 is weakly preferred” falls as attributes are added with that lead held fixed. Along the product route, attending to Suit 1—whose value of thirteen is the highest—the fit falls as suits are added with that value held fixed.

Attribute route: price in mind (two suits)		Product route: Suit 1 in mind (two attributes)	
# of attributes	Fit	# of suits	Fit
2	0.98	2	0.48
3	0.88	3	0.34
4	0.81	4	0.22
5	0.75	5	0.11

Table 3: Fit of “Suit 1 is weakly preferred” as attributes or suits are added

Dealing with difficult choices

What does an agent do when they struggle to choose? We do not take a firm position, but offer two possibilities. One is that they fall back on a default—in effect, declining to choose at all. This may help explain why defaults are so influential in complex settings such as retirement-plan enrollment (see Madrian and Shea, 2001; Thaler and Benartzi, 2004).

A second possibility is that the agent may gradually lower their fit threshold α —becoming more willing to extrapolate—until a choice becomes feasible.³²

This second possibility can be made precise in the product-choice setting, and it yields intuitive predictions. Suppose the consumer, facing a picture P , proceeds as follows. Starting from a threshold of one, they lower α until, for some product a , some extrapolation reaches the decision-relevant feature f_a —that is, until α reaches

$$\alpha^*(P) = \max_a \max_{\substack{D \subseteq K_0 \\ |D| \leq L}} \Psi(f_a; D).$$

They evaluate the f_a 's in a predetermined order, starting at $a = 1$ and concluding at $a = n$, and adopt the first feature attaining the maximum, choosing the corresponding product.³³

The stopping threshold $\alpha^*(P)$ measures how much time it takes the consumer to reach a decision: the further the threshold must fall, the longer they deliberate. Let $E(P)$ equal one if the product chosen is not optimal, and zero otherwise. Deliberation time and error rates can also be examined across problems of a given size, by averaging $\alpha^*(P)$ and $E(P)$ over all pictures in which n products are described by m attributes, every rating being equally likely. We write α^* and E for these averages too.

The following proposition considers what happens as the number of attributes grows, taking the simplest description—a single attribute—as the benchmark.

³²An agent can always reach a choice once the threshold is low enough.

³³The evaluation order serves as a tie-breaking rule: when several features tie for the highest fit, the consumer adopts the one they evaluate first. Drawing on nothing at all is permitted and has fit zero, so the threshold need never fall below zero: at zero, every feature can be reached, and the evaluation order decides.

Proposition 6. *Consider the product-choice setting with $n \geq 2$, $V \geq 2$, and a consumer with processing capacity $L \geq 1$ and no intermediate concepts.*

1. *If $L \geq n \times m$, then $\alpha^* = 1$ and $E = 0$: when the full problem fits in mind, deliberation is fast and the consumer never errs.*
2. *As the number of attributes m increases:*
 - (i) *Deliberation lengthens: $\alpha^* \rightarrow 0$, eventually falling strictly below its value at $m = 1$.*
 - (ii) *Mistakes increase: $E \rightarrow 1 - 1/n$ —the error rate of choosing randomly—eventually rising strictly above its value at $m = 1$.*

The message is simple. When problems are more complex—when m is greater—the consumer deliberates more and makes more mistakes.³⁴ Both patterns have empirical support. Decision times rise with the complexity of the choice (Payne et al., 1993; Wilcox, 1993), and subjects choosing from larger, more complex sets more often fail to find the best option (Caplin et al., 2011).³⁵

Extrapolating to save steps

One reason to lower the threshold α is that a conclusion is otherwise out of reach. Another is to save time. Even when a conclusion is reachable by deduction, it may be reachable only after many steps, and rather than work through them the agent may lower their threshold and extrapolate, reaching a conclusion in fewer. On this view, the agent extrapolates not only when a problem cannot be solved but when solving it is taking too long. This suggests a prediction. Section 2.4 measured a problem’s difficulty by the number of steps deduction requires. Problems that are difficult in this sense should be the ones on

³⁴An exactly parallel result for n is not available, for an instructive reason. As options multiply, some option carrying top ratings on L attributes appears with probability approaching one, and the consumer comes to select such an option. Deliberation then depends on the ratio L/m —the share of a product’s attributes the consumer can take in—with the threshold settling at L/m rather than falling to zero; and the error rate likewise settles below one, at a level governed by how much the L inspected attributes reveal about the rest.

³⁵Complexity in Caplin et al. (2011) varies along two dimensions: the number of options in the set and the complexity of each object. The first corresponds to growing n , the second to growing m . The failure rate of their subjects rises with both.

which people extrapolate most, and so the ones on which the consequences of extrapolation traced in this section—from selective attention to systematic bias—should be most visible.

3.4 Systematic Biases

A consumer who extrapolates can make mistakes. In the example we began with (Figure 3), Mr. Stitch chooses correctly (Suit 1) if he attends to price, but wrongly (Suit 2) if he attends to quality. Do such reasoning difficulties merely make choice noisier, or can they push the consumer toward systematically incorrect conclusions? The distinction is central to the behavioral tradition initiated by Kahneman, Tversky, and Thaler (Kahneman and Tversky, 1979; Thaler, 1980; Tversky and Kahneman, 1974). Random errors favor no one option in particular and tend to cancel in the aggregate, leaving the rational model approximately correct. By contrast a systematic bias reliably favors one suboptimal option, survives aggregation, and can move markets, distort prices, and generate persistent welfare losses.

Limited processing capacity produces the latter and the direction of the bias is determined endogenously by the structure of the problem. To see how, consider the version of Mr. Stitch’s problem in Figure 5, with two suits and five attributes; as before, Mr. Stitch has a processing capacity of two and no intermediate concepts. Suit 1 leads Suit 2 considerably on attribute 1 but trails it slightly on each of the other four, so that Suit 1’s value (28) falls just short of Suit 2’s (29). Suit 2 is the better buy.

$$P = \begin{array}{c} \text{Attr. 1} \quad \text{Attr. 2} \quad \text{Attr. 3} \quad \text{Attr. 4} \quad \text{Attr. 5} \\ \text{Suit 1} \left[\begin{array}{ccccc} 10 & 6 & 3 & 4 & 5 \end{array} \right] \\ \text{Suit 2} \left[\begin{array}{ccccc} 2 & 7 & 8 & 5 & 7 \end{array} \right] \end{array}$$

Figure 5: An Example of Systematic Bias

Under the noise story, Mr. Stitch would sometimes pick the wrong suit but would not

favor it. In our model he may choose the inferior Suit 1 every time. His highest fit for f_1 (“Suit 1 is weakly preferred”) comes from attending to attribute 1, the ratings 10 and 2, and equals 0.75; his highest fit for f_2 (“Suit 2 is weakly preferred”) comes from attending to attribute 3, the ratings 3 and 8, and equals only 0.55. So for any threshold between the two ($0.55 < \alpha < 0.75$), he chooses Suit 1 over Suit 2 systematically.

The source of this bias is the asymmetry in how the two suits’ advantages are arranged. Suit 1’s edge is concentrated on a single attribute and so it is easy to see, whereas Suit 2’s edges are spread thinly across four attributes and therefore easy to miss given Mr. Stitch’s limited processing capacity.

In the example, we placed no restriction on Mr. Stitch’s thoughts. The fits just reported are the highest attainable over *every* thought his capacity permits, including those where he draws on one suit’s rating on one attribute and the other suit’s rating on a *different* attribute. Comparisons of that kind—one suit’s price against the other’s quality—seem unnatural. Setting them aside, and restricting the consumer to the two routes of Proposition 4—attending to an attribute, or to a product—systematic bias admits a clean characterization.

Proposition 7. *Consider a choice between two products, and suppose product 2 has the higher value ($u_2 > u_1$). Suppose the consumer has processing capacity $L \geq 2$ and their thoughts draw on either both products’ ratings for a single attribute, or on a single product’s ratings in full—the routes of Proposition 4. Let $\Delta_1 \equiv \max_y \Delta_{12}^y$ and $\Delta_2 \equiv \max_y \Delta_{21}^y$ denote each product’s largest lead on any single attribute. By Proposition 4, each route’s fit depends only on a single statistic: write $F_m(\Delta)$ for the attribute-route fit at lead Δ , $G_m(u)$ for the product-route fit attending to a product of value u , and $H_m(u')$ for it attending to a rival of value u' . Set $G_m \equiv H_m \equiv -\infty$ when $L < m$ (i.e. when the product route is infeasible).*

Then there exist thresholds α at which the consumer can conclude “product 1 is weakly preferred” but not “product 2 is weakly preferred”—choosing the inferior product whenever they choose—if and only if

- (a) $\Delta_1 > \Delta_2$: *product 1 has the strictly larger maximal lead;*

(b) $G_m(u_2) < F_m(\Delta_1)$: product 2's value is not too high; and

(c) $H_m(u_1) < F_m(\Delta_1)$: product 1's value is not too low.

The set of such thresholds is exactly $(\max\{F_m(\Delta_2), G_m(u_2), H_m(u_1)\}, F_m(\Delta_1)]$.

Condition (a) is the concentration asymmetry of the example: the inferior product's best selling point is the more striking one. Conditions (b) and (c) say the values are middling—product 2 not so good that the consumer settles on it from its value alone, product 1 not so bad that they reject it out of hand. When $L < m$ (so that the product route is infeasible), Conditions (b) and (c) hold automatically, and only Condition (a) has bite.

Two other accounts of bias under complexity are complementary to ours. In the cognitive-uncertainty framework of Enke and Graeber (2023), complexity raises the noise in the agent's internal signal about each option's value, leading them to lean instead on a mental default. Where that default favors one option, a systematic bias toward it follows.³⁶ In our model the direction of bias arises more endogenously. Because the consumer reasons from a subset of the facts, and different subsets point different ways, the structure of the problem itself sets the direction—a conclusion is easier to reach when the evidence for it is concentrated in a few facts than when it is dispersed across many.

The second account is the salience-based framework of BGS, together with the closely related focusing model of Kőszegi and Szeidl (2013) and the relative-thinking model of Bushong et al. (2021), in which the agent's weight on an attribute depends on how it stands out within the choice set. There the bias runs through a posited weighting of attributes, whose direction the models dispute. In ours, salience plays no role in the bias, which turns instead on whether an option's advantages fit within the agent's capacity.³⁷

³⁶Formally, in Enke and Graeber (2023) the agent's perceived value of option i is $\hat{V}_i = \lambda(V_i + \epsilon_i) + (1 - \lambda)\bar{V}_i$, with V_i the true value, ϵ_i mean-zero noise, and $\bar{V}_i$ a default; the weight λ falls as cognitive uncertainty rises. The default $\bar{V}_i$ shapes the direction of bias.

³⁷Salience and limited processing can be hard to tell apart since both work by confining the agent to part of the picture. Salience does this through what captures the agent's attention. In our framework it occurs through what the agent can hold in mind at once. But the accounts can be distinguished. Bushong et al. (2021) is already distinct from our account by the direction of choice. Relative thinking underweights

3.5 Belief Inconsistency

In most models of belief an agent cannot believe inconsistent things. In the extrapolation version of our model they can, and the inconsistency may go unnoticed.³⁸ A range of economic behaviors seem to reflect inconsistencies of this kind. Households carry high-interest credit-card debt while holding low-yielding liquid savings (Gross and Souleles, 2002; Telyukova, 2013). Investors judge the market overpriced while still expecting it to rise (Shiller et al., 1996). Experiments on reasoning find this pattern directly. When shown sets of assertions that cannot all be true at once, subjects routinely and predictably judge them consistent (Johnson-Laird et al., 2000; Legrenzi et al., 2003).

The same can happen in the product-choice setting.

Proposition 8. *Define the feature*

$$f_{i \succ j} = \{Q \in \mathcal{P} : u(i, Q) > u(j, Q)\}$$

as the event that product i is strictly preferred to product j , and suppose consumers can conceptualize such features. A consumer may reach a knowledge state K_t containing $f_{i \succ j}$, $f_{j \succ k}$, and $f_{k \succ i}$: they simultaneously believe i is preferred to j , j to k , and k to i , which cannot all hold.

Note that the agent’s utility function is entirely standard, ranking the products completely and transitively in every picture; $f_{i \succ j}$ is a *belief* about which product would yield more utility. What can be inconsistent are the agent’s beliefs about which products they prefer.

To prove the proposition, consider an example. Suppose Mr. Stitch has processing capacity of two ($L = 2$) and faces a choice between three suits, each with three attributes

the wide-spread attribute on which Suit 1 leads, and so selects Suit 2—the correct choice—where our account selects Suit 1. BGS and Köszegi and Szeidl (2013) are the closer cases, since both overweight that attribute, as we effectively do, and predict the same biased choice. What most separates our account from these is that neither predicts *struggling to choose*. While a fixed weighting always leads to a conclusion, our consumer may reach none—deferring choice, or falling back on a default.

³⁸An exception in spirit is the fact-free learning of Aragones et al. (2005) in which agents may likewise fail to see implications of what they know.

rated in $\{1, 2, 3\}$:

$$P = \begin{array}{c} \text{Suit 1} \\ \text{Suit 2} \\ \text{Suit 3} \end{array} \begin{bmatrix} \text{Attr. 1} & \text{Attr. 2} & \text{Attr. 3} \\ 2 & 3 & 1 \\ 3 & 1 & 2 \\ 1 & 2 & 3 \end{bmatrix}$$

Figure 6: An example of inconsistent beliefs

Evaluating $f_{1 \succ 2}$ on attribute 2, where Suit 1 leads 3 to 1, gives a fit of about 0.77; by cyclic symmetry the same holds for $f_{2 \succ 3}$ on attribute 1 and $f_{3 \succ 1}$ on attribute 3. So if $\alpha \leq 0.77$, Mr. Stitch extrapolates to all three, and K_t contains all three beliefs at once, though they cannot all hold. Indeed, if he extrapolates only from attribute comparisons (drawing both suits' ratings on one attribute), then for $0.47 < \alpha \leq 0.77$ these are the only pairwise comparisons he can reach.

Whether an agent can recognize such inconsistencies, and how they respond, is taken up in Appendix A: an agent who reevaluates beliefs with $L = 2$ catches this three-belief cycle but not a four-belief one.

These cyclic beliefs drive choice. Mr. Stitch picks Suit 1 over 2, 2 over 3, and 3 over 1. Extrapolation thus offers an account of intransitive choice, a pattern documented since May (1954) and Tversky (1969) (see also Loomes et al., 1991), and one that differs from existing accounts in locating the inconsistency in the agent's beliefs at a single moment.^{39,40} It also predicts that agents who recognize an inconsistency will act to resolve it, and experiments bear this out: confronted with intransitivities in their own choices, subjects overwhelmingly disavow and revise them (MacCrimmon, 1968; Nielsen and Rehbeck,

³⁹Regenwetter et al. (2011) argue that many apparent cycles are consistent with stochastic but transitive preferences.

⁴⁰Existing accounts build cycles either into preference—regret theory (Loomes and Sugden, 1982), SSB utility (Fishburn, 1982), similarity-based procedures (Rubinstein, 1988)—or into context-dependent attention—salience (Bordalo et al., 2012), focusing (Kőszegi and Szeidl, 2013), relative thinking (Bushong et al., 2021), sequential shortlisting (Manzini and Mariotti, 2007). In neither is there an inconsistency in beliefs at any single moment: in the preference-based theories the cycle is the agent's preference, coherent by the theory's own lights; in the attention-based theories beliefs may shift from one occasion to the next, but the agent holds only one at a time.

2022)).⁴¹ Our account also predicts that greater processing capacity, more intermediate concepts (which expertise may supply), or a push toward deduction over extrapolation should all reduce inconsistency.

3.6 Filling in the Blanks

So far, extrapolation has helped the consumer reach conclusions that follow from K_0 but are hard to see by deduction. It has a second role. When K_0 does not fully specify the picture ($\bigcap_{f \in K_0} f \supset \{P\}$), deduction is silent about what K_0 leaves open; extrapolation lets the consumer make an educated guess.

To see this, suppose the consumer does not know, *ex ante*, how ratings translate into utility. The picture now carries an extra column recording the utility u_x each product delivers, alongside its ratings $v_{x1}, \dots, v_{xm}$. The consumer knows every rating, but only the utilities of products they have tried before; for the rest, u_x is a blank.

From the products they know, the consumer may extrapolate to a feature of the form “ $u_x = U(v_{x1}, \dots, v_{xm})$ for all x ”—a *rule* for how ratings translate into utility, additive or with attributes interacting. With the rule in hand, they can fill in the blanks: for an untried product, they apply the rule to its known ratings and work out u_x . K_0 never pinned u_x down; the consumer gets there by first inferring a rule and then building on it.

This kind of reasoning is ubiquitous. Shown the partial phrase “_OOD _OR_ING” on *Wheel of Fortune*, a contestant first works out the whole phrase—“GOOD MORNING”—and then uses it to fill in the G, M, and N, exactly as the consumer does.⁴² What this adds is the capacity to *predict*: to reach past what the starting facts contain and fill in what one was never told, as when we finish a half-heard sentence, recognize an object from a

⁴¹In MacCrimmon (1968), 36 of 38 subjects, shown their intransitive choices, described them as mistakes. In Nielsen and Rehbeck (2022), whose subjects faced conflicts between their choices and axioms they had endorsed, only 17 percent of transitivity violations were left unreconciled: subjects revised the offending choices 66 percent of the time and renounced the axiom only 5 percent of the time. (Revision was less lopsided for other axioms.)

⁴²These two directions correspond to a distinction familiar in cognitive psychology between *bottom-up* and *top-down* thinking (see Marr, 2010; Neisser, 2014): reasoning from pixels toward higher-level features, or from a higher-level feature toward the pixels.

glimpse, or judge how an unfamiliar situation will unfold.

What the agent predicts, though, depends on the concepts they hold. The contestant lands on “GOOD MORNING” rather than “WOOD FORMING” because only the former is a concept for them; a consumer who can conceptualize only additive rules will infer an additive rule, and predict accordingly, even if their true preferences are more tangled. Conceptualization thus does more than limit what the agent can deduce; it shapes what they predict.

4 Path-dependent Thinking

Look at the rabbit–duck illusion (Figure 1) and you will see a rabbit or a duck, but not both. And whichever you see first, you tend to keep seeing. What you perceive depends on where perception started—a *path dependence* psychologists call *hysteresis* (Leopold and Logothetis, 1999). The same path dependence appears in reasoning more generally as *primacy effects*. Information encountered early weighs more heavily than information encountered later. Asch (1946) found that subjects formed a more favorable impression of a person when a list of traits ran from positive to negative than when the same traits were read in reverse. A Bayesian would not expect order to matter—the same evidence yields the same posterior however it arrives—yet it clearly does.

So far we have made no assumption about which features are *salient* to the agent—which of the facts they hold come to mind when they think. Salience is plainly a crucial aspect of how people think and choose. It runs through the work of Kahneman and Tversky—the availability and anchoring heuristics, for instance (Tversky and Kahneman, 1974)—and is the foundation of the body of work by BGS et al., which places it at the center of a theory of choice. Once we make a salience assumption, path dependence follows. Where the agent’s reasoning begins shapes where it ends. We demonstrate this with an example, then generalize and draw out the implications for persuasion.

4.1 Mr. Judge

Suppose a voter, Mr. Judge, is deciding whether to vote for a politician described by m characteristics:

$$P = \begin{bmatrix} p_1 & p_2 & \cdots & p_m \end{bmatrix},$$

where $p_i \in \{\text{good}, \text{bad}\}$ and $m > 4$. The politician is a “good egg” if at most two characteristics are bad, and a “bad egg” if at most two are good—so he cannot be both. Mr. Judge prefers to vote for the politician if he is a good egg and for an alternative candidate if he is a bad egg, and is indifferent otherwise.

Mr. Judge has a processing capacity of two ($L = 2$), an extrapolation threshold $\alpha = 1/2$, and no intermediate concepts. He observes four of the politician’s characteristics—two good (p_1, p_2) and two bad (p_3, p_4)—so his initial knowledge K_0 consists of the pixel values for these four.

All Mr. Judge can do at time 1 is evaluate “good egg” or “bad egg” against two of the characteristics he has observed. Take $m = 7$. Evaluated against the two good characteristics, “good egg” has fit 0.53, which is above his fit threshold. By symmetry, “bad egg” has fit 0.53 against the two bad characteristics. He can thus extrapolate to either verdict, depending on which pair of characteristics he considers.

Does his verdict depend on which he considers first? As the model stands, no. Nothing stops him from focusing on the good characteristics at time 1 and reaching “good egg,” then focusing on the bad characteristics at time 2 and reaching “bad egg.” What is missing is salience. If “good egg,” once reached, stays in Mr. Judge’s mind, “bad egg”—which contradicts it—becomes unreachable, and the bad characteristics are brushed aside as the foibles of a good egg.

4.2 Salience and Path Dependence

To formalize this idea, we do three things. First we define the *support* of a conjecture as the pixels it concerns. We then define the *power* of a set of features on a region of the

picture—how much the features narrow down that region. Finally, we assume that when the agent evaluates a conjecture, features that are powerful on its support—particularly explanatory of what the conjecture concerns—tend to be drawn into their thinking.

Definition 3. Pixel p_{xy} is in the support of conjecture ϕ_t if, for some realization of the other pixels, knowing that ϕ_t applies to P narrows the values p_{xy} could take.

The conjecture “ $p_{11} = 10$ ” has only pixel p_{11} in its support, and “ $p_{11} + p_{12} = 10$ ” has p_{11} and p_{12} . The conjectures Mr. Judge is considering, “good egg” and “bad egg,” have the whole picture in their support.

Definition 4. Let a region R be a nonempty set of pixel locations. Let $\mathcal{R}$ be the set of subpictures on R : the classes of pictures that agree at every pixel in R . For a set of features F , let $\delta_R(F) \subseteq \mathcal{R}$ be the subpictures consistent with each feature in F —those $P_R \in \mathcal{R}$ with $P_R \cap f \neq \emptyset$ for each $f \in F$. The power of F on R is

$$\Gamma_R(F) \equiv 1 - \frac{\log |\delta_R(F)|}{\log |\mathcal{R}|}.$$

Power measures how much the features F narrow down the region R : it is zero when F rules out nothing on R and one when F pins R down completely.

Our salience assumption is that features that are powerful on the support of the conjecture command a place in the agent’s thinking.

Assumption 1. Suppose $f \in K_t \setminus \{\phi_t\}$. If the agent draws on a feature $g \in K_t \setminus \{\phi_t\}$ that is strictly less powerful than f on the support of ϕ_t , they must also draw on f . The same holds if they draw on features $g_1, \dots, g_n$ ($n > 1$) that taken together are weakly less powerful than f on the support of ϕ_t .

Under this assumption, Mr. Judge’s thinking is path dependent when $m = 7$. On the whole picture—the support of both “good egg” and “bad egg”—“good egg” has power 0.31, while two pixel values have power 0.29 and one has 0.14. Now suppose Mr. Judge

concludes “good egg” at time 1, drawing on the two good characteristics. Can he conclude “bad egg” at time 2? To evaluate it, he must draw on the bad characteristics, p_3 and p_4 . But either of these alone, or the two together, is less powerful on the support of “bad egg” than “good egg,” so by Assumption 1 he cannot draw on them without also drawing on “good egg.” Since “good egg” contradicts “bad egg,” the conjecture is rejected. Having concluded “good egg,” Mr. Judge cannot go on to “bad egg.” By symmetry, having concluded “bad egg” at time 1, he cannot go on to “good egg.” This proves, by example, the following proposition.

Proposition 9. *Under Assumption 1, the agent’s thinking may exhibit path dependence: for some (α, L, P, C, K_0) , there exist features f and f' and finite thought sequences (i) $\theta_0, \dots, \theta_t$ and (ii) $\theta'_0, \dots, \theta'_t$ such that under sequence (i), $f \in K_\tau$ and $f' \notin K_\tau$ for all $\tau > t$, and under sequence (ii), $f' \in K_\tau$ and $f \notin K_\tau$ for all $\tau > t$, whatever the continuation of the thought sequence.*

4.3 Implications for Persuasion

Path dependence gives a persuader channels beyond Bayesian persuasion (Kamenica and Gentzkow, 2011). The first is immediate: *order matters*. Because the framework predicts primacy, a persuader can manipulate the order in which information arrives—a politician displays his strengths before opponents can air his flaws, so that “good egg” takes hold and hardens against later attacks.

A persuader can also try to *anchor* the agent. Anchoring effects—the pull of an initial value on subsequent judgments—are well documented and remarkably robust, across purchase decisions, legal sentencing, real-estate appraisal, and negotiation (Furnham and Boo, 2011; Tversky and Kahneman, 1974). In our framework, an anchor is an initial conjecture, and path dependence makes the starting point decisive: if a persuader gets Mr. Judge to entertain “good egg” rather than “bad egg” at the outset, then once it takes hold it cannot be dislodged.

A third channel is to suggest a *narrative*—an interpretation of the evidence, together with a few facts that support it. A persuader might propose “the politician is a good egg,”

say, and point to the politician’s good characteristics. In terms of the framework, think of this—like anchoring—as planting an initial conjecture and, alongside it, some facts for the agent to draw on: the persuader proposes a whole thought, $\theta_t = (\phi_t, D_t)$. It is a form of “model persuasion” (Schwartzstein and Sunderam, 2021), though here the persuader supplies a feature rather than a complete model. Path dependence is why “controlling the narrative”—being first to set it—is treated as so valuable in politics. The opening frame shapes how later information is interpreted (Entman, 1993). A narrative can be supplied, however, only if the agent can conceptualize it, which in practice means it must be *simple*. As the political consultant Frank Luntz puts it, “The most memorable political language is rarely longer than a sentence.”⁴³ One corollary is that a simple falsehood can outcompete a complex truth—or, as Swift put it, “Falsehood flies, and the Truth comes limping after it.”

4.4 Resolving Inconsistencies by Deleting Features

So far, a feature the agent has accepted stays in their thinking for good. A natural extension lets the agent reevaluate a conclusion they have drawn and delete it if, on reevaluation, its fit falls below a threshold. One rationale for deletion is to resolve inconsistencies. As Section 3.5 showed, an agent who extrapolates can come to hold beliefs that cannot all be true. Appendix A develops the extension and finds that deletion does this only in part. An inconsistency among k features can be detected and eliminated if and only if $k \leq L + 1$, since recognizing that a feature contradicts the others requires holding those others in mind at once. Blatant contradictions like “good egg” and “bad egg” are caught, while the four-suit cycle of Section 3.5 escapes an agent with $L = 2$. Deletion has a further effect: it can make thinking cyclic. Mr. Judge, reevaluating “good egg” against the bad characteristics, may delete it, extrapolate to “bad egg,” and later reverse again. The framework thus predicts alternation between readings—a pattern documented both in perception and in judgment, where people alternate between opposing positions rather

⁴³Luntz (2007), p. 7.

than holding a stable view (Leopold and Logothetis, 1999; Rothman et al., 2017).

5 Discussion

We have taken the agent’s conceptualization C as given. Yet it is plainly endogenous. An expert chess player holds concepts, such as common board configurations, that a novice lacks. We close by sketching two ways the framework might endogenize concepts.

Concept selection. One approach is to suppose the agent can hold up to Λ *intermediate concepts*, and selects those that let them solve as many problems as possible.⁴⁴ In this sense, the agent selects concepts that are useful for solving the types of problems they face.⁴⁵ An example illustrates this. Ms. Lodge is deciding whether to rent a house, described by whether it is urban, close to cafes, has a garden, and has a garage. A garage is essential, but beyond that she wants the house to be either urban and close to cafes, or suburban with a garden. She has capacity $L = 2$ and may hold $\Lambda = 2$ intermediate concepts. Without intermediate concepts, she can recognize a house to pass on—for instance, a missing garage settles it—but cannot deduce that a house is worth renting, which requires holding three features at once. Two concepts bridge this gap: “good urban location” (urban and close to cafes) and “good suburban location” (suburban with a garden). These are the useful concepts for the problems she faces, and so the ones we might expect her to hold.⁴⁶

Codes. A feature, so far, is either a concept or it is not. A richer approach lets the agent *encode* features more or less efficiently, with efficiency governing how readily a feature

⁴⁴Intermediate concepts are those beyond the pixel values and decision-relevant features, which are always available. Formally, the agent faces a class of problems, each a pair (P, K_0) . They *solve* a problem if they can reach a decision-relevant feature from K_0 within their processing capacity, and C is *optimal* if none solves more problems. A concept comes paired with its complement at no extra cost.

⁴⁵The agent need not literally optimize. For instance, C might evolve toward useful concepts; the optimum then indicates where this evolution is headed.

⁴⁶The analysis applies generally. A choice problem induces a function from pixels to optimal action, and useful concepts can be found by decomposing it. For Ms. Lodge, with $p_1, \dots, p_4$ the four features, renting is optimal when $F = p_4 \wedge ((p_1 \wedge p_2) \vee (\neg p_1 \wedge p_3))$, which decomposes as $p_4 \wedge (G_1 \vee G_2)$ with $G_1 = p_1 \wedge p_2$ and $G_2 = \neg p_1 \wedge p_3$ the two location concepts. Decomposing such functions into simpler pieces is the central concern of circuit design (Sasao, 1999; Wegener, 1987).

comes to mind.⁴⁷ Concept selection becomes code selection. Short strings are scarce, so which features get short codewords is the coding counterpart of which concepts are held—the concept limit now built into the code rather than imposed by fiat. And when the world has regularities, a *compositional* code—codewords assembled from reusable parts, such as a shape, a position, and a size—represents a large world with a small vocabulary, and leads the agent to reason with objects and operators rather than whole features (see Bengio et al., 2013).

Codes also say what it means to have a *model*. Two strings for the same feature are two models of it, differing in length, and the minimum description length principle—prefer the most compact (Rissanen, 1978)—selects among them. What makes a model compact is usually an assumption about how the data were generated. The Necker cube is far shorter to describe as a cube projected onto the plane than line by line, and cheaper precisely because that description posits structure the figure does not require. Causal models compress in the same way, treating some variables as caused by others when the correlations alone do not say (Pearl, 2009). The attraction of a generative model is, in the end, due to economy of thought.

6 Conclusion

Standard economic theory assumes that people draw all the conclusions their information supports, and draw them immediately. Yet a person with every fact about their own finances in front of them—their income and the tax rules—may still fail to see what those facts jointly imply. Faced with such problems, people tend to extrapolate—to settle on a conclusion from part of the evidence—and the mistakes they make are not random. They depend on the aspects of the problem to which they attend.

Economics has largely treated information processing as a black box. Given the facts, the implications follow. This paper offers a way to open that box. Our central assumption

⁴⁷Give the agent a code $C : B \rightarrow \mathcal{F}$ from binary strings to features. Drawing on facts $f_1, \dots, f_n$ then means drawing on encodings $b_1, \dots, b_n$, and the demand on processing capacity is their total length: $\sum_i \ell(b_i) \leq L$.

is that agents have limited *processing capacity*. They can hold only a few facts in mind and combine them at once. This constraint has predictive implications. More complex decisions take longer to make and lead to more mistakes—and, because the agent must attend selectively, the mistakes are systematically and predictably biased. The order in which information arrive matters. Early information and well-timed narratives move the conclusions people reach, and it pays to make one’s case first. The same constraint accounts for choice overload, satisficing, and beliefs that are inconsistent without the agent noticing.

Our framework leaves open a number of issues. One question is how the contents of an agent’s thoughts come to be determined—which facts they attend to, and which conjectures they entertain first. We have remained largely agnostic, imposing only the assumption of Section 4.2 that powerful features command a place in the agent’s thinking. Theories of attention are the natural place to look for more complete answers. On the goal-driven side, an agent’s grasp of the problem can steer their thoughts toward what is useful for solving it—a form of *metacognition* (Flavell, 1979) that could be modeled within the framework, under the same capacity constraint. On the stimulus-driven side, our salience assumption is one formalization among several. The salience theory of BGS et al., in which prominence, contrast, and surprise govern what captures attention, could equally govern which facts are drawn into a thought.

A second question concerns the *speed* of reasoning. Reasoning unfolds in real time, and a conclusion reached too late is of little use. What an agent can work out *in time* to act may matter as much as what they can work out at all. The difficulty measure of Section 2.4 and the stopping threshold of Section 3 are first steps toward modeling this—and they suggest that agents extrapolate not only when a conclusion is out of reach, but when deduction would take too long.

Appendix

A Deleting Features

Here we consider the idea, remarked on in Section 4.4, that the agent may delete features as well as add them. As Section 3.5 showed, an agent who extrapolates can come to hold mutually inconsistent beliefs—formally, there may be no picture consistent with every feature they have accepted: $\bigcap_{f \in K_t} f = \emptyset$. Mr. Stitch, for instance, can simultaneously believe Suit 1 is better than Suit 2, Suit 2 better than Suit 3, and Suit 3 better than Suit 1. It is natural to suppose the agent would try to resolve such inconsistencies, and one way they can do so is by reevaluating and deleting features. We show that this resolves inconsistencies only in part. Deletion also has a further consequence: it can make the agent’s thinking cyclic as well as path dependent.

A.1 Eliminating Inconsistencies

Suppose that at each time t the agent draws on facts $D_t \subseteq K_t \setminus \{\phi_t\}$ and either evaluates a new conjecture ($\phi_t \notin K_t$) or reevaluates a conclusion they have already drawn ($\phi_t \in K_t \setminus K_0$).⁴⁸ A new conjecture is accepted—added to K_t —if its fit is at least α ($\Psi(\phi_t; D_t) \geq \alpha$); an existing conclusion is discarded—deleted from K_t —if its fit is at most β ($\Psi(\phi_t; D_t) \leq \beta$), where $\alpha > \max\{\beta, 0\}$.⁴⁹

Does deletion, so defined, rid the agent of their inconsistencies? Not always: whether an inconsistency can be eliminated depends on how many features it involves. Call a set F of k features an *inconsistency of order k* if the features are jointly inconsistent ($\bigcap_{f \in F} f = \emptyset$) while every proper subset is consistent (for each $f' \in F$, $\bigcap_{f \in F \setminus \{f'\}} f \neq \emptyset$). “Good egg” and “bad egg” form an inconsistency of order two; “Suit 1 is better than Suit 2,” “Suit 2 is better than Suit 3,” and “Suit 3 is better than Suit 1” form one of order three—any two are consistent, but not all three.

Proposition 10. *If an inconsistency of order $k \leq L + 1$ exists at time t , there is a thought θ_t that eliminates it at time $t + 1$. However, for any $L \geq 1$ and any $k > L + 1$, for some $(\alpha, \beta, P, C, K_0)$ and finite thought sequence $\theta_0, \dots, \theta_{t-1}$, an inconsistency of order k exists at time t that no subsequent thought sequence can eliminate.*

Intuitively, the agent can detect and delete an inconsistency only when it is small enough to fit within their processing capacity. To recognize a feature f as one to discard, they must draw on the other features that contradict it; an order- k inconsistency requires holding the other $k - 1$ alongside f , which their capacity permits only when $k - 1 \leq L$. A capacity of one suffices to resolve order-two inconsistencies, so we should not expect agents to harbor contradictions as blatant as “good egg” and “bad egg.” Higher-order inconsistencies, however, can escape detection and persist: with $L = 2$, Mr. Stitch cannot detect and eliminate the order-four inconsistency “Suit 1 is better than 2,” “2 is better than 3,” “3 is better than 4,” and “4 is better than 1.”

⁴⁸Think of the agent’s initial knowledge K_0 as axiomatic and not subject to reevaluation.

⁴⁹The deletion threshold β may be positive or negative.

A.2 Cycling

Once features can be deleted, an agent’s thinking may cycle. Consider Mr. Judge with $m = 7$, and suppose he extrapolates to “good egg” at time 1 by focusing on the politician’s two good characteristics. At time 2 he reevaluates “good egg,” now focusing on the two bad characteristics; its fit against them is -1.33 , so if $\beta > -1.33$ he deletes “good egg.” With it gone, at time 3 he can extrapolate to “bad egg” from the bad characteristics—and by symmetry can later delete “bad egg” and return to “good egg.” Whether this happens turns on β . When $\beta < -1.33$, each verdict is stable once reached, and the model delivers path dependence as in Section 4.2; when $\beta > -1.33$, neither is stable, and Mr. Judge oscillates between them, never settling. This proves, by example, the following proposition.

Proposition 11. *Under Assumption 1, cycling is possible: for some $(\alpha, \beta, L, P, C, K_0)$, there exist features f and f' and a thought sequence $\theta_0, \theta_1, \dots$ such that $f \in K_{t_1}$ and $f' \notin K_{t_1}$, $f' \in K_{t_2}$ and $f \notin K_{t_2}$, and $f \in K_{t_3}$ and $f' \notin K_{t_3}$, for some $t_1 < t_2 < t_3$.*

What does the agent choose while their thinking cycles? A natural answer is that they act on whichever verdict they hold at the moment of choosing—taking action a when “action a is optimal” is in mind, and reversing later if they can.

B Proofs

Propositions 8, 9 and 11 are proved by example in Sections 3.5 and 4.2 and in Appendix A, respectively. Throughout, the *draw* of a thought is the set D_t of facts on which it draws.

Proof of Proposition 1. Part 1. (i) Take $F^* \subseteq K_0$ specifying P with $|F^*| = s(K_0) \leq L$. Every $\phi \in C(P)$ is deducible from F^* in one thought, as $\bigcap_{f \in F^*} f = \{P\} \subseteq \phi$. *(ii)* With $L = 1$, a draw is either empty (its intersection is $\mathcal{P}$, from which no feature is deducible) or a single fact g , from which ϕ_t is deducible only if $g \subseteq \phi_t$. By induction, every $f \in K_t$ contains some $g_0 \in K_0$: it is derivative.

Part 2. Fix $L > 1$ and $N = L(L + 1) + 1$. Let P be the all-zero $1 \times N$ picture over $\{0, 1\}$ and $K_0 = F_{\text{pixels}}(P) = \{p_1, \dots, p_N\}$, where $p_i = \{Q : q_i = 0\}$, so $s(K_0) = N > L$. Let $C = F_{\text{pixels}} \cup \{h, g, f_{\text{full}}\}$ with $h = p_1 \cap \dots \cap p_L$, $g = h \cap p_{L+1}$ and $f_{\text{full}} = \{P\}$. *(i)* $h \in R_1 \setminus K_0$; but $f_{\text{full}} \notin R$, as any L known facts other than f_{full} fix at most $L(L + 1) < N$ pixels. *(ii)* $g \notin R_1$, as L pixel values leave one of $q_1, \dots, q_{L+1}$ free; but h, p_{L+1} give $g \in R_2$. $\square$

Proof of Proposition 2. Thoughts feasible under C are feasible under $C' \supset C$, which gives both weak relations. For strictness, let $L = 3$, let P be the all-zero 1×5 picture over $\{0, 1\}$ and $K_0 = F_{\text{pixels}}(P)$. Write f_S for the feature that the pixels in S are zero (f_{12} for $f_{\{1,2\}}$, etc.), and let $C_1 = F_{\text{pixels}} \cup \{f_{12}, f_{12345}\}$, $C_2 = C_1 \cup \{f_{34}\}$ and $C_3 = C_2 \cup \{f_{123}\}$. Three known facts other than $f_{12345} = \{P\}$ fix at most four pixels unless they include f_{123} or both f_{12} and f_{34} . So f_{12345} is unreachable under C_1 , takes three steps under C_2 (f_{12} , f_{34} , then f_{12345}) and two under C_3 (f_{123} , then f_{12345}): the inclusion in Part 2 is strict for $C_1 \subset C_2$, the inequality in Part 1 for $C_2 \subset C_3$.

Part 3. Let $g_1, \dots, g_n \in K_0$ specify P . Each $h_i = g_1 \cap \dots \cap g_i$ is a feature applying to P , hence a concept; $h_1 = g_1$ is known and, as $L \geq 2$, each further h_i is deduced from h_{i-1} and g_i . From $h_n = \{P\}$ every $f \in C(P)$ follows. $\square$

Proof of Proposition 3. Lower bound. Define $s(K, f)$ for any knowledge set K as for K_0 . A thought lowers $s(K_t, f)$ by at most $L - 1$: in a smallest subset of K_{t+1} whose intersection is inside f , the new fact ϕ_t , if present, can be replaced by the at most L facts $D_t \subseteq K_t$ it was deduced from. So if n thoughts reach f , then $s(K_n, f) \leq 1$ gives $s(K_0, f) \leq 1 + n(L - 1)$, i.e. $n \geq \lceil (s(K_0, f) - 1) / (L - 1) \rceil$; and $n \geq 1$ as $f \notin K_0$.

Upper bound. Let $g_1, \dots, g_c \in K_0$ attain $c = s(K_0, f)$ and put $h_i = g_1 \cap \dots \cap g_i$. For $i < c$ this is a feature (it lies inside $g_1 \neq \mathcal{P}$ and, as c is minimal, not inside f , so it is nonempty), hence a concept as $C = \mathcal{F}$. From $h_1 = g_1$, each thought deduces h_{i+L-1} from h_i and the next $L - 1$ premises $g_{i+1}, \dots, g_{i+L-1}$; once at most $L - 1$ premises are left, a last thought deduces f from h_i and them, as together they intersect in $h_c \subseteq f$. The number of thoughts is the least $n \geq 1$ with $1 + n(L - 1) \geq c$, namely $\max\{1, \lceil (c - 1) / (L - 1) \rceil\}$.

Difficulty. An f_a that does not follow from K_0 has $s(K_0, f_a) = \infty$ and is never reached (lower bound). By nontriviality the others are outside K_0 , and by well-posedness there is at least one; their step count is nondecreasing in $s(K_0, f_a)$, so least at s^* . $\square$

Propositions 4–7 and 10: Preliminaries. Probabilities are under the uniform distribution over pictures, so in product choice ratings are i.i.d. uniform on $\{1, \dots, V\}$. Then $\psi(S) = -\log \Pr(\delta(S))$ and, for every feature f and consistent draw D_t ($\delta(D_t) \neq \emptyset$),

$$\Psi(f; D_t) = 1 - \frac{-\log \Pr(f \mid D_t)}{-\log \Pr(f)}, \quad (1)$$

where $\Pr(f \mid D_t)$ conditions on $\delta(D_t)$ and $-\log 0 = \infty$; so $\Psi(f; D_t) \leq 0$ iff $\Pr(f \mid D_t) \leq \Pr(f)$, and $\Psi(f; D_t) = -\infty$ iff $\Pr(f \mid D_t) = 0$. Let T be the value of a random product, H its c.d.f., and $w_n := \mathbb{E}[H(T)^{n-1}] = \Pr(f_i)$. Let $R_0, \dots, R_{n-1}$ be i.i.d. sums of $m - 1$ ratings and, for an integer vector $\Delta = (\Delta_1, \dots, \Delta_{n-1})$, $A_{m-1}(\Delta) := \Pr(R_j - R_0 \leq \Delta_j \text{ for all } j)$; define $A_m(\Delta)$ likewise from sums of m ratings, so $A_m(\mathbf{0}) = w_n$. Throughout, $n, V \geq 2$ by non-degeneracy, $m \geq 2$ on the attribute route (as $n \leq L < nm$), and a sum of k ratings has full support on the integers in $[k, kV]$.

Proof of Proposition 4. Part 1. Given the attended ratings, the totals R_0 of product i and R_j of each rival j over the unattended attributes are i.i.d. sums of $m - 1$ ratings, and f_i holds iff $R_j - R_0 \leq \Delta_{ij}^y$ for every j . So $\Pr(f_i \mid D_t) = A_{m-1}(\Delta)$ with $\Delta = (\Delta_{ij}^y)_{j \neq i}$, $\Pr(f_i) = A_m(\mathbf{0})$, and by (1) the fit depends on D_t only through Δ :

$$\Psi(f_i; D_t) = 1 - \frac{-\log A_{m-1}(\Delta)}{-\log A_m(\mathbf{0})}. \quad (2)$$

The fit is strictly increasing in each advantage: if the attended ratings give Δ' and $\Delta \leq \Delta'$, $\Delta \neq \Delta'$, is an integer vector, then $A_{m-1}(\Delta') - A_{m-1}(\Delta) \geq \Pr(R_j - R_0 = \Delta'_j \text{ for all } j)$, the probability that all n products then tie in value. This is positive by full support: with $\bar{v}$ the

largest attended rating, the totals $R_0 = m - 1 + \bar{v} - v_{iy}$ and $R_j = m - 1 + \bar{v} - v_{jy}$ produce the tie and lie in $[m - 1, (m - 1)V]$, as $\bar{v} - v_{xy} \leq V - 1 \leq (m - 1)(V - 1)$ when $m \geq 2$.

Part 2. (i) Given product i 's ratings, f_i holds iff each of the $n - 1$ independent rivals has value at most u_i ; so $\Pr(f_i | D_t) = H(u_i)^{n-1}$ and, by (1),

$$\Psi(f_i; D_t) = 1 - \frac{(n - 1)[- \log H(u_i)]}{- \log w_n}, \quad (3)$$

which depends on the ratings only through u_i and strictly increases in it, as H does by full support. *(ii)* Given rival j 's ratings, f_i holds iff product i 's value T_0 is at least u_j and the other $n - 2$ rivals' values are at most T_0 ; so $\Pr(f_i | D_t) = \mathbb{E}[\mathbf{1}\{T_0 \geq u_j\} H(T_0)^{n-2}] =: J(u_j)$ and the fit is $1 - [- \log J(u_j)] / [- \log w_n]$; by full support, both strictly decrease in u_j . $\square$

Propositions 5–7: Preliminaries. Let $q^{(k)}$ and $H^{(k)}$ be the p.m.f. and c.d.f. of a sum of k ratings (so $H = H^{(m)}$, and each R_j has c.d.f. $H^{(m-1)}$), and $\bar{H}^{(k)} = H^{(k)} - \frac{1}{2}q^{(k)}$ the midpoint c.d.f.; let $\mu = (V + 1)/2$ and $\sigma^2 = (V^2 - 1)/12$ be the mean and variance of a rating, $h_k = 1/(\sigma\sqrt{k})$ the standardized lattice mesh, and $I_{r,s} = \int \phi^r \Phi^s dz > 0$ ($r \geq 1, s \geq 0$) for the standard normal density ϕ and c.d.f. Φ . A rating has span one, so its characteristic function satisfies $|\chi(t)| \leq e^{-cvt^2}$ on $[-\pi, \pi]$, and Fourier inversion gives the *anticoncentration* bound $\sup_x q^{(k)}(x) \leq C_V/\sqrt{k}$; by convolution, it also holds for the difference W_k of two independent such sums.

Proof of Proposition 5. Part 1 (more attributes). Fix $\Delta = (\Delta_j)_{j=1}^{n-1}$ with $\Delta_j = \Delta_{ij}^y > 0$, and write Ψ_m for the fit (2). By anticoncentration, W_m and W_{m-1} take each value with probability $O(m^{-1/2})$. The number of weakly best products has mean $nA_m(\mathbf{0})$, is at least one, and exceeds one only if two values tie; so $A_m(\mathbf{0}) = 1/n + O(m^{-1/2})$. Likewise $A_{m-1}(\Delta) = A_{m-1}(\mathbf{0}) + O(m^{-1/2}) = 1/n + O(m^{-1/2})$, as $0 < R_j - R_0 \leq \Delta_j$ has probability $O(m^{-1/2})$; so $\Psi_m \rightarrow 0$. For the strict decrease, expand both probabilities to third order in h_m . Conditioning on R_0 gives $A_{m-1}(\Delta) = \sum_x q^{(m-1)}(x) \prod_j H^{(m-1)}(x + \Delta_j)$, and conditioning on T gives $A_m(\mathbf{0}) = \sum_x q^{(m)}(x) H^{(m)}(x)^{n-1}$. Write $z = (x - k\mu)/(\sigma\sqrt{k})$. The lattice Edgeworth expansion gives $q^{(k)}(x) = h_k[\phi(z) + h_k^2 p(z)] + O(h_k^5)$ uniformly in x , where $p(z)$ is a fixed polynomial times $\phi(z)$, with no odd powers of h_k in the brackets because a rating is symmetric about its mean (Petrov, 1975, Ch. VII). Summing over $|z| \leq \log k$ by the Euler–Maclaurin formula (Hoeffding's inequality bounds the discarded tails), the half-atom $\frac{1}{2}q^{(k)}(x)$ subtracted in $\bar{H}^{(k)}$ cancels the formula's order- h_k correction at the endpoint x , and $\bar{H}^{(k)}(x) = \Phi(z) + h_k^2[\int_{-\infty}^z p + \phi'(z)/12] + O(h_k^4 \log k)$. Inserting $H^{(k)} = \bar{H}^{(k)} + \frac{1}{2}q^{(k)}$ and expanding the products turns each probability into a combination of lattice sums; by Taylor expansion of each shifted factor about z and Poisson summation,

$$\sum_x \prod_{i=1}^r q^{(k)}(x + d_i) \prod_{l=1}^s \bar{H}^{(k)}(x + e_l) = h_k^{r-1} [I_{r,s} + \gamma_1 h_k + \gamma_2 h_k^2 + \gamma_3 h_k^3 + O(h_k^4 \log^2 k)]$$

for integer shifts $d_1 = 0, d_2, \dots, d_r$ and $e_1, \dots, e_s$, where the γ_ℓ are Gaussian integrals depending on r, s, V and the shifts, and $\gamma_1 = I_{r+1,s-1} \sum_l e_l$ when every $d_i = 0$ and $s \geq 1$. The term with no factor $\frac{1}{2}q^{(k)}$ has $r = 1$: its leading coefficient is $I_{1,n-1} = 1/n$, and its γ_1 is $I_{2,n-2} \sum_j \Delta_j$ for $A_{m-1}(\Delta)$, whose c.d.f. factors are shifted by the Δ_j , and zero for $A_m(\mathbf{0})$. The $n-1$ terms with one factor $\frac{1}{2}q^{(k)}$ have $r = 2$ and together contribute $\frac{n-1}{2} I_{2,n-2} h_k$ to either probability; all other contributions are $O(h_k^2)$. Using $h_{m-1} = h_m + \frac{\sigma^2}{2} h_m^3 + O(h_m^5)$, write $\xi_m := nA_{m-1}(\Delta) - 1 = \sum_{\ell=1}^3 a_\ell h_m^\ell + o(h_m^3)$ and $\eta_m := nA_m(\mathbf{0}) - 1 = \sum_{\ell=1}^3 b_\ell h_m^\ell + o(h_m^3)$, where $a_1 - b_1 = n I_{2,n-2} \sum_j \Delta_j > 0$. Expanding $\Psi_m = [\log(1 + \xi_m) - \log(1 + \eta_m)] / [\log n - \log(1 + \eta_m)]$ gives $\Psi_m = c_1 m^{-1/2} + c_2 m^{-1} + c_3 m^{-3/2} + o(m^{-3/2})$ with $c_1 = (a_1 - b_1) / (\sigma \log n) > 0$. Since $m^{-1/2} - (m+1)^{-1/2} = \frac{1}{2} m^{-3/2} + O(m^{-5/2})$, the other two terms change by $O(m^{-2})$ and both remainders are $o(m^{-3/2})$, $\Psi_m - \Psi_{m+1} = \frac{1}{2} c_1 m^{-3/2} + o(m^{-3/2}) > 0$ for all large m .

Part 2 (more products). Let $X = H(T) \in (0, 1]$, so that $w_n = \mathbb{E}[X^{n-1}]$; $X = 1$ exactly on $\{T = mV\}$, an event of probability $V^{-m} \in (0, 1)$. Attending to product i , where $u_i < mV$ so $-\log H(u_i) > 0$, (3) gives the fit $\Psi_n^{\text{own}} = 1 - [-\log H(u_i)] / [-\log \|X\|_{n-1}]$, where $\|X\|_k = \mathbb{E}[X^k]^{1/k}$. By Lyapunov's inequality $\|X\|_k$ is strictly increasing in k (X is not a.s. constant), and $V^{-m/k} \leq \|X\|_k < 1$; so $-\log \|X\|_{n-1}$ strictly decreases to zero, and Ψ_n^{own} strictly decreases to $-\infty$. Attending to a rival j , the proof of Part 2(ii) of Proposition 4 gives the fit $\Psi_n^{\text{rival}} = 1 - [-\log J(u_j)] / [-\log w_n]$ with $J(u_j) = \mathbb{E}[\mathbf{1}\{T_0 \geq u_j\} H(T_0)^{n-2}]$; as $n \rightarrow \infty$, $H(T_0)^n \rightarrow \mathbf{1}\{T_0 = mV\}$, so $J(u_j)$ and w_n tend to V^{-m} and $\Psi_n^{\text{rival}} \rightarrow 0$. $\square$

Proof of Proposition 6. Part 1. If $L \geq nm$, the draw of all nm ratings gives each true f_a fit one, so $\alpha^*(P) = 1$; no false f_a attains this, as fit one requires deducibility from ratings true of P , so $E(P) = 0$.

Part 2. Write $\alpha^*(m)$ and $E(m)$ for the averages over pictures with m attributes. Let S_a be product a 's rating total and $\pi_m = \Pr(f_a) = A_m(\mathbf{0})$, the same for all a by exchangeability. A draw $D \subseteq K_0$ of at most L ratings of P pins those ratings; given D , the unpinned ratings are i.i.d. uniform. Let $\rho_a(P) = \max_D \Pr(f_a | D)$. By (1), $\Psi(f_a; D) = 1 - \log \Pr(f_a | D) / \log \pi_m$ is strictly increasing in $\Pr(f_a | D)$, with the same π_m for every product, so the consumer chooses the first maximizer of ρ_a .

We use three facts. First, $\pi_m = 1/n + O(m^{-1/2})$, by the proof of Proposition 5. Second, *stability under pins*: $|\Pr(f_a | D) - \pi_m| \leq c/\sqrt{m}$ for some constant c , uniformly in the picture, product and draw. Indeed, let Q be a uniformly random picture and Q^D be Q with the ratings that D pins overwritten by their pinned values, so that Q^D has the conditional distribution given D . Overwriting moves each difference $S_a - S_x$ by at most $L(V-1)$, so f_a holds in one of Q, Q^D but not the other only if some $S_a - S_x$ is within $L(V-1)$ of zero in Q , which has probability $O(m^{-1/2})$ by anticoncentration and a union bound. Third, *saturation*: call a picture saturated if every row contains every rating value at least L times; by the law of large numbers, the probability of this tends to one. Given a draw, $\Pr(f_a | D)$ depends only on a and on the multiset of values pinned in each row, and by symmetry it is unchanged if rows a and b exchange their multisets and f_b replaces f_a . At a saturated picture each row contains any multiset of at most L values, so some draw

pins the exchanged multisets; hence $\rho_b \geq \rho_a$ for all a and b : all ρ_a are equal, and product 1 is chosen.

(i) By the first two facts, $\Pr(f_a \mid D)$ and π_m tend to $1/n \in (0, 1)$ uniformly, so every fit $\Psi(f_a; D)$ tends to zero uniformly, and $\alpha^*(m) \rightarrow 0$. At $m = 1$, pinning a rating equal to V certifies its product, and the empty draw has fit zero; so $\alpha^*(1) \geq 1 - (1 - 1/V)^n > 0$.

(ii) At a saturated picture $E(P) = \mathbf{1}\{\text{product 1 not optimal}\}$, an indicator with mean $1 - \pi_m$ over all pictures; so $|E(m) - (1 - \pi_m)|$ is at most the probability of non-saturation, and $E(m) \rightarrow 1 - 1/n$. At $m = 1$, let a 's rating be at least b 's, and let P' swap these two ratings, so that $\rho_b(P) = \rho_a(P')$ by exchangeability. The same products' ratings, drawn in P and in P' , pin the same values, except that a 's rating, if pinned, is weakly higher in P and b 's weakly lower; for any values of the unpinned ratings, f_a then holds in the first case whenever it holds in the second, so $\rho_a(P) \geq \rho_a(P') = \rho_b(P)$. Thus product 1 is chosen whenever optimal, and $E(1) \leq 1 - \pi_1 < 1 - 1/n$, as the maximum can be tied. $\square$

Proof of Proposition 7. By Proposition 4, F_m and G_m are strictly increasing and H_m is strictly decreasing (or $G_m \equiv H_m \equiv -\infty$, when $L < m$). Product i 's best attribute is thus one on which its lead is largest, and the consumer can conclude “product i is weakly preferred” iff $\alpha \leq B_i := \max\{F_m(\Delta_i), G_m(u_i), H_m(u_{-i})\}$, where u_{-i} is the other product's value; the thresholds sought are those with $B_2 < \alpha \leq B_1$. As $u_1 < u_2$, $G_m(u_1) \leq G_m(u_2) \leq B_2$ and $H_m(u_2) \leq H_m(u_1) \leq B_2$, so for $\alpha > B_2$, $\alpha \leq B_1$ iff $\alpha \leq F_m(\Delta_1)$: the thresholds form the interval $(B_2, F_m(\Delta_1)]$ of the statement. It is nonempty iff $F_m(\Delta_1)$ exceeds $G_m(u_2)$, $H_m(u_1)$ and $F_m(\Delta_2)$, which are Conditions (b), (c) and, as F_m is strictly increasing, (a). It lies in the model's range $(0, 1]$: no fit exceeds one, and $F_m(\Delta_2) > 0$ by (2), because $A_{m-1}(\Delta_2) > A_{m-1}(0) \geq A_m(0)$. The strict inequality holds by full support, as $u_2 > u_1$ forces $\Delta_2 \geq 1$; for the weak one, two independent sums of k ratings tie with probability $\|q^{(k)}\|_2^2$, so $A_k(0) = \frac{1}{2}(1 + \|q^{(k)}\|_2^2)$ by symmetry, which is nonincreasing in k because convolution with a probability distribution contracts the ℓ^2 norm. $\square$

Proof of Proposition 10. Let $F \subseteq K_t$ be an inconsistency of order $k \leq L + 1$. Every feature in K_0 applies to P , so some $f \in F$ lies outside K_0 ; reevaluated against the other $k - 1 \leq L$ members of F , which are consistent but contradict it, f has fit $-\infty \leq \beta$ and is deleted.

Now let $k > L + 1$. Let P have $k - 1$ pixels, each equal to a , over $\Omega = \{a, b, c\}$; let $K_0 = F_{\text{pixels}}(P) = \{f_1, \dots, f_{k-1}\}$, where f_i is “pixel i equals a ”; and let the only non-pixel concept be f_0 , “exactly one pixel differs from a .” Then $F = K_0 \cup \{f_0\}$ is an inconsistency of order k . A draw of $\ell \leq L \leq k - 2$ true pixel values leaves $x \geq 1$ pixels unfixed, and f_0 holds in $2x$ of the 3^x consistent pictures, a share that strictly decreases in x ; so by (1) the fit of f_0 is zero if $\ell = 0$ and positive otherwise. Let $\beta < 0 < \alpha$, with α at most the fit of f_0 with $D_0 = \{f_1, \dots, f_L\}$, so that the thought (f_0, D_0) adds f_0 : $K_1 = F$.

No later thought changes K_1 . The only other concepts are false pixel values h (“pixel i equals ω ,” $\omega \neq a$). Given a draw $D \subseteq K_1$ (consistent, as $|D| < k$), $\Pr(h \mid D)$ is 0 if $f_i \in D$ and $\Pr(h) = 1/3$ if D contains neither f_i nor f_0 ; otherwise h holds in one of the $2x \geq 4$ pictures in $\delta(D)$, as D contains f_0 and at most $L - 1 \leq k - 3$ pixel values, leaving $x \geq 2$ pixels unfixed. So h has fit at most zero and is rejected. Only f_0 is open to reevaluation, and its fit with true pixel values is at least $0 > \beta$. So F is never eliminated. $\square$

References

- Abiteboul, Serge, Richard Hull, and Victor Vianu**, *Foundations of databases*, Vol. 8, Addison-Wesley Reading, 1995.
- Agranov, Marina and Pietro Ortoleva**, “Stochastic Choice and Preferences for Randomization,” *Journal of Political Economy*, 2017, 125 (1), 40–68.
- and —, “Revealed Preferences for Randomization: An Overview,” *AEA Papers and Proceedings*, 2022, 112, 426–430.
- Alaoui, Larbi and Antonio Penta**, “Endogenous Depth of Reasoning,” *Review of Economic Studies*, 2016, 83 (4), 1297–1333.
- Ambuehl, Sandro, B Douglas Bernheim, and Annamaria Lusardi**, “Evaluating deliberative competence: A simple method with an application to financial choice,” *American Economic Review*, 2022, 112 (11), 3584–3626.
- Anderson, John R.**, *Rules of the Mind*, Hillsdale, NJ: Lawrence Erlbaum Associates, 1993.
- , **Daniel Bothell, Michael D. Byrne, Scott Douglass, Christian Lebiere, and Yulin Qin**, “An Integrated Theory of the Mind,” *Psychological Review*, 2004, 111 (4), 1036–1060.
- Aragones, Enriqueta, Itzhak Gilboa, Andrew Postlewaite, and David Schmeidler**, “Fact-free learning,” *American Economic Review*, 2005, 95 (5), 1355–1368.
- Asch, Solomon E**, “Forming impressions of personality,” *The journal of abnormal and social psychology*, 1946, 41 (3), 258.
- Backus, John**, “Can Programming Be Liberated from the von Neumann Style? A Functional Style and Its Algebra of Programs,” *Communications of the ACM*, 1978, 21 (8), 613–641.
- Baddeley, A D and G Hitch**, “Working Memory,” in G H Bower, ed., *The psychology of learning and motivation: Advances in research and theory*, Vol. 8, New York: Academic Press, 1974, pp. 47–89.
- Baumeister, RE, E Bratslavsky, M Muraven, and DM Tice**, “Ego depletion: is the active self a limited resource?,” *Journal of Personality and Social Psychology*, 1998, 74 (5), 1252–1265.
- Bénabou, Roland, Armin Falk, and Jean Tirole**, “Narratives, imperatives, and moral reasoning,” *NBER Working Paper 24798*, 2018.
- Bengio, Yoshua, Aaron Courville, and Pascal Vincent**, “Representation Learning: A Review and New Perspectives,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35 (8), 1798–1828.
- Bordalo, Pedro, John J Conlon, Nicola Gennaioli, Spencer Y Kwon, and Andrei Shleifer**, “Memory and probability,” *Quarterly Journal of Economics*, 2023, 138 (1), 265–311.
- , **Nicola Gennaioli, and Andrei Shleifer**, “Salience theory of choice under risk,” *Quarterly Journal of Economics*, 2012, 127 (3), 1243–1285.
- , —, and —, “Salience and asset prices,” *American Economic Review*, 2013, 103 (3), 623–628.
- , —, and —, “Salience and consumer choice,” *Journal of Political Economy*, 2013, 121 (5), 803–843.
- , —, and —, “Competition for attention,” *Review of Economic Studies*, 2016, 83 (2), 481–513.
- , —, and —, “Memory, attention, and choice,” *Quarterly Journal of Economics*, 2020, 135 (3), 1399–1442.
- , —, **Giacomo Lanzani, and Andrei Shleifer**, “A Cognitive Theory of Reasoning and Choice,” *Quarterly Journal of Economics*, 2026. Advance article, qjag023.
- Broadbent, Donald E.**, *Perception and Communication*, London: Pergamon Press, 1958.
- Bushong, Benjamin, Matthew Rabin, and Joshua Schwartzstein**, “A model of relative thinking,”

- The Review of Economic Studies*, 2021, 88 (1), 162–191.
- Caplin, Andrew, Mark Dean, and Daniel Martin**, “Search and Satisficing,” *American Economic Review*, 2011, 101 (7), 2899–2922.
- Chang, Ruth**, “Hard choices,” *Journal of the American Philosophical Association*, 2017, 3 (1), 1–21.
- Chase, William G and Herbert A Simon**, “Perception in chess,” *Cognitive Psychology*, 1973, 4 (1), 55–81.
- Chater, Nick**, “Reconciling simplicity and likelihood principles in perceptual organization,” *Psychological review*, 1996, 103 (3), 566.
- **and Paul Vitányi**, “Simplicity: a unifying principle in cognitive science?,” *Trends in cognitive sciences*, 2003, 7 (1), 19–22.
- Chernev, Alexander, Ulf Böckenholt, and Joseph Goodman**, “Choice Overload: A Conceptual Review and Meta-Analysis,” *Journal of Consumer Psychology*, 2015, 25 (2), 333–358.
- Costa-Gomes, Miguel A and Vincent P Crawford**, “Cognition and behavior in two-person guessing games: An experimental study,” *American economic review*, 2006, 96 (5), 1737–1768.
- Cowan, Nelson**, “The Magical Number 4 in Short-Term Memory: A Reconsideration of Mental Storage Capacity,” *Behavioral and Brain Sciences*, 2001, 24 (1), 87–114.
- Cremer, Jacques, Luis Garicano, and Andrea Prat**, “Language and the Theory of the Firm,” *Quarterly Journal of Economics*, 2007, 122 (1), 373–407.
- De Groot, A D**, *Thought and choice in chess*, Mouton, 1965.
- Dehaene, Stanislas**, *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*, New York: Viking, 2014.
- Dhar, Ravi**, “Consumer preference for a no-choice option,” *Journal of Consumer Research*, 1997, 24 (2), 215–231.
- Eliaz, Kfir and Ran Spiegler**, “A model of competing narratives,” *American Economic Review*, 2020, 110 (12), 3786–3816.
- Enke, Benjamin and Thomas Graeber**, “Cognitive uncertainty,” *Quarterly Journal of Economics*, 2023, 138 (4), 2021–2067.
- Entman, Robert M**, “Framing: Toward clarification of a fractured paradigm,” *Journal of communication*, 1993, 43 (4), 51–58.
- Farah, Martha J, Kevin D Wilson, Maxwell Drain, and James N Tanaka**, “What is” special” about face perception?,” *Psychological review*, 1998, 105 (3), 482.
- Fishburn, Peter C.**, “Nontransitive Measurable Utility,” *Journal of Mathematical Psychology*, 1982, 26, 31–67.
- Flavell, John H.**, “Metacognition and Cognitive Monitoring: A New Area of Cognitive–Developmental Inquiry,” *American Psychologist*, 1979, 34 (10), 906–911.
- Furnham, Adrian and Hua Chu Boo**, “A literature review of the anchoring effect,” *The journal of socio-economics*, 2011, 40 (1), 35–42.
- Gabaix, Xavier**, “A sparsity-based model of bounded rationality,” *Quarterly Journal of Economics*, 2014, 129 (4), 1661–1710.
- **and Thomas Graeber**, “The Complexity of Economic Decisions,” 2024. Working paper.
- Gagnon-Bartsch, Tristan, Matthew Rabin, and Joshua Schwartzstein**, “Channeled Attention and Stable Errors,” 2023. Working paper, Harvard Business School.
- Gennaioli, Nicola and Andrei Shleifer**, “What comes to mind,” *The Quarterly journal of economics*, 2010, 125 (4), 1399–1433.

- Gibbons, Robert, Marco LiCalzi, and Massimo Warglien**, "What situation is this? Shared frames and collective performance," *Strategy Science*, 2021, 6 (2), 124–140.
- Grether, David M. and Charles R. Plott**, "Economic Theory of Choice and the Preference Reversal Phenomenon," *American Economic Review*, 1979, 69, 623–638.
- Gross, David B. and Nicholas S. Souleles**, "Do Liquidity Constraints and Interest Rates Matter for Consumer Behavior? Evidence from Credit Card Data," *Quarterly Journal of Economics*, 2002, 117 (1), 149–185.
- Ilut, Cosmin and Rosen Valchev**, "Economic Agents as Imperfect Problem Solvers," *Quarterly Journal of Economics*, 2023, 138 (1), 313–362.
- Iyengar, Sheena S and Mark R Lepper**, "When choice is demotivating: Can one desire too much of a good thing?," *Journal of Personality and Social Psychology*, 2000, 79 (6), 995.
- Johnson-Laird, Philip N., Paolo Legrenzi, Vittorio Girotto, and Maria Sonino Legrenzi**, "Illusions in Reasoning about Consistency," *Science*, 2000, 288, 531–532.
- Johnson, Samuel GB, Avri Bilovich, and David Tuckett**, "Conviction narrative theory: A theory of choice under radical uncertainty," *Behavioral and Brain Sciences*, 2023, 46, 1–26.
- Kahneman, Daniel**, *Attention and Effort*, Englewood Cliffs, NJ: Prentice-Hall, 1973.
- and **Amos Tversky**, "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 1979, 47 (2), 263–292.
- Kamenica, Emir and Matthew Gentzkow**, "Bayesian persuasion," *American Economic Review*, 2011, 101 (6), 2590–2615.
- Kőszegi, Botond and Adam Szeidl**, "A Model of Focusing in Economic Choice," *The Quarterly Journal of Economics*, 2013, 128 (1), 53–104.
- Legrenzi, Paolo, Vittorio Girotto, and Philip N. Johnson-Laird**, "Models of Consistency," *Psychological Science*, 2003, 14, 131–137.
- Leopold, David A and Nikos K Logothetis**, "Multistable phenomena: changing views in perception," *Trends in cognitive sciences*, 1999, 3 (7), 254–264.
- Lichtenstein, Sarah and Paul Slovic**, "Reversals of Preference between Bids and Choices in Gambling Decisions," *Journal of Experimental Psychology*, 1971, 89, 46–55.
- Loomes, Graham and Robert Sugden**, "Regret theory: An alternative theory of rational choice under uncertainty," *The economic journal*, 1982, 92 (368), 805–824.
- , **Chris Starmer, and Robert Sugden**, "Observing Violations of Transitivity by Experimental Methods," *Econometrica*, 1991, 59, 425–439.
- Luntz, Frank**, *Words That Work: It's Not What You Say, It's What People Hear*, Hachette UK, 2007.
- MacCrimmon, Kenneth R.**, "Descriptive and Normative Implications of the Decision Theory Postulates," in Karl Borch and Jan Mossin, eds., *Risk and Uncertainty*, London: Macmillan, 1968, pp. 3–32.
- Madrian, Brigitte C and Dennis F Shea**, "The power of suggestion: Inertia in 401 (k) participation and savings behavior," *Quarterly Journal of Economics*, 2001, 116 (4), 1149–1187.
- Manzini, Paola and Marco Mariotti**, "Sequentially Rationalizable Choice," *American Economic Review*, 2007, 97, 1824–1839.
- Marr, David**, *Vision: A computational investigation into the human representation and processing of visual information*, MIT press, 2010.
- May, Kenneth O.**, "Intransitivity, Utility, and the Aggregation of Preference Patterns," *Econometrica*, 1954, 22, 1–13.

- Miller, George A**, "The magical number seven, plus or minus two: Some limits on our capacity for processing information.," *Psychological Review*, 1956, 63 (2), 81.
- Mullainathan, Sendhil**, "A memory-based model of bounded rationality," *Quarterly Journal of Economics*, 2002, 117 (3), 735–774.
- , **Joshua Schwartzstein**, and **Andrei Shleifer**, "Coarse thinking and persuasion," *Quarterly Journal of Economics*, 2008, 123 (2), 577–619.
- , —, and —, "Coarse Thinking and Persuasion," *Quarterly Journal of Economics*, 2008, 123 (2), 577–619.
- Neisser, Ulric**, *Cognitive psychology: Classic edition*, Psychology press, 2014.
- Newell, Allen**, *Unified Theories of Cognition*, Cambridge, MA: Harvard University Press, 1990.
- and **Herbert A. Simon**, *Human Problem Solving*, Englewood Cliffs, NJ: Prentice-Hall, 1972.
- Nielsen, Kirby and John Rehbeck**, "When Choices Are Mistakes," *American Economic Review*, 2022, 112, 2237–2268.
- Oprea, Ryan**, "What Makes a Rule Complex?," *American Economic Review*, 2020, 110 (12), 3913–3951.
- , "Decisions under risk are decisions under complexity," *American Economic Review*, 2024, 114 (12), 3789–3811.
- Payne, John W, James R Bettman, and Eric J Johnson**, *The Adaptive Decision Maker*, Cambridge University Press, 1993.
- Pearl, Judea**, *Causality: Models, Reasoning, and Inference*, 2 ed., Cambridge: Cambridge University Press, 2009.
- Petrov, Valentin V.**, *Sums of Independent Random Variables*, Berlin: Springer-Verlag, 1975.
- Regenwetter, Michel, Jason Dana, and Clinton P. Davis-Stober**, "Transitivity of Preferences," *Psychological Review*, 2011, 118, 42–56.
- Rissanen, Jorma**, "Modeling by shortest data description," *Automatica*, 1978, 14 (5), 465–471.
- Rothman, Naomi B, Michael G Pratt, Laura Rees, and Timothy J Vogus**, "Understanding the dual nature of ambivalence: Why and when ambivalence leads to good and bad outcomes," *Academy of Management Annals*, 2017, 11 (1), 33–72.
- Rubinstein, Ariel**, "Similarity and Decision-Making under Risk (Is There a Utility Theory Resolution to the Allais Paradox?)," *Journal of Economic Theory*, 1988, 46, 145–153.
- Sasao, Tsutomu**, *Switching Theory for Logic Synthesis*, Boston: Kluwer Academic Publishers, 1999.
- Scheibehenne, Benjamin, Rainer Greifeneder, and Peter M. Todd**, "Can There Ever Be Too Many Options? A Meta-Analytic Review of Choice Overload," *Journal of Consumer Research*, 2010, 37 (3), 409–425.
- Schwartzstein, Joshua**, "Selective Attention and Learning," *Journal of the European Economic Association*, 2014, 12 (6), 1423–1452.
- and **Adi Sunderam**, "Using models to persuade," *American Economic Review*, 2021, 111 (1), 276–323.
- and —, "Sharing Models to Interpret Data," 2024. Working paper, Harvard Business School.
- Shafir, Eldar, Itamar Simonson, and Amos Tversky**, "Reason-based choice," *Cognition*, 1993, 49 (1-2), 11–36.
- Shiller, Robert J., Fumiko Kon-Ya, and Yoshiro Tsutsui**, "Why Did the Nikkei Crash? Expanding the Scope of Expectations Data Collection," *Review of Economics and Statistics*, 1996, 78 (1), 156–164.

- Simon, Herbert A.**, "A behavioral model of rational choice," *Quarterly Journal of Economics*, 1955, pp. 99–118.
- Simon, Herbert A.**, "From Substantive to Procedural Rationality," in Spiro J. Latsis, ed., *Method and Appraisal in Economics*, Cambridge: Cambridge University Press, 1976, pp. 129–148.
- Sims, Christopher A.**, "Implications of rational inattention," *Journal of Monetary Economics*, 2003, 50 (3), 665–690.
- Stahl, Dale O. and Paul W. Wilson**, "Experimental evidence on players' models of other players," *Journal of Economic Behavior and Organizations*, 1994, 25 (3), 309–327.
- Sweller, John**, "Cognitive load during problem solving: Effects on learning," *Cognitive Science*, 1988, 12 (2), 257–285.
- Telyukova, Irina A.**, "Household Need for Liquidity and the Credit Card Debt Puzzle," *Review of Economic Studies*, 2013, 80 (3), 1148–1177.
- Thaler, Richard**, "Toward a positive theory of consumer choice," *Journal of economic behavior & organization*, 1980, 1 (1), 39–60.
- Thaler, Richard H and Shlomo Benartzi**, "Save more tomorrowTM: Using behavioral economics to increase employee saving," *Journal of Political Economy*, 2004, 112 (S1), S164–S187.
- Tversky, Amos**, "Intransitivity of Preferences," *Psychological Review*, 1969, 76, 31–48.
- , "Elimination by aspects: A theory of choice," *Psychological review*, 1972, 79 (4), 281.
- **and Daniel Kahneman**, "Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty," *science*, 1974, 185 (4157), 1124–1131.
- **and Eldar Shafir**, "Choice under conflict: The dynamics of deferred decision," *Psychological Science*, 1992, 3 (6), 358–361.
- , **Shmuel Sattath, and Paul Slovic**, "Contingent Weighting in Judgment and Choice," *Psychological Review*, 1988, 95, 371–384.
- Wegener, Ingo**, *The complexity of Boolean functions*, John Wiley & Sons, Inc., 1987.
- Wilcox, Nathaniel T.**, "Lottery Choice: Incentives, Complexity and Decision Time," *Economic Journal*, 1993, 103 (421), 1397–1417.
- Wilson, Andrea**, "Bounded memory and biases in information processing," *Econometrica*, 2014, 82 (6), 2257–2294.
- Wojtowicz, Zachary**, "Model Diversity and Dynamic Belief Formation," *Working paper*, 2024.
- Yin, Robert K**, "Looking at upside-down faces," *Journal of experimental psychology*, 1969, 81 (1), 141.