Getting the Picture

Robert Akerlof, Richard Holden, and Hongyi Li*

July 14, 2026

Abstract

Standard economic theory assumes that people have perfect information processing ability: they can derive *all* logical consequences of information *immediately*. In reality, however, people struggle to work out the consequences of information. Faced with a decision such as whether to refinance a mortgage or when to realize a capital gain, a person may hold every relevant fact and still fail to see what follows. In this paper, we offer a theory of the reasoning process—how people go about "connecting 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*. We show that limited processing capacity makes sense of why people suffer from choice overload, are influenced by defaults, engage in satisficing, selectively attend to attributes, are subject to primacy effects, and vacillate between interpretations. We apply the framework to political persuasion, showing how supplying a simple narrative can lock in an interpretation of ambiguous evidence.

JEL Classification: D01, D80, D90.

Keywords: Cognition, reasoning, perception, narratives.

*Akerlof: UNSW Business School and University of Warwick, 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, 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—notably the work of Bordalo, Gennaioli, Shleifer, and coauthors—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. Miller (1956) argued that people can hold only about seven items in mind at once; Broadbent (1958) modeled attention as a limited-capacity channel; Kahneman (1973), whose capacity model of attention we draw on, treated attention and effort as the allocation of a limited mental capacity; and Cowan (2001) places the capacity of the "focus of attention" closer to four items. We borrow the term *processing capacity* from this literature, using it in a broad sense: the limited resource that can be brought to bear on combining information at any one moment.

We deliberately remain agnostic about whether this processing is conscious or unconscious—our mechanism turns only on its being capacity-limited, however it is realized in the brain. The two may nonetheless be related: there is some evidence that the processing which rises to consciousness is precisely that which integrates information across distant, otherwise-separate regions of the brain (see Dehaene, 2014).

To illustrate the role of processing capacity, consider how ordinary financial decisions strain the limits of what people can keep in mind. Deciding whether to refinance a mortgage, how much to save for retirement, when to claim Social Security, whether to buy or rent a home, how to allocate a portfolio, or how to time the sale of an asset—each of these requires the decision-maker to hold several facts in mind at once and combine them.

For instance, suppose 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, his 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 is larger, and he seeks the year of sale that minimizes his total tax across the two years. His situation can be summarized in five facts:

F1. His ordinary income is $100,000 this year and $200,000 next year, the capital gain is $400,000, the ordinary tax rates are 30% on income and 15% on the gain, and the minimum tax has a 25% rate and a $100,000 exemption.

F2. The gain falls entirely in whichever year he chooses.

F3. The ordinary tax is the tax on income plus the tax on the gain.

F4. The minimum tax combines ordinary income and gain into one base and taxes the amount above the exemption.

F5. In each year he pays whichever is larger, the ordinary tax or the minimum tax.

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. Yet almost no one can simply read this answer off the list. Each fact is individually plain—the sort of thing a sophisticated investor grasps on its own. The difficulty lies not in any one fact but in putting them together: in getting, from five separate pieces, the picture they jointly form.

Mr. Green reasons by combining facts, and combining facts yields new ones. Some follow from just two. From F1 and F4 he can see that this year only the gain can push his minimum-tax base above the exemption: his \$100,000 of income exactly reaches the \$100,000 threshold, so on its own it leaves nothing above it. Others require three: to find the minimum tax if he sells this year, he combines F1, F2, and F4—the figures, the gain's landing this year, and the minimum-tax rule—arriving at 25% of (\$100,000 + \$400,000 − \$100,000), or \$100,000. And the conclusion that matters most—which year to sell—draws on all five. Combining facts is harder the more of them must be combined at once. To combine several facts, he needs to hold them in mind together and process them at the same time; a new fact that follows from two facts is one he can often see at a glance, while one that follows from three, or five, is progressively harder to reach—not because the additional facts are themselves difficult, but because each is one more thing to keep active in mind alongside the others.

One way Mr. Green can ease this difficulty is to proceed in *manageable steps*, reaching the right answer—but not *immediately*. A step combines a few facts into a new one; Mr. Green retains the new fact in place of the originals, so he can build on previous thoughts without needing to process every fact at once. To find the minimum tax for selling this year, he might first add his income and the gain (\$100,000 + \$400,000 = \$500,000), then subtract the exemption (\$500,000 − \$100,000 = \$400,000), then apply the rate (25% ×

$400,000 = $100,000). Each step takes only one or two quantities and yields a single new one, which he carries forward in place of the materials that produced it: having reached $100,000, he no longer needs to hold the income, the gain, and the exemption together. Earlier conclusions serve the same role. For example, having established that the minimum tax does not bind in whichever year the gain does not land, he can determine that the tax is the ordinary tax without revisiting the minimum tax. Working in steps thus enlarges what Mr. Green can figure out: conclusions he could not see in a single stroke become attainable once decomposed.

Even with the help of steps, the demands of such a problem may lead some investors to take shortcuts, *extrapolating* from a salient piece of the problem to a judgment about the whole. Mr. Green might notice that the ordinary tax takes the gain at 15% whichever year he sells, and conclude that the timing makes no difference—so he may as well sell now. The reasoning sounds sensible, but it quietly drops a pertinent fact—that realizing the gain triggers the minimum tax, so that the timing does affect the tax on his ordinary income. The shortcut costs him $5,000. As we will see, extrapolations of this kind are not random error; they give rise to biases that are systematic and predictable.

The model we develop captures the key aspects of Mr. Green's cognitive process. An agent faces a choice problem in which the optimal choice depends on the state of the world. We refer to the state of the world as the "picture" and the elements that make it up as "pixels." In Mr. Green's case, the pixels are quantities such as his ordinary income in each year, the size of the capital gain, and the tax rates.

We define a "feature" as an event in picture-space—for instance, "the tax paid equals the larger of the ordinary tax and the minimum tax"—representing something the agent might know about the picture. To make a choice, the agent must work out a *decision-relevant* feature—one that reveals which action is best. For Mr. Green, the decision-relevant feature is "the total tax is higher if the gain is realized in year 1 than in year 2."

The agent's time-t "knowledge" is the set of features they take to apply to the picture at time t—a store of everything they have worked out so far, which they can draw on freely in their reasoning. We assume the agent starts with some initial knowledge. Even when this initial knowledge contains all the *information* needed to work out the decision-relevant feature, further work is required to learn it. In Mr. Green's case, his starting facts F1–F5 already entail that "the total tax is higher if the gain is realized in year 1 than in year 2"—but he must still reason his way to that conclusion.

The agent learns by drawing on facts they already know to reach new conclusions. Mr. Green, for instance, might combine the figures (F1), the fact that the gain lands in the year he sells (F2), and the rule that the ordinary tax is the tax on income plus the tax on the gain (F3) to conclude that, if he sells this year, his ordinary tax for the year is \$90,000—30% of his \$100,000 income plus 15% of the \$400,000 gain. Each new conclusion is added to the agent's knowledge.

The agent has limited processing capacity, which restricts how many facts they can draw on at any one time. This is the central mechanism of our model: in computing terms, the agent has scarce RAM but a vast hard drive—knowledge is boundless, processing scarce.[1] Working alongside this limit is a second: the agent has only a limited set of *concepts*, and can draw only on facts they can conceptualize.[2] Concepts shape what capacity can accomplish: without intermediate ones such as "the minimum-tax base" or "the total tax across both years" to build on, even an agent with ample capacity would struggle to assemble the raw facts into a decision.

We consider two variants of the model. In the first, the agent is purely deductive; in the second, the agent extrapolates. In the version with extrapolation, the agent accepts conclusions when they "fit the data" sufficiently well—in the sense that relatively few alternative explanations remain. Mr. Green's shortcut above—to the conclusion "the timing makes no difference"—is an example.

Extrapolation makes sense of a variety of phenomena—such as satisficing behavior (in the sense of Simon (1955)), selective attention to attributes, and choice overload. Choice

[1]We treat this store of knowledge as unbounded; it may be bounded in reality, but treating it so isolates the role of limited processing capacity.

[2]Cognitive 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)).

overload reflects the idea that as a choice problem gets more complex, it becomes harder to extrapolate to a conclusion about which action is best.

Moreover, under extrapolation, the agent's limited processing capacity can lead to systematic biases in behavior, with the direction of bias determined endogenously by the structure of the choice problem. For instance, Mr. Green might consistently choose to realize capital gains early, when it is better to wait, if the benefit of selling early is relatively clear (e.g. it frees up capital to make other investments) while the cost is less clear (e.g. the alternative minimum tax binds).

Extrapolation lets the same picture be read in more than one way. Multiple readings are familiar in economic life: before the 2008 crisis, some saw rising house prices as driven by fundamentals, others as a bubble—one set of facts, two conclusions. With one further ingredient—an assumption that "powerful" features are particularly salient—the reading the agent reaches becomes *path dependent*: where reasoning begins shapes where it ends, and the first reading tends to *stick*. This is well recognized in cognitive psychology; a vivid example is the rabbit–duck illusion (Figure 1), where, having seen a rabbit, one finds it hard to see a duck (and vice versa).[3] Path dependence makes sense of a large body of evidence on primacy effects, in which information encountered early shapes judgments more than information encountered later. It also bears on how one agent can persuade another: by suggesting a *narrative*—a thought for the agent to consider—a persuader can steer how the agent makes sense of the picture. In politics, this is the advantage of being first to set the narrative, before one's opponent can offer a rival one.

Related Literature

There is, of course, a vast literature in economics on bounded rationality, starting with the work of Simon (1955) who pointed to "limits on computational capacity" as an important constraint on "actual human choice." The highly influential line of work on heuristics and

[3]The Gestalt school studied such figures closely, taking from them the lesson that perceiving the whole of a picture is distinct from perceiving its parts—the same parts, differently organized, becoming a rabbit or a duck (see Wertheimer, 1912). Seeing all the parts without seeing the whole is the visual counterpart of knowing all the facts without seeing what follows from them.

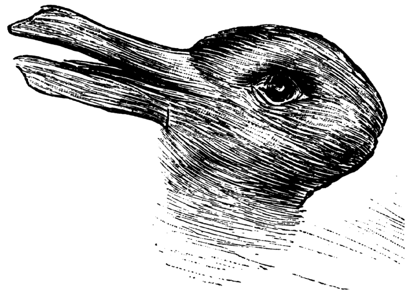

Figure 1: The Rabbit-Duck Illusion

biases, initiated by Kahneman and Tversky, suggests that people rely on mental shortcuts to make decisions due to cognitive limitations, though these shortcuts can sometimes lead to systematic errors (see Tversky and Kahneman (1974) and Thaler (1980)).

Much of the work on bounded rationality falls into one of two broad approaches. The first posits a behavioral tendency as a primitive—present bias, say, or loss aversion—and traces its consequences. The second retains a Bayesian agent but subjects them to a cognitive constraint: a limit on the information they can attend to, as in the rational-inattention literature (Sims, 2003),[4] or a bound on memory (Mullainathan, 2002; Wilson, 2014), or a cost of deliberate thinking, as in Ilut and Valchev (2023).[5] Our approach differs from both. We do not posit biases—they emerge from the reasoning process—and our agent is not a Bayesian optimizer: in fact, they have no priors to guide them.

Two strands of work bear more closely on ours. The closest is the body of work by Bordalo, Gennaioli, Shleifer, and coauthors (hereafter BGS), which seeks to "get inside people's heads" and draw out the consequences for decision-making (Bordalo et al., 2023, 2012, 2013a,b, 2016, 2020, 2026; Gennaioli and Shleifer, 2010; Mullainathan et al., 2008). 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

[4]Relatedly, Gabaix (2014) models an agent who copes with such limits by attending *sparsely*—building a simplified representation that omits variables of little importance. The agent optimizes, but over a deliberately simplified model rather than as a full Bayesian.

[5]A 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.

most salient attributes of a problem loom largest in choice. On memory, they model how cues shape what an agent recalls, and how those recollections in turn color judgment.[6]

Our framework is highly 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*. The binding limit is that the agent can hold only a few facts together: Mr. Green's difficulty is not that some fact escapes his notice, but that he cannot assemble the facts he already has. 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.[7]

A second framework that bears closely on ours 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 agree that complexity distorts choice, but locate the distortion differently. In theirs, the *direction* of any systematic bias is set by the default, supplied exogenously; in ours, it is set endogenously by 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—to which Enke and Graeber (2023) themselves contribute, alongside Oprea (2024) and Ambuehl et al. (2022)—showing that much of what looks like nonstandard preference is the imprint of computational constraints (for an overview, see

[6]Closest to our extrapolation is the local-thinking model of Gennaioli and Shleifer (2010), in which the agent also judges from a partial representation of the problem. In their model, the agent considers only the scenarios that come most readily to mind; in ours, the agent draws on only the few facts they can hold in mind at once.

[7]Limited *depth* of reasoning is another such phenomenon. In level-k and cognitive-hierarchy models, agents reason only a bounded number of steps about others' strategies (Costa-Gomes and Crawford, 2006; Stahl and Wilson, 1994). This literature takes a distinct approach to reasoning limits, developed specifically for strategic settings, where the relevant bound is on iterated beliefs about others; the limit nonetheless resonates with ours, reflecting how far an agent can carry a chain of reasoning under limited processing capacity. Within this strand, Alaoui and Penta (2016) endogenize the depth of reasoning through a cost-benefit tradeoff—reasoning proceeds while its value exceeds its cost—which parallels the possibility, discussed in Section 3, that an agent facing a difficult choice lowers their extrapolation threshold until a conclusion becomes reachable.

Gabaix and Graeber, 2024).[8]

Our paper also relates to a literature on narratives and mental models (e.g. Eliaz and Spiegler (2020); Gibbons et al. (2021); Schwartzstein and Sunderam (2021); Shiller (2017); Wojtowicz (2024)).[9] Much like Schwartzstein and Sunderam (2021), we think of a narrative as something the agent entertains and that a persuader can suggest. In our framework it is a conjecture, together with facts that support it. Our result on path dependence helps explain why narratives are sticky, and why it is valuable to "control the narrative."

Aragones et al. (2005) also begin from the observation that an agent may hold all the relevant facts yet not know what follows from them. They show that extracting regularities from a database is computationally hard (NP-complete), and note that persuasion can consist of pointing out an overlooked regularity. But theirs is a complexity result about reasoning's endpoint. They do not model the reasoning process: what features of the data an agent knows and how knowledge grows as features are combined, step by step, under limited processing capacity.

We draw on a large literature in cognitive psychology and computer science. Closest to our model of the reasoning process is the tradition of Newell and Simon (1972), in which an agent solves problems step by step under limited working memory, storing intermediate results for later use (see also Anderson, 1993); we embed this process view in an economic model of choice and derive its behavioral consequences. A second strand concerns *chunking* (Chase and Simon, 1973; Miller, 1956), the compression of several pieces of information into a single more memorable unit (see Section 2.3). A related literature on *minimum description length* holds that people favor explanations that encode

[8]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.

[9]Schwartzstein and Sunderam (2021) conceive of an agent as choosing a mental model from an available set that is potentially influenced by a persuader. Eliaz and Spiegler (2020) conceive of narratives as causal interpretations of events (specifically, directed acyclic graphs). Gibbons et al. (2021) conceive of narratives in terms of categorizations and examine how a group's shared narrative affects their capacity to cooperate. Bénabou et al. (2018) model the process by which narratives disseminate. Perhaps most closely related is Wojtowicz (2024), who considers a different mechanism that can lead to path dependence: the speed at which new data arrives. In Wojtowicz (2024), an agent repeatedly evaluates their existing model against "nearby" alternatives, and switches whenever the alternative has better fit. If data arrives too quickly, path dependence may arise and the agent may not converge to the maximum-likelihood model.

the data simply (Chater, 1996; Chater and Vitányi, 2003; Feldman, 2000)—an intuition our framework shares, since processing capacity puts a premium on economical representations.[10] From computer science, we draw on database theory (Abiteboul et al., 1995), where the central problem is to store data and retrieve useful conclusions from it efficiently; its distinction between storage and retrieval parallels ours between knowledge and processing, the bottleneck in each case lying not in what is stored but in what can be brought to bear at once.

Roadmap

The paper proceeds as follows. Section 2 presents a version of the model in which the agent is purely deductive, and shows how limited processing capacity—together with the agent's stock of concepts—bounds the conclusions they can reach. Section 3 allows the agent to extrapolate—to accept conclusions that fit the facts well enough, rather than only those that follow from them with certainty—and shows how this gives rise to satisficing, choice overload, and systematic bias. Section 4 shows how the model gives rise to path-dependent thinking—where what an agent concludes depends on where their reasoning begins—and draws out the implications for persuasion. Section 5 takes up three extensions: the deletion of features and the cycling it can produce, the selection of the agent's concepts, and their generalization to an internal code. Section 6 concludes. Proofs are in the appendix.

[10] Our approach differs from an influential tradition that casts induction as Bayesian inference guided by fixed, hard-wired priors (Oaksford and Chater, 2007; Tenenbaum et al., 2006): in our framework the agent has no priors to guide them and must work out which model they are in. More broadly, most work in psychology, like economics, emphasizes the outputs of inductive processes—the conclusions people draw—whereas we examine how they build up partial insights step by step.

2 Deduction

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 xth vector is $P_x = (p_{x1}, p_{x2}, \ldots, p_{xM_x})$. When the vectors share a common length ($M_1 = \cdots = M_N = M$), 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 taking 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, τ_I^{ord}, τ_K^{ord}, τ^{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 $T_j(i)$.

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

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 $\{Q \in \mathcal{P} : T_j(i) = \max(T_j^{\text{ord}}(i), T_j^{\text{amt}}(i))$ 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 $u : A \times \mathcal{P} \to \mathbb{R}$ 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."[11] 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.

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

[11] We 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.

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 ($K_0 \subseteq C$).[12] 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, by contrast, consists of the five facts F1–F5 he is told about the problem. 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 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

[12]Since 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.

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. Two forces shape the answer: 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.[13]

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 $\theta_0, ..., \theta_{\tau-1}$ 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:*

$$s(K) \; = \; \min\Big\{|F| : F \subseteq K, \, \textstyle\bigcap_{f \in F} f = \{P\}\Big\},$$

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 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.

[13] Section 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.

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. When $L \geq s(K_0)$, the agent can draw at once on a subset of K_0 that specifies P, and from it deduce any concept that applies—so everything is reachable in a single step. When $L = 1$, each thought draws on just one fact, and deduction from a single fact yields only what that fact already implies; nothing new can be built, so only derivative concepts are reachable. Part 2 is the substantive intermediate case: there are problems in the range $1 < L < s(K_0)$ where the agent can deduce some but not all concepts that apply to P, and some deductions require multi-step reasoning chains.

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 "$T_1(1) = \$100{,}000$."

Each step stays within the budget of five, and the most demanding—Steps 2 and 3—use exactly five facts. 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: if the agent has processing capacity of two or more and their conceptualization is rich enough to contain every feature that applies to P, then every such feature is reachable, $R(C) = C(P)$. Drawing on two facts at a time, the agent can intersect known features pairwise—each intersection is again a feature that applies to P, and so is itself a concept the agent can hold—until P is pinned down exactly, after which any feature containing P follows in one further step. Some features may nonetheless be reachable only after an extremely large number of steps, and in this sense may be *close* to unreachable.

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.[14]

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

An important finding in cognitive psychology is that short-term memory is limited not so much by the sheer volume of information we are trying to hold as by how that information is organized. This idea was pioneered by Miller (1956), who introduced the term *chunking* to describe the process of binding small, individual pieces of information into a smaller number of meaningful units, or “chunks,” that can be held more readily in mind.

[14]For 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 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).

A more striking illustration of chunking comes from chess. In a well-known experiment, Chase and Simon (1973) showed expert and novice chess players a board position for five seconds and then asked them to reconstruct it from memory.[15] When the position was taken from an actual game, experts dramatically outperformed novices: on average, experts correctly recalled the locations of about 16 out of 20 pieces, while novices recalled only about 4. The standard interpretation is that experts learn to chunk familiar chess configurations, which allows them to store and retrieve complex positions efficiently.

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 shows how the Chase and Simon experiment can be cast in these terms. Panel (a) is a 5×10 picture whose pixels take the values "black" or "white," standing in for the positions of pieces on a board. Think of the expert as an agent whose conceptualization includes the concepts "the left-hand side is an H" and "the right-hand side is an I"; drawing on just these two, they can hold the entire board in mind, as in Panel (b). The novice, by contrast, can conceptualize only pixel values, and so must draw on all fifty separately—far exceeding any plausible L. The expert's edge thus reflects a difference in conceptualization, not in 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

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

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.

(a) Game Board (b) Game Board - Chunked

Figure 2: Chunking a Game Board

3 Extrapolation

Through deduction alone, an agent may be unable to reach a decision, for either of two reasons. Sometimes the conclusion follows from the agent's 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 do not pin down the picture fully—and what they leave unknown, no deduction can recover.

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 the conjecture when it is merely *suggested* by those facts. Extrapolation addresses both limits. The first function is to reach conclusions that follow from the agent's knowledge but lie beyond what they can deduce with limited processing capacity. The second is to *predict*: when the agent's initial knowledge does not fully specify the picture ($\bigcap_{f \in K_0} f \supset \{P\}$), extrapolation lets them make an educated guess at what is left open—filling in blanks the facts alone leave empty. Both functions extend the agent's reach, but the same license to go beyond the facts opens the door to error—and, as we will see, the errors are not merely random noise. Extrapolation introduces biases that are systematic and predictable.

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

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

Intuitively, $\psi\big(D_t \cup \{\phi_t\}\big) - \psi(D_t)$ 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.[16]

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 $\delta\big(D_t \cup \{\phi_t\}\big) = \delta(D_t)$.

3. Fit equals $-\infty$ if D_t and ϕ_t are incompatible ($\delta\big(D_t \cup \{\phi_t\}\big) = \varnothing$).

4. Fit equals zero when D_t is uninformative about ϕ_t (for instance, if $D_t = \varnothing$).

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 $\alpha = 1$.

[16]More 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 $\psi\big(D_t \cup \{\phi_t\}\big) - \psi(D_t) = \log|\delta(D_t)| - \log\Big|\delta\big(D_t \cup \{\phi_t\}\big)\Big|$ 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.

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.[17]

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

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

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

[17]The 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.

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.[18] 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 2×2 matrix

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

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.[19]

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?

[18]This 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.

[19]Mr. 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.

Consider first the conjecture that "Suit 1 is weakly preferred" ($\phi_t = f_1$). Its fit with the two price ratings is

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

The pieces are these: $|\mathcal{P}| = 10^4$, 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. The fit is near one because, given Suit 1's large price advantage, just one case remains 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 preferred" and choose Suit 1.[20]

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.[21]

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 $\alpha \leq 0.83$.

Putting the cases together: with a very high threshold ($\alpha > 0.98$) Mr. Stitch is stuck, unable to extrapolate to either conclusion. With a moderate threshold ($0.83 < \alpha \leq 0.98$) he can reach only "Suit 1 is weakly preferred," by attending to price. And with a low

[20] Here 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.

[21] When 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.

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

enough threshold ($\alpha \leq 0.83$) he can reach either—attending to price yields "Suit 1 is weakly preferred," attending to quality yields "Suit 2 is weakly preferred."

That Mr. Stitch may land either way when his threshold is low makes his choice to some degree indeterminate: it turns on which attribute he happens to attend to. This accords with an experimental literature documenting a random element in choice. Agranov and Ortoleva (2017) posed the same questions to subjects repeatedly and found that a large majority chose differently across repetitions; a follow-up survey by Agranov and Ortoleva (2022) documents such randomization across a range of domains, including objective lotteries, ambiguous gambles, social preferences, and time preferences.

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

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

Here $\psi(D_t)$ and $\psi(\{f_1\})$ are as before; what differs is $\psi(D_t \cup \{f_1\}) = \log 10^4 - \log 72$, where 72 is the number of Suit 2 price–quality rating combinations for 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 shows.

Proposition 3. *Consider the choice among n products described by m attributes, with ratings in* $\{1, \ldots, V\}$*, 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* $v_{1y}, \ldots, v_{ny}$*. 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*. If* u_j *is low enough, the consumer can extrapolate to "product* i *is weakly preferred" and choose* i*.*

As in the Mr. Stitch example, along the attribute route the consumer can choose product i only if it leads by a wide enough margin on the attended attribute. Likewise, along the product route, the consumer can choose it only if its own value is high enough or a rival's value 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.[22]

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

[22] That decision-making falters when an agent cannot reach a conclusion about which option is best is supposed across disciplines. Shafir et al. (1993)'s notion of "reason-based choice" holds that people struggle to choose when they cannot find a unifying rationale for one option over another. The cognitive-load theory of Sweller (1988) likewise links the failure to integrate disparate information to heightened cognitive burden and the risk of "analysis paralysis" (Baumeister et al., 1998). Relatedly, Chang (2017) examines the difficulty of choosing in the absence of clear evaluative criteria, and a literature on conviction narratives (Johnson et al., 2023) argues that people reach confident decisions only once they can form an integrated understanding of their situation.

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).[23]

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.

$$\begin{array}{c} \\ \text{Suit 1} \\ \text{Suit 2} \end{array} \begin{array}{c} \begin{array}{cc} \text{Price Rating} & \text{Quality Rating} \end{array} \\ \left[\begin{array}{cc} 7 & 5 \\ 2 & 5 \end{array}\right] \end{array} \qquad\qquad \begin{array}{c} \\ \text{Suit 1} \\ \text{Suit 2} \end{array} \begin{array}{c} \begin{array}{cc} \text{Price Rating} & \text{Quality Rating} \end{array} \\ \left[\begin{array}{cc} 7 & 5 \\ 6 & 5 \end{array}\right] \end{array}$$

(a) (b)

Figure 4: Two choice problems for Mr. Stitch

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 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

[23] A 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.

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.[24]

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 4. *Consider the product-choice setting of Proposition 3. 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:*

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.*

[24]Subsequent meta-analyses find that choice overload is conditional rather than universal (Chernev et al., 2015; Scheibehenne et al., 2010), and the conditions under which it appears are those the model predicts. Overload is strongest when options are numerous and attributes many—as we show below, fit falls as n and m grow; when no option stands out—when options are similar, 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 before a conclusion is reachable (see Section 3.5). Conversely, overload should recede with expertise, which supplies the intermediate concepts that ease integration.

The intuition is one of 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.

We can illustrate both aspects of the proposition with Mr. Stitch's original problem. Suppose Mr. Stitch follows the attribute route and attends to price, on which Suit 1 leads by eight. Holding that lead fixed, suppose we add further attributes beyond the original two. The fit of "Suit 1 is weakly preferred" falls as their number grows (Table 3).

# of suits	# of attributes	Fit
2	2	0.98
2	3	0.88
2	4	0.81
2	5	0.75

Table 3: Fit of "Suit 1 is weakly preferred" when one attribute is in mind

Now suppose Mr. Stitch follows the product route and attends to Suit 1, which has the highest value—thirteen. Holding that value fixed, suppose we add further suits beyond the original two. The fit of "Suit 1 is weakly preferred" again falls as their number grows (Table 4).

# of suits	# of attributes	Fit
2	2	0.48
3	2	0.34
4	2	0.22
5	2	0.11

Table 4: Fit of "Suit 1 is weakly preferred" when one suit is in mind

In short, adding products or attributes makes the problem harder to resolve. And it grows harder still with further richness—for instance, when the consumer must also extrapolate the relationship between attributes and value, a problem we take up below.

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 they gradually lower their fit threshold α—becoming more willing to extrapolate—until a choice becomes feasible.[25] This accords with a large experimental literature finding that harder choices take longer to make (e.g. Payne et al., 1993).

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. A natural question is whether such reasoning difficulties merely make choice noisier or whether they can push the consumer toward systematically wrong conclusions. Noise leaves deviations from the optimal choice undirected, favoring no option in particular; a systematic bias, by contrast, reliably favors one suboptimal option. We show that limited processing capacity produces the latter, and that the direction of the bias is fixed endogenously by the structure of the problem.

The distinction between random error and systematic bias is central to the behavioral tradition initiated by Kahneman, Tversky, and Thaler (Kahneman and Tversky, 1979; Thaler, 1980; Tversky and Kahneman, 1974): random errors would tend to cancel in the aggregate, leaving the rational model approximately right, whereas systematic biases survive aggregation and can move markets, distort prices, and generate persistent welfare losses.

Our model generates such biases. 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 pro-

[25] An agent can always reach a choice once the threshold is low enough.

cessing 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.

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

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 3—attending to an attribute, or to a product—systematic bias admits a clean characterization.

Proposition 5. *Consider a choice between* two *products, and suppose product 2 has the higher value ($u_2 > u_1$). Suppose the consumer's thoughts draw on either both products' ratings for a* *single attribute, or on a single product's ratings in full—the routes of Proposition 3. 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 3, 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'*.*

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 $\big(\max\{F_m(\Delta_2),\, G_m(u_2),\, H_m(u_1)\},\ F_m(\Delta_1)\big]$.

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.

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.[26] 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.

[26] Formally, 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.

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. These are complementary in a different way. 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.[27]

3.5 Filling in the Blanks

So far, extrapolation has played one role: helping the consumer reach conclusions that follow from K_0 but are hard to see by deduction alone. It has a second role as well. When K_0 does not fully specify the picture ($\bigcap_{f \in K_0} f \supset \{P\}$)—fixing most of the pixels, say, but leaving some open—deduction is silent about what K_0 does not determine. Extrapolation, however, lets the consumer make an educated guess at what is left open.

To see this, extend the product-choice model we have been considering. Until now the consumer has valued a product by the total of its ratings—and has known that this is how they value it. Suppose instead that they do not know, ex ante, how ratings translate into utility. The picture now carries an extra column recording the utility each product delivers:

$$P = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1m} & u_1 \\ v_{21} & v_{22} & \cdots & v_{2m} & u_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ v_{n1} & v_{n2} & \cdots & v_{nm} & u_n \end{bmatrix}.$$

[27] Salience and limited processing can be hard to tell apart, since both work by confining the agent to part of the picture—salience through what draws the agent's notice, ours through what the agent can hold at once. But the accounts can be distinguished. Bushong et al. (2021) is already set apart from our account by the direction of choice: relative thinking underweights 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 Kőszegi 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*: where a fixed weighting always delivers a verdict, our consumer may reach none—deferring, or falling back on a default.

The consumer knows every rating v_{xy}, but only some of the utilities u_x—those of products they have tried before. For the rest, u_x is a blank.

From the products they know—those for which they know both the ratings and the utility they yield—the consumer may be able to extrapolate to a feature describing how ratings translate into utility in general: one of the form "$u_x = U(v_{x1}, \ldots, v_{xm})$ for all x." Settling on such a feature amounts to settling what their utility function looks like—whether the attributes enter additively or interact, some complementary and others substitutes. What they have inferred is not a missing number but a *rule*.

With the rule in hand, the consumer can fill in the blanks. For a product x whose utility they do not know, they draw on the rule and the product's known ratings $v_{x1}, \ldots, v_{xm}$ and work out u_x, the product's value. Notice that K_0 did not pin u_x down; the consumer gets there only by first inferring a rule and then building on it.

This kind of reasoning is ubiquitous. Take the game show *Wheel of Fortune*. Shown the partial phrase "_OOD _OR_ING," a contestant supplies the missing letters by first working out the whole phrase—"GOOD MORNING"—and then using it to fill in the G, M, and N. First a guess at the whole, then the whole used to recover the parts—the same two steps as the consumer's.[28]

What this adds to the agent is the capacity to *predict*. An agent who only deduces is confined to what their starting facts already contain; they can draw out their implications, but they cannot reach past them. By extrapolating to a rule and then applying it, the agent goes further—using the structure they have inferred to fill in what they were never told. This is one of reasoning's main functions. We rarely have complete information about the situations we face; what lets us act anyway is the ability to read a pattern from the part of the picture we can see and use it to anticipate the part we cannot. The same move that lets the consumer value a product they have never tried lets us finish a half-heard

[28]These two directions correspond to a distinction familiar in cognitive psychology between *bottom-up* and *top-down* thinking (see Marr, 2010; Neisser, 2014). In our framework, bottom-up thinking starts from pixels and builds toward higher-level features that speak to many of them; top-down thinking starts from a higher-level feature and reasons from it about the individual pixels.

sentence, recognize an object from a glimpse, or judge how an unfamiliar situation is likely to unfold.

What the agent predicts, though, depends on the concepts they hold. The contestant lands on "GOOD MORNING" rather than "WOOD FORMING" because the former is a concept for them and the latter is not. They fill the blanks, that is, in a way that accords with their conceptualization—reaching for a familiar pattern that fits what they can see. The consumer is no different: if the only utility rule they can conceptualize is additive, that is the rule they will infer, and their predicted valuations will take its form 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 either a rabbit or a duck, but not both at once; and whichever you see first, you tend to keep seeing—it takes real effort to make the other appear. What you end up perceiving thus depends on where your perception started. This is a kind of *path dependence*, and a robust feature of perception: psychologists call it *hysteresis*—the tendency of a current interpretation to persist against its rival (Blake and Logothetis, 2002; Leopold and Logothetis, 1999).

The same path dependence appears in reasoning more generally. A large body of evidence documents *primacy effects*—the tendency for information encountered early to weigh more heavily than information encountered later. In a classic study, Asch (1946) read subjects a list of personality traits and found that they formed a more favorable impression when the list ran from positive traits to negative than when the same traits were read in reverse; Anderson (1965) likewise found a trait's influence on impressions declining with its position in the sequence. A Bayesian would not expect order to matter at all—the same evidence yields the same posterior however it arrives—yet it plainly does.

Path dependence of this kind arises naturally within our framework. We illustrate with a simple example, and draw out its 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. Table 5 reports, for various m, the fit of "good egg" when evaluated against two good characteristics and against two bad ones. By symmetry, swapping "good" for "bad" throughout leaves the fits unchanged—so the table gives the fit of "bad egg" as well.

	Fit("good egg" | 2 good)	Fit("good egg" | 2 bad)
$m = 5$	0.81	-2.00
$m = 6$	0.65	-1.60
$m = 7$	0.53	-1.33
$m = 8$	0.45	-1.15
$m = 9$	0.38	-1.01
$m = 10$	0.33	-0.91

Table 5: Fit of "good egg" against two good or two bad characteristics

The table shows that when $m \leq 7$, Mr. Judge can make either of two extrapolations, depending on which characteristics he considers. Considering two good characteristics, he can extrapolate to "good egg"; considering two bad ones, he can extrapolate to "bad egg." Both extrapolations clear his threshold.

Path Dependence

Does Mr. Judge's eventual verdict depend on whether he attends to the good characteristics first or the bad? As the model stands, it does not—nothing stops him from reaching the conclusion "good egg" at time 1 by focusing on the good characteristics, and "bad egg" at time 2 by focusing on the bad.

To generate genuine path dependence, we need one further ingredient: an assumption about the *salience* of features. Cognitive psychologists observe that once a person has settled on an interpretation, it becomes hard to set aside (see Leopold and Logothetis, 1999)—it is difficult, for instance, to keep "rabbit" out of mind once one has seen the rabbit-duck figure as a rabbit. We formalize this idea below and show that it generates path dependence.

The idea, informally, is this. Suppose "good egg" and "bad egg," once learned, are salient—hard to keep out of mind. If Mr. Judge reaches "good egg" at time 1, it stays with him at time 2; and with "good egg" in mind, he can no longer extrapolate to "bad egg," since the two are contradictory. The politician's bad characteristics are then brushed aside as the small foibles of an otherwise good egg. We now make this precise.

4.2 Salience and Path Dependence

To incorporate salience into the model, we proceed in three steps: we first define the *power* of a feature, then relate power to salience, and finally show that, under the resulting assumption, an agent's thinking can exhibit path dependence.

Power

The *power* of a feature f on a region R of the picture measures how much f narrows the possibilities on R. To formalize this, call any nonempty set S of pixel locations a *region*.

Two pictures are *equivalent on* R if they agree at every pixel in R; this partitions $\mathcal{P}$ into equivalence classes, which we call the *subpictures of* $\mathcal{P}$ *on* R and denote $\mathcal{R}$. Given a set of features F, let $\delta_R(F) \subseteq \mathcal{R}$ be the subpictures on R consistent with each feature in F—those $P_R \in \mathcal{R}$ for which $P_R \cap f$ is nonempty for each $f \in F$.

Definition 3. *The* power *of a set of features F on region R is*

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

So $\Gamma_R(F) \approx 0$ when F rules out little on R, and $\Gamma_R(F) = 1$ when F pins down R completely—only one assignment on R remains.[29]

Application to Mr. Judge

Return to the example, and take R to be the whole picture. The power of "good egg" is

$$1 - \frac{\log\big(\binom{m}{2} + m + 1\big)}{\log 2^m},$$

since there are 2^m politician types ex ante and $\binom{m}{2} + m + 1$ types consistent with "good egg." By symmetry, "bad egg" has the same power. The power of k pixel values is

$$1 - \frac{\log 2^{m-k}}{\log 2^m} = \frac{k}{m},$$

since 2^{m-k} types remain once k pixel values are known. Table 6 compares the power of "good egg"—and, by symmetry, "bad egg"—with that of two pixel values and of one. When the politician has few characteristics, a pixel value or two is the more powerful in pinning down his type; but once he has enough characteristics ($m \geq 7$), "good egg"

[29] Equivalently, the power of F is the reduction in entropy on R relative to complete ignorance. Note that $\delta_R(F)$ imposes consistency feature by feature: a subpicture is included if, for each feature in F, it contains some picture to which that feature applies—possibly a different picture for different features. An alternative reading would require a single picture to which all the features apply at once. These two readings coincide whenever the region R is the whole picture—as it is in every power calculation we use, including the Mr. Judge example below—so nothing in our results turns on the choice. We think of our approach as the more natural one: the agent assesses each feature on its own, and so can miss a contradiction that surfaces only when the features are taken together.

becomes more powerful than even two. Intuitively, when there are many characteristics to account for, the compact summary "good egg" narrows down the politician's type more than a couple of individual facts about him do.

m	Power of "good egg"	Power of two pixels	Power of one pixel
5	0.20	0.40	0.20
6	0.26	0.33	0.17
7	0.31	0.29	0.14
8	0.35	0.25	0.13
9	0.39	0.22	0.11
10	0.42	0.20	0.10

Table 6: Power of "good egg" versus one or two pixels

Power and Salience

We assume the agent tends to draw on features that are powerful with respect to the conjecture being evaluated (ϕ_t)—more precisely, powerful on its *support* (defined below). Powerful features of this kind are, in this sense, *salient*: they command a place in the agent's thinking.

The support of a conjecture is, informally, the set of pixels it concerns. We define it formally as follows.

Definition 4. *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.*

For instance, the conjecture "$p_{11} = 10$" has only pixel p_{11} in its support, while "$p_{11} + p_{12} = 10$" has both p_{11} and p_{12}.

We can now state our assumption that powerful features on the support of ϕ_t are drawn into 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 $g_1, \ldots, g_n$ ($n > 1$) that taken together are weakly less powerful than f on the support of ϕ_t.*

The idea is that the agent prioritizes powerful features on the conjecture's support, and among equally powerful ones, those that take up less processing capacity. The assumption thus lets the agent replace a set of features with an equivalent but simpler "chunked" one: in the game-board example, an agent who has concluded that "the left-hand side is an H" draws on that feature in place of the pixels that compose it, economizing on processing capacity.[30]

Establishing Path Dependence

Let us now show that Mr. Judge's thinking exhibits path dependence when $m = 7$.

Suppose he reaches "good egg" at time 1. Could he reach "bad egg" at time 2? Its support is all of P, and on that support "good egg"—which, at $m = 7$, is more powerful than one or two pixel values—is forced into his thinking by Assumption 1: he cannot draw on the bad pixel values, whether one or two, without also drawing on "good egg." But "good egg" contradicts "bad egg," so the latter is rejected. Having concluded "good egg," then, Mr. Judge cannot go on to "bad egg"—and, by symmetry, having concluded "bad egg" he cannot go on to "good egg." His verdict depends on where he begins.

This establishes, by example, the following proposition.

Proposition 6. *Under Assumption 1, the agent's thinking may exhibit "path dependence." That is, for some (α, L, P, C, K_0), there exist features f and f' and finite thought sequences (i) $\theta_0, ..., \theta_t$ and (ii) $\theta'_0, ..., \theta'_t$ such that:*

- *Under sequence (i), $f \in K_\tau$ and $f' \notin K_\tau$ for all $\tau > t$, under any continuation of the thought sequence,*
- *Under sequence (ii), $f' \in K_\tau$ and $f \notin K_\tau$ for all $\tau > t$, under any continuation of the thought sequence.*

[30] They may still draw on the unchunked pixels, but only if they also draw on the chunked feature.

4.3 Implications for Persuasion

Path dependence has implications for how one agent can influence another—implications that reach beyond the Bayesian-persuasion framework, in which a principal sways an agent only by choosing what information to disclose (e.g. Kamenica and Gentzkow, 2011).

The first is immediate: *the order in which information is presented matters*. Because our framework makes sense of primacy effects, it points to a simple persuasive strategy—manipulating the order in which information reaches the agent. A politician, for instance, will want to display his strengths early, before opponents can air his flaws, so that a favorable view—"good egg"—takes hold and hardens against later attacks.

Our framework also makes sense of *anchoring effects*, which can themselves be harnessed for persuasion. Anchoring effects are the well-documented pull of an initial value on subsequent judgments, remarkably robust across purchase decisions, legal sentencing, real-estate appraisal, and negotiation (Furnham and Boo, 2011; Tversky and Kahneman, 1974).[31] A persuader can exploit them: a seller names a high price before negotiating, or a prosecutor proposes a long sentence before the judge deliberates, each setting a starting point that pulls the eventual judgment toward it. In our framework, think of an anchor as an initial conjecture; which one the agent starts from can decide where they end up. If a persuader gets Mr. Judge to entertain "good egg" rather than "bad egg" at the outset, path dependence makes that starting point decisive: once "good egg" takes hold it cannot be dislodged.

A further channel is to suggest a *narrative* to the agent—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: in other words, the persuader proposes a whole thought, $\theta_t = (\phi_t, D_t)$. This is why "controlling the narrative"—being first to set

[31] In the classic demonstration, subjects who first considered whether the share of African nations in the UN was above or below a randomly chosen number gave estimates biased toward that number.

it—is treated as so valuable in politics: the opening frame shapes how later information is read (Entman, 1993; Lakoff, 2014). It is also why a campaign may release damaging information about itself—a tactic known as "stealing thunder" (Arpan and Roskos-Ewoldsen, 2005)—to deny opponents the chance to break the story and frame it first.

A narrative can be supplied, however, only if the agent can grasp it: it must be something they can conceptualize. In practice, this means it must be *simple*. Effective political messages are famously compact—as the consultant Frank Luntz puts it, the most memorable political language is rarely longer than a sentence.[32] 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."[33]

5 Discussion

The framework we have developed admits a variety of extensions. We sketch three here: how an agent might resolve inconsistencies in their thinking, and how it can lead them to cycle; how the agent's concepts might themselves be selected rather than taken as given; and how those concepts might be replaced by a richer notion of an internal code.

5.1 Inconsistencies and the Deletion of Features

In the model, an agent can come to hold mutually inconsistent beliefs—formally, there may be no picture consistent with every feature they have accepted: $\bigcap_{f \in K_t} f = \emptyset$. Take Mr. Judge with $m = 5$: he can extrapolate to "good egg" at time 1 from the good characteristics and then to "bad egg" at time 2 from the bad, leaving him holding both at once.

It is natural to suppose the agent would try to resolve such inconsistencies. One way they could do so is by *deleting* features as well as adding them. Suppose, at each time t,

[32] Luntz (2007), p. 7.

[33] A modern restatement is Brandolini's law: "The amount of energy needed to refute bullshit is an order of magnitude bigger than to produce it."

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 ($\phi_t \in K_t \setminus K_0$).[34] A new conjecture is accepted—added to K_t—if its fit is above α ($\Psi(\phi_t; D_t) \geq \alpha$); an existing conclusion is discarded—deleted from K_t—if its fit is below β ($\Psi(\phi_t; D_t) \leq \beta$), where $\alpha > \max\{\beta, 0\}$.[35]

Eliminating Inconsistencies

Whether the agent can rid themselves of an inconsistency turns on how many features are involved. Call a set of k features, F, 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" are an inconsistency of order two. By contrast, "$A > 0$," "$B > 0$," and "$A + B = 0$" are order three: any two statements are consistent, but the three together are not.

If the agent has an inconsistency of order $k \leq L + 1$, there is a thought that eliminates it. However, for $k > L + 1$, there are inconsistencies that may persist in knowledge. 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 just 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 in the agent's thinking.

Cycling

Once features can be deleted, an agent's thinking may cycle as well as exhibit path dependence. To see how, 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; by Table 5, its

[34] Think of the agent's initial knowledge K_0 as axiomatic and not subject to reevaluation.
[35] The deletion threshold β may be positive or negative.

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." Thus, he oscillates between "good egg" and "bad egg," never settling.

Whether this happens turns on β. When $\beta < -1.33$, "good egg" and "bad egg" are each stable once reached, and the model delivers path dependence as before; when $\beta > -1.33$, neither is stable and Mr. Judge cycles.

Cycling of this kind is well documented in perception: viewing an ambiguous figure such as the rabbit–duck illusion for an extended period, subjects often report that perception alternates between the two interpretations (Leopold and Logothetis, 1999). Analogous patterns appear in judgment and decision-making—Rothman et al. (2017), for example, find that people often alternate between opposing positions rather than holding a stable view, a pattern they call "vacillation."

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.

5.2 Concept Selection

So far we have treated the agent's conceptualization C as exogenous. Yet C is plainly endogenous—an expert chess player holds concepts, such as common board configurations, that a novice lacks—so an account of reasoning limits should say why some features come to be conceptualized and others not.

One approach—which we call *concept selection*—is to suppose that the agent can hold up to Λ *intermediate concepts*, and selects those that let them solve as many problems as possible.[36] We assume a concept comes paired with its complement at no extra cost—to

[36]By *intermediate concepts* we mean 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) of a picture and the agent's initial knowledge of it. Holding a conceptualization C, the agent *solves* a problem if they can reach a decision-relevant feature from K_0 within their processing capacity; C is *optimal* for the class if none solves more problems.

grasp "rabbit" is to grasp "not a rabbit."

While we do not consider it realistic that the agent literally selects concepts optimally—this optimization would itself be hard for a boundedly rational agent—the approach still has purchase: we think of C as evolving toward useful concepts, so that characterizing the optimum sheds light on the conceptualization an agent actually comes to hold.

Ms. Lodge

A simple example shows how useful concepts can be found—and connects the problem to a question studied in computer science. Ms. Lodge is deciding whether to rent a house. She observes four binary features: whether it is urban (p_1), close to cafes (p_2), has a garden (p_3), and has a garage (p_4). A garage is essential; beyond that, she wants the house either to be urban and close to cafes, or suburban with a garden. Renting yields utility 1 if these requirements are met and -1 otherwise; not renting yields 0. She has processing capacity $L = 2$, may hold $\Lambda = 2$ intermediate concepts, and faces the class of all pictures with $p_i \in \{0, 1\}$.

Without intermediate concepts, she cannot always decide. Write $F(p_1, p_2, p_3, p_4)$ for the Boolean function that equals 1 exactly when renting is optimal, so that the decision-relevant feature f_1 ("renting is optimal") applies to P if and only if $F(P) = 1$. Here

$$F \;=\; p_4 \wedge \big((p_1 \wedge p_2) \vee (\neg p_1 \wedge p_3)\big),$$

where $\wedge$, $\vee$, and $\neg$ denote "and," "or," and "not." A house she should pass on she can always recognize: a missing garage settles it by itself, as does a failed location pair. But a house worth renting she can never certify, for at such a house F turns on three pixels at once, and no two of them determine it.

Intermediate concepts close the gap, and the Boolean function shows which ones will do. F decomposes as $F = p_4 \wedge (G_1 \vee G_2)$, with $G_1 = p_1 \wedge p_2$ and $G_2 = \neg p_1 \wedge p_3$. Each piece corresponds to an intermediate feature,

$$g_1 = \{Q : q_1 = q_2 = 1\}, \qquad g_2 = \{Q : q_1 = 0,\ q_3 = 1\},$$

a "good urban location" and a "good suburban location." Each depends on only two pixels, so each is deducible within her capacity; and with them she can settle every house. If g_1 or g_2 applies, she draws on it together with the garage (p_4) and concludes; if neither applies—drawing on the complements $\bar{g}_1$ and $\bar{g}_2$, which cost her nothing extra—she deduces f_0, that renting is not optimal.

The decomposition of F is what did the work, and the move is general. Any binary choice problem—rent or not—induces a Boolean function of the pixels that encodes which action is optimal in which state, and choosing good intermediate concepts amounts to *decomposing* that function into pieces each simple enough to fall within the agent's capacity. This is a problem computer scientists have studied deeply: the decomposition of Boolean functions into simpler components is the central concern of circuit design (Wegener, 1987). With more than two actions the relevant object is no longer a Boolean function but a multi-valued one, and the same decomposition problem has been taken up for these as well (Sasao, 1999). We do not pursue the connection further here, but it supplies both a disciplined way to think about conceptualization under constraints and a body of tools for identifying which intermediate concepts are effective.

5.3 From Concepts to Codes

So far we have assumed that a feature is either a concept—in which case the agent can think about it—or it is not. A richer approach lets the agent *encode* features more or less efficiently, with the efficiency of the encoding governing how readily they can bring a feature to mind. Moving from concepts to codes gives a finer-grained picture of how an agent models their environment.

The idea that cognition operates over internal codes has deep roots. In neuroscience, a large literature studies how populations of neurons encode stimuli, with sparse, distributed codes a leading account of how sensory systems represent high-dimensional environments efficiently (Barlow, 2001; Olshausen and Field, 1996). In artificial intelligence, much recent work seeks *disentangled* representations—codes whose separate dimensions track separate factors of variation in the data (Bengio et al., 2013).

We can capture encoding as follows. Suppose the agent has a *code* $\mathcal{C} : B \to \mathcal{F}$ that maps finite-length binary strings to features. A feature f is represented by any string b with $\mathcal{C}(b) = f$. When the agent draws on facts $D_t = \{f_1, \ldots, f_n\}$, they are really drawing on a set of encodings $\{b_1, \ldots, b_n\}$ of those facts, and the demand on processing capacity is their total length:

$$\sum_{i=1}^{n} \ell(b_i) \;\leq\; L,$$

where $\ell(\cdot)$ denotes the length of a string and L is the agent's processing capacity, as before.

The problem of concept selection now becomes one of *code selection*: which string(s) to assign to each feature. The concept limit reappears—not imposed by fiat, as with the bound Λ above, but built into the code itself. Short strings are scarce, so any code can assign short codewords to only so many features—and only those can be drawn on within L. Which features get short codewords is thus the coding counterpart of which concepts the agent selects.

When the picture-space has regularities, the agent can use a *compositional code*—in which codewords are built from reusable parts—at no cost in codeword length. Suppose, for instance, that features f_1, f_2, f_3 are represented by the strings

$$b_1 : \underbrace{11}_{\text{square}}\ \underbrace{000}_{\text{position 1}}\ \underbrace{10}_{\text{size 1}}, \qquad b_2 : \underbrace{11}_{\text{square}}\ \underbrace{111}_{\text{position 2}}\ \underbrace{01}_{\text{size 2}}, \qquad b_3 : \underbrace{00}_{\text{circle}}\ \underbrace{000}_{\text{position 1}}\ \underbrace{01}_{\text{size 2}}.$$

Each string is assembled from a shape, a position, and a size. A handful of such parts, recombined, can represent a vast range of features, so the agent codes a large world with a small vocabulary. Building features this way also changes how the agent thinks—leading them to reason not with whole features (a particular square, of a particular size, in a particular place) but with objects (squares, circles) and operators (position, size, rotation). The analogy is to modular design in computer science, where complex systems are built by reusing well-defined components (Gamma et al., 1994).

Codes also formalize what it means for the agent to have a *model*—and some models are more efficient than others. A feature f may be represented by more than one string: if $\mathcal{C}(b) = \mathcal{C}(b') = f$, the strings b and b' are two models of the same feature, differing in how they describe it, and in length. This is the idea behind the *minimum description length* principle: among models consistent with the data, prefer the one that describes it most compactly (Grünwald, 2007; Rissanen, 1978).

What makes one model more compact is often that it builds in an assumption about how the data were generated. Take the Necker cube (Figure 6). Encoding each line's position and orientation is costly; describing it as a cube in a particular three-dimensional orientation, projected onto the plane, is far shorter—and cheaper precisely because it assumes a generative process the data do not require, since the figure could equally be a flat arrangement of lines.

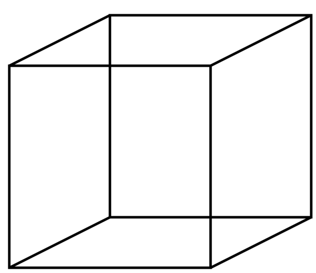

Figure 6: The Necker cube

Causal modeling works the same way. Faced with co-varying data, the agent encodes it most compactly by treating some variables as exogenous and others as caused by them—in effect assigning a direction of causation, in the sense of Pearl's *do*-operator (Eliaz and Spiegler, 2020; Pearl, 2009). As with the cube, this imposes structure the data do not pin down: many causal orderings fit the same correlations. In both cases the agent reaches for the same device—inferring a process that could have generated the data—

because a world with hidden structure is cheaper to encode than one without. The pull toward a data-generating process is, in the end, a pull toward 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, the tax rules, the figures that bear on whether to refinance or when to realize a gain—may still fail to see what those facts jointly imply. The difficulty lies not in any single fact but in putting them together. That integrating information is itself hard—seeing how the parts fit together is distinct from, and often harder than, grasping the parts—is a foundational idea in cognitive psychology and neuroscience. It is also a central concern of computer science and information theory. Economics, by contrast, 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 is that agents have limited *processing capacity*: they can hold only a few facts in mind and combine them at once. We show how this constraint makes sense of a range of behavior, from choice overload and satisficing to why early information and well-timed narratives shape the conclusions people reach.

We also study what happens when agents *extrapolate*—accepting conclusions that fit the facts well enough rather than only those the facts entail. Extrapolation lets an agent reach past what limited processing capacity would otherwise allow, and predict where the facts run out. But the same license to go beyond the facts is what makes their conclusions systematically, and predictably, biased.

The framework leaves much open. One question is how the contents of an agent's thoughts come to be determined—which facts they draw on, and which conjectures they entertain first. We have deliberately remained agnostic, imposing only the assumption of Section 4.2 that powerful features command a place in the agent's thinking. But psychology suggests where fuller answers lie, in the standard division of attention into a *goal-driven* system, directed by the task at hand, and a *stimulus-driven* system, captured by what stands out (Corbetta and Shulman, 2002; Kahneman, 1973).[37] Both systems have natural counterparts here. On the goal-driven side, the agent's grasp of the problem can steer their thoughts toward what is useful for solving it: Mr. Green, having seen that the minimum tax is what makes timing matter, knows where to point his thinking. Such thoughts about how to direct thought are a form of *metacognition* (Ackerman and Thompson, 2017; Flavell, 1979), and modeling them within the framework—the agent reasoning about its own reasoning, under the same capacity constraint—strikes us as a promising direction. Rational inattention (Sims, 2003) and the directed cognition of Gabaix et al. (2006) can both be read as accounts of this goal-driven side. On the stimulus-driven side, our salience assumption is one formalization among several the framework admits: the prominence-, contrast-, and surprise-based salience of BGS could equally govern which facts are drawn into a thought.

A second question concerns the *speed* of reasoning. Our focus has largely been on what an agent can ultimately conclude. Yet reasoning takes 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.

[37] This attentional sense of goal- and stimulus-driven is distinct from the inferential top-down and bottom-up of Section 3.5—reasoning from pixels toward wholes, or from wholes toward pixels—though the two are related.

Appendix

A.1 Proofs

Proof of Proposition 1. Part 1(i). Since $L \geq s(K_0)$, pick a minimal specifying subset $F^* \subseteq K_0$: $|F^*| = s(K_0) \leq L$ and $\bigcap_{f \in F^*} f = \{P\}$. For any $\phi \in C(P)$, the thought (ϕ, F^*) is feasible ($|F^*| \leq L$) and deducible ($\bigcap_{f \in F^*} f = \{P\} \subseteq \phi$, as $P \in \phi$). Hence $R_1 = C(P)$.

Part 1(ii). With $L = 1$ each thought draws on at most one fact. An empty draw gives $\bigcap = \mathcal{P}$ and deduces only $\mathcal{P}$, not a feature; a draw $\{g\}$ deduces ϕ_t only if $g \subseteq \phi_t$. So each new concept is a superset of some feature already known, and by induction a superset of some $g \in K_0$—that is, derivative.

Part 2. Fix $L > 1$, let $N = L(L+1) + 1$, and let P be the all-zero $1 \times N$ picture over $\Omega = \{0, 1\}$ with $K_0 = F_{\text{pixels}}(P)$; then $s(K_0) = N > L$, so $1 < L < s(K_0)$. Let C contain every feature depending on at most $L + 1$ pixels, together with $f_{\text{full}} = \{P\}$.

(i). Until f_{full} is deduced, every concept the agent can draw on depends on at most $L + 1$ pixels, so any L of them jointly constrain at most $L(L+1) < N$ pixels; their intersection leaves a pixel free and cannot equal $\{P\}$. Thus f_{full} is never reachable, and $R \subset C(P)$. Yet R exceeds K_0—e.g. $\{Q : q_1 = q_2 = 0\}$, not a pixel value, is deduced from p_1, p_2—so $K_0 \subset R \subset C(P)$.

(ii). Let $g = \{Q : q_1 = \cdots = q_{L+1} = 0\}$. In one step the agent draws only on pixel values, and L of them leave one of $q_1, \ldots, q_{L+1}$ free, so $g \notin R_1$. But $g \in R_2$: deduce $h = \{Q : q_1 = \cdots = q_L = 0\}$ from $p_1, \ldots, p_L$, then g from h and p_{L+1}. Hence $R_1 \subset R_2$. □

Proof of Proposition 2. Feasibility is monotone in the conceptualization: any thought sequence feasible under C is feasible under any $C' \supset C$. This gives the weak relations $k(f, C') \leq k(f, C)$ (part 1) and $R(C') \cap C \supseteq R(C)$ (part 2). We show each can be strict; in the two examples below, pixels are binary, $K_0 = F_{\text{pixels}}(P)$, and f_S denotes the feature that the pixels in S are white.

Part 1. Let $L = 3$ and P the all-white 5×1 picture, with $C = F_{\text{pixels}} \cup \{f_{12}, f_{34}, f_{12345}\}$ and $C' = C \cup \{f_{123}\}$. Under C' the agent reaches $f_{12345} = \{P\}$ in two steps: f_{123} from p_1, p_2, p_3, then f_{12345} from f_{123}, p_4, p_5. Under C it needs three: one step yields at most one of f_{12}, f_{34}, which together with two pixels pins only four of the five, so f_{12345} requires both f_{12} and f_{34} and a third step combining them with p_5. Thus $k(f_{12345}, C) = 3 > 2 = k(f_{12345}, C')$.

Part 2. Let $L = 2$ and P the all-white 4×1 picture, with $C = F_{\text{pixels}} \cup \{f_{12}, f_{1234}\}$ and $C' = C \cup \{f_{34}\}$. Under C', $f_{1234} = \{P\}$ follows from f_{12} and f_{34}. Under C the only nonpixel intermediate is f_{12}, which with one further pixel pins only three of the four, so f_{1234} is unreachable. Hence $f_{1234} \in (R(C') \cap C) \setminus R(C)$.

Part 3. As K_0 specifies P, $s(K_0) < \infty$: pick $g_1, \ldots, g_n \in K_0$ with $\bigcap_i g_i = \{P\}$ and set $h_i = g_1 \cap \cdots \cap g_i$. Each h_i applies to P, hence is a concept by hypothesis; since $L \geq 2$, the agent deduces h_i from h_{i-1} and g_i, reaching $h_n = \{P\}$. Any $f \in C(P)$ then follows from $\{P\}$ alone. Hence $R(C) = C(P)$. □

Throughout the proofs of the Section 3 propositions, probabilities refer to the uniform distribution over pictures, under which the ratings v_{xy} are i.i.d. Uniform$\{1, \ldots, V\}$; we write $U, U_1, U_2, \ldots$ for generic i.i.d. ratings. Since $\delta(S \cup \{f\}) = \delta(S) \cap f$, we have $|\delta(D_t \cup \{f\})|/|\delta(D_t)| = \Pr(f \mid D_t)$ and $|\delta(\{f\})|/|\mathcal{P}| = \Pr(f)$, so whenever $\Pr(f \mid D_t) > 0$ and $0 < \Pr(f) < 1$ the fit takes the form

$$\Psi(f; D_t) = 1 - \frac{\log|\delta(D_t)| - \log|\delta(D_t \cup \{f\})|}{\log|\mathcal{P}| - \log|\delta(\{f\})|} = 1 - \frac{-\log \Pr(f \mid D_t)}{-\log \Pr(f)}. \tag{1}$$

Let T denote the sum of m i.i.d. ratings (the value of a random product), $H(v) := \Pr(T \leq v)$ its distribution function, and $w_n := \mathbb{E}\big[H(T)^{n-1}\big]$ the unconditional probability that a given product is weakly preferred to $n-1$ independent rivals. For the attribute route, let $R_0, R_1, \ldots, R_{n-1}$ be i.i.d. sums of $m-1$ ratings (the products' values over the *unattended* attributes) and, for an integer vector $\Delta = (\Delta_1, \ldots, \Delta_{n-1})$, let

$$A_{m-1}(\Delta) \ := \ \Pr\big(R_j - R_0 \leq \Delta_j \text{ for all } j\big),$$

and analogously $A_m(\Delta)$ with sums of m ratings. Throughout, $m \geq 2$ and $n \geq 2$ (the section's standing assumption $L < n \times m$ forces $m \geq 2$ on the attribute route, where $L = n$, and $n \geq 2$ on the product route, where $L = m$), and we use repeatedly that a sum of k ratings attains every integer in $[k, kV]$ with positive probability.

Proof of Proposition 3. Part 1 (attribute route). The consumer draws on the n ratings on the attended attribute y, $D_t = \{v_{1y}, \ldots, v_{ny}\}$; index the rivals by $j = 1, \ldots, n-1$. Since $u(x, Q) = \sum_z v_{xz}$, product i is weakly preferred iff, for every rival j, $\sum_{z \neq y}(v_{jz} - v_{iz}) \leq \Delta^y_{ij} := v_{iy} - v_{jy}$. Conditional on D_t, the ratings on the other $m-1$ attributes are free and independent across products, so with R_0 (resp. R_j) denoting product i's (resp. rival j's) total over the unattended attributes, $\Pr(f_i \mid D_t) = A_{m-1}(\Delta)$ and $\Pr(f_i) = A_m(\mathbf{0})$; by (1),

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

(well defined: $A_{m-1}(\Delta) > 0$ and $0 < A_m(\mathbf{0}) < 1$ by full support). The denominator does not depend on the attended ratings, and the numerator depends on them only through $\Delta = \{\Delta^y_{ij}\}_{j \neq i}$; this proves the first clause.

For strict monotonicity, fix a ratings profile and raise Δ^y_{ij} by one—i.e., lower rival j's attended rating v_{jy} by one, admissible whenever $v_{jy} \geq 2$. Then

$$A_{m-1}(\Delta + e_j) - A_{m-1}(\Delta) = \Pr\big(R_j - R_0 = \Delta^y_{ij} + 1,\ R_l - R_0 \leq \Delta^y_{il}\ \forall\, l \neq j\big) \ > \ 0:$$

writing $v'_{jy} := v_{jy} - 1 \geq 1$ and $M := \max\big(\{0,\ v'_{jy} - v_{iy}\} \cup \{v_{ly} - v_{iy} : l \neq j\}\big)$, the event contains the point $R_0 = (m-1) + M$, $R_j = R_0 + v_{iy} - v'_{jy}$, $R_l = m-1$ $(l \neq j)$, whose coordinates lie in $[m-1, (m-1)V]$—for R_0 because $0 \leq M \leq V-1$ and $m \geq 2$; for R_j because $0 \leq M + v_{iy} - v'_{jy} \leq V-1$, whichever term attains M—and satisfy every constraint (as $R_0 \geq m - 1 + v_{ly} - v_{iy}$). Hence $A_{m-1}(\Delta)$, and with it the fit, is strictly increasing in each Δ_{ij}^{y} over admissible profiles. The "if and only if" clause follows: for fixed α the set of advantage vectors whose fit clears α is an upper set.

Part 2 (product route). (i) The consumer draws on product i's m ratings. Conditional on them, product i is weakly preferred iff each of the $n-1$ independent rivals has value at most $u_i = \sum_y v_{iy}$, so $\Pr(f_i \mid D_t) = H(u_i)^{n-1}$ and $\Pr(f_i) = w_n$; by (1),

$$\Psi(f_i; D_t) = 1 - \frac{(n-1)\left[-\log H(u_i)\right]}{-\log w_n}. \tag{3}$$

This depends on the attended ratings only through u_i, and since H is strictly increasing on $[m, mV]$, the fit is strictly increasing in u_i (the denominator is a positive constant). The threshold characterization follows as in Part 1.

(ii) Now the consumer attends a rival $j \neq i$, whose ratings pin u_j. Conditional on them, f_i holds iff product i's value T_0 satisfies $T_0 \geq u_j$ and every one of the other $n-2$ rivals' values is at most T_0:

$$\Pr\big(f_i \mid D_t\big) = \mathbb{E}\Big[\mathbf{1}\{T_0 \geq u_j\}\, H(T_0)^{\,n-2}\Big] =: J(u_j),$$

which depends on j's ratings only through u_j. Raising u_j by one subtracts $\Pr(T_0 = u_j)\, H(u_j)^{n-2} > 0$ from J, so J is strictly decreasing on $[m, mV]$, and the fit $\Psi = 1 - \big[-\log J(u_j)\big] / \big[-\log w_n\big]$ is strictly decreasing in u_j. (For $n = 2$, $J(u_j) = \Pr(T_0 \geq u_j)$.) $\square$

Lemma 1. *Let T be the sum of m i.i.d. Uniform$\{1, \ldots, V\}$ ratings, $H(v) := \Pr(T \leq v)$, and $\mathcal{U} := H(T) \in (0,1]$, a non-degenerate random variable (since $\Pr(\mathcal{U} = 1) = \Pr(T = mV) = V^{-m} \in (0,1)$). Then $w_n := \mathbb{E}[\mathcal{U}^{n-1}]$ satisfies*

$$-\frac{\log w_n}{n-1} \text{ is strictly decreasing in } n, \qquad n = 2, 3, \ldots$$

Proof. Since $\mathcal{U}$ is not a.s. constant, strict monotonicity of L^p norms (Lyapunov's inequality) gives $w_n^{1/(n-1)} = \|\mathcal{U}\|_{n-1} < \|\mathcal{U}\|_n = w_{n+1}^{1/n}$; taking logs and rearranging yields $-\log(w_n)/(n-1) > -\log(w_{n+1})/n$. $\square$

Proposition 4: Preliminaries

Endow $\mathcal{P}$ with the uniform measure $\mathbb{P}$, so that $\psi(\{g\}) = -\log \mathbb{P}(g)$ and $\psi(D \cup \{g\}) - \psi(D) = -\log \mathbb{P}(g \mid D)$; with $L(p) \equiv -\log p$,

$$\Psi(f_i; D) = 1 - \frac{L(\mathbb{P}(f_i \mid D))}{L(\mathbb{P}(f_i))}.$$

Ratings are i.i.d. uniform on $\{1, \ldots, V\}$, $V \geq 2$. Let $T^{(k)}$ denote a sum of k i.i.d. ratings, D_ℓ i.i.d. copies of the difference of two independent ratings (symmetric, lattice, full support on $\{-(V-1), \ldots, V-1\}$), and $W_k = D_1 + \cdots + D_k$, with $f_k(j) = \mathbb{P}(W_k = j)$, $\rho_k = f_k(0)$,

$\pi_k(c) = \mathbb{P}(|W_k| \le c)$. The characteristic function of D_ℓ is the Fejér kernel

$$\varphi(t) = \left(\tfrac{\sin(Vt/2)}{V\sin(t/2)}\right)^2 \in [0,1], \qquad \rho_k = \int \varphi^k d\mu, \quad \pi_k(c) = \int \varphi^k K_c\, d\mu,$$

where $d\mu = dt/2\pi$ on $[-\pi,\pi]$ and $K_c(t) = \sum_{|j|\le c} e^{ijt} = \sin((c+\frac{1}{2})t)/\sin(t/2)$ is the Dirichlet kernel. Since $\varphi < 1$ a.e., ρ_k is strictly decreasing. For symmetric integer W and integer $c \ge 0$, $\mathbb{P}(W \le c) = \frac{1}{2} + \frac{1}{2}\mathbb{P}(|W| \le c)$ (add $\mathbb{P}(W \le c)$ and $\mathbb{P}(W \ge -c)$).

Lemma 2 (Reduction)**.** *Fix the attended attribute and the advantages $\Delta_j \equiv \Delta^y_{ij}$, $j \ne i$. Then*

$$p_c(m) \equiv \mathbb{P}(f_i \mid D) = \mathbb{P}\big(S_j - S_i \le \Delta_j\ \forall j \ne i\big), \qquad p_u(m) \equiv \mathbb{P}(f_i) = \mathbb{P}\big(T_j \le T_i\ \forall j \ne i\big),$$

with $S_1,\ldots,S_n$ i.i.d. $T^{(m-1)}$ and $T_1,\ldots,T_n$ i.i.d. $T^{(m)}$. Along the product route with $u_i < mV$ fixed,

$$\mathbb{P}(f_i \mid D) = q^{n-1}, \quad q = \mathbb{P}(T^{(m)} \le u_i) \in (0,1), \qquad \mathbb{P}(f_i) = \mathbb{E}\big[F(T)^{n-1}\big],$$

where $T \sim T^{(m)}$ with c.d.f. F.

Proof. Condition on D. On the attribute route the unfixed ratings are i.i.d. uniform, $u(x,Q) = v_{xy} + S_x$, and f_i holds iff $S_j - S_i \le v_{iy} - v_{jy} = \Delta_j$ for all j. On the product route the rivals' values are i.i.d. $T^{(m)}$ independent of the fixed u_i; $q < 1$ since $u_i < mV$, $q > 0$ since $u_i \ge m$. □

Lemma 3 (Anticoncentration)**.** *There is $C_V < \infty$ with $\sup_x \mathbb{P}(W_k = x) \le \rho_k \le C_V/\sqrt{k}$ for all $k \ge 1$.*

Proof. $\sup_x f_k(x) \le \frac{1}{2\pi}\int |\varphi|^k dt = \rho_k$ because $\varphi \ge 0$. Since D_ℓ is nondegenerate on the span-1 lattice, $\varphi(t) < 1$ for $0 < |t| \le \pi$ and $1 - \varphi(t) \ge c_V t^2$ near 0; hence $\varphi(t) \le e^{-c'_V t^2}$ on $[-\pi,\pi]$ for some $c'_V > 0$, and $\rho_k \le \int e^{-c'_V k t^2} dt/2\pi = C_V/\sqrt{k}$. □

The general-n half of Part 1 below rests on classical local expansions for the lattice sum $T^{(k)}$; we collect them now, in Lemmas 4–6. Let $q_k(x) = \mathbb{P}(T^{(k)} = x)$ and $G_k(x) = \mathbb{P}(T^{(k)} \le x)$ be the p.m.f. and c.d.f. of a sum of k ratings; let $\mu = (V+1)/2$ and $\sigma^2 = (V^2-1)/12$ be the mean and variance of one rating, and set

$$z_k(x) = \frac{x - k\mu}{\sigma\sqrt{k}}, \qquad h_k = \frac{1}{\sigma\sqrt{k}},$$

the standardized coordinate and the lattice mesh. Write ϕ, Φ for the standard normal density and c.d.f. (the straight ϕ; the Fejér kernel of the Preliminaries is the curled φ), and define the *midpoint* c.d.f.

$$\bar{G}_k(x) \;:=\; \mathbb{P}\big(T^{(k)} < x\big) + \tfrac{1}{2}q_k(x), \qquad \text{so that} \qquad G_k = \bar{G}_k + \tfrac{1}{2}q_k \text{ exactly.}$$

A rating is symmetric about μ, so $T^{(k)} - k\mu$ is a symmetric lattice variable with real characteristic function $r(t)^k$, $r(t) = \sin(Vt/2)/(V\sin(t/2))$; as in Lemma 3, $|r(t)| \le e^{-c_V t^2}$ on $[-\pi, \pi]$. Finally put

$$I_{r,s} \;:=\; \int_{-\infty}^{\infty} \phi(z)^r\,\Phi(z)^s\,dz \;\in (0,\infty), \qquad r \ge 1,\ s \ge 0.$$

Below, $O(\cdot)$ is uniform in x, and all displayed constants depend only on (n, V, Δ).

Lemma 4 (Local expansion)**.** *Uniformly in $x \in \mathbb{Z}$, with $z = z_k(x)$,*

$$q_k(x) \;=\; h_k\Big[\phi(z) + \frac{\eta(z)}{k}\Big] + O(k^{-5/2}), \qquad \eta(z) = \frac{\kappa_4}{24\sigma^4}\,(z^4 - 6z^2 + 3)\,\phi(z),$$

where κ_4 is the fourth cumulant of a rating (so κ_4/σ^4 is its excess kurtosis). There is no $k^{-1/2}$ term: odd cumulants vanish by symmetry. Moreover $\sup_x q_k(x) \le Ch_k$ (the inversion bound, as in Lemma 3), and by Hoeffding's inequality for bounded summands, $\mathbb{P}\big(|T^{(k)} - k\mu| \ge \lambda\sigma\sqrt{k}\big) \le 2e^{-c\lambda^2}$ for all $\lambda > 0$.

Proof. Fourier inversion gives $q_k(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-i(x-k\mu)t}\,r(t)^k\,dt$ with r real and even. Split at $|t| \le k^{-2/5}$: outside, $|r|^k \le e^{-c_V k^{1/5}}$; inside, expand $\log r(t) = -\frac{\sigma^2 t^2}{2} + \frac{\kappa_4 t^4}{24} + O(t^6)$ (odd terms vanish by symmetry), so $r(t)^k = e^{-k\sigma^2 t^2/2}\big(1 + \frac{k\kappa_4 t^4}{24} + O(kt^6 + k^2 t^8)\big)$, and the Gaussian integrals produce the display. (This is the standard span-one lattice local expansion; see Petrov (1975, Ch. VII) for expansions to arbitrary order. After the substitution $s = \sigma\sqrt{k}\,t$ the neglected terms are $O(s^6/k^2) + O(s^8/k^2)$ against a Gaussian weight, giving the stated $O(k^{-5/2})$; outside the central region $|r|^k \le e^{-c_V k^{1/5}}$.) The sup bound is the inversion bound $\sup_x q_k(x) \le \frac{1}{2\pi}\int_{-\pi}^{\pi} |r|^k \le Ch_k$, and the tail bound is Hoeffding's inequality for bounded summands. ☐

Lemma 5 (Midpoint c.d.f.)**.** *Uniformly in $x \in \mathbb{Z}$, with $z = z_k(x)$,*

$$\bar{G}_k(x) \;=\; \Phi(z) + \frac{\vartheta(z)}{k} + O(k^{-2}\log k),$$

with ϑ bounded and vanishing at $\pm\infty$. There is no $k^{-1/2}$ term—the half-cell correction of the Riemann sum is cancelled exactly by the subtracted half-atom—and no $k^{-3/2}$ term, since Euler–Maclaurin corrections come in even powers of h_k and the expansion of q_k carries no odd terms.

Proof. $\bar{G}_k(x) = \sum_{y \le x} q_k(y) - \frac{1}{2}q_k(x)$. Substituting Lemma 4 and applying the Euler–Maclaurin (midpoint) formula to the smooth summands, whose derivatives are integrable,

$$h_k \sum_{y \le x} \phi(z_k(y)) = \int_{-\infty}^{z + h_k/2} \phi + O(h_k^2) = \Phi(z) + \tfrac{1}{2}h_k\phi(z) + O(h_k^2),$$

with the $O(h_k^2)$ terms carrying bounded coefficient functions of z; the correction term η

receives the same midpoint treatment,

$$h_k k^{-1} \sum_{y \le x} \eta(z_k(y)) - \tfrac{1}{2} h_k k^{-1} \eta(z) = k^{-1} \int_{-\infty}^{z} \eta + O\big(k^{-1} h_k^2\big),$$

so it contributes at orders k^{-1} and k^{-2} only—no h_k^3 term; and the accumulated local errors are $O(\sqrt{k} \log k \cdot k^{-5/2}) + O(e^{-c \log^2 k}) = O(k^{-2} \log k)$ after restricting to $|z_k(y)| \le \log k$ (Hoeffding outside). The subtracted half-atom $\frac{1}{2} q_k(x) = \frac{1}{2} h_k \phi(z) + \frac{1}{2} h_k k^{-1} \eta(z) + O(k^{-5/2})$ supplies exactly the two half-cell terms netted above. What remains is $\Phi(z)$ plus terms in $h_k^2 = 1/(\sigma^2 k)$ and h_k^4 with bounded coefficient functions (the η-integral and the even Euler–Maclaurin corrections), i.e. $\Phi(z) + \vartheta(z)/k + O(k^{-2} \log k)$. $\square$

Lemma 6 (Lattice sums)**.** *Fix $r \ge 1$, $s \ge 0$, and integer shifts $d_1 = 0, d_2, \ldots, d_r$ and $e_1, \ldots, e_s$. Then*

$$\Sigma_k \ := \ \sum_{x \in \mathbb{Z}} \prod_{i=1}^{r} q_k(x + d_i) \prod_{l=1}^{s} \bar{G}_k(x + e_l) \ = \ h_k^{r-1} \Big[I_{r,s} + \gamma_1 h_k + \gamma_2 h_k^2 + \gamma_3 h_k^3 + o(h_k^3) \Big]$$

for constants $\gamma_1, \gamma_2, \gamma_3$ depending on (r, s, V) and the shifts, with

$$\gamma_1 \ = \ \int_{-\infty}^{\infty} \Big[\sum_{i=1}^{r} d_i \, \phi'(z) \phi(z)^{r-1} \Phi(z)^s + \sum_{l=1}^{s} e_l \, \phi(z)^{r+1} \Phi(z)^{s-1} \Big] dz.$$

In particular, the leading coefficient $I_{r,s}$ never depends on the shifts, and $\gamma_1 = 0$ when all shifts vanish.

Proof. Restrict to $|z_k(x)| \le \log k$; by the tail bounds of Lemma 4 (and $r \ge 1$), the discarded sum is $O(e^{-c \log^2 k})$. On the retained range substitute Lemmas 4–5 and Taylor-expand each shift, $\phi(z + d_i h_k) = \phi(z) + d_i h_k \phi'(z) + \frac{1}{2} d_i^2 h_k^2 \phi''(z) + O(h_k^3)$ (to third order) and likewise $\Phi(z + e_l h_k)$: the summand becomes

$$h_k^r \Big[F_0(z) + h_k F_1(z) + h_k^2 F_2(z) + h_k^3 F_3(z) + o(h_k^3) \Big], \qquad F_0 = \phi^r \Phi^s,$$

where F_1 (whose integral is γ_1) collects the first-order shift terms displayed above, and F_2, F_3 collect higher-order shift terms together with the η/k and ϑ/k corrections ($k^{-1} = \sigma^2 h_k^2$) and their cross terms; each F_j is smooth with Gaussian decay (η, and ϑ with its derivatives, decay at a Gaussian rate by construction—for ϑ this uses $\int_{\mathbb{R}} \eta = 0$—so the shifted expansions are uniform on the retained range). Finally, for such F the midpoint lattice sum has no algebraic corrections: $h_k \sum_{x \in \mathbb{Z}} F(z_k(x)) = \int F + O(h_k^N)$ for every N, by Poisson summation (the Fourier transform of F decays super-polynomially). Collecting powers of h_k gives the display: each $h_k^{r+j} F_j(z)$ sums to $h_k^{r+j-1} \int F_j$, while the accumulated pointwise errors—at worst $h_k^r \cdot O(k^{-2} \log k)$ per point, over $O(\sqrt{k} \log k)$ points—total $O(h_k^{r+3} \log^2 k) = o(h_k^{r+2})$, as required. $\square$

Proof of Proposition 4. Part 2 (product route: attending product i). By Lemma 2, $\Psi_n = 1 - (n -$

$1)L(q)/g(n-1)$ with $g(k) \equiv -\log \mathbb{E}[X^k]$, $X \equiv F(T) \in (0,1]$, $g(0) = 0$. Note $\mathbb{E}[X^k] < 1$ (since $X < 1$ on $\{T < mV\}$, an event of positive probability), so $g(k) > 0$ for $k \geq 1$, and $L(q) > 0$.

Monotonicity. By Lyapunov's inequality $k \mapsto \log \mathbb{E}[X^k]$ is convex, and strictly so because X is not a.s. constant ($X = 1$ exactly on $\{T = mV\}$, which has probability $V^{-m} \in (0,1)$). Hence g is strictly concave with $g(0) = 0$, so $g(k)/k$ is strictly decreasing, so $(n-1)L(q)/g(n-1)$ is strictly increasing in n: Ψ_n is strictly decreasing.

Divergence and the uniform $\bar{n}$. $\mathbb{E}[X^{n-1}] > \mathbb{P}(T = mV) = V^{-m}$ (values $T < mV$ also contribute) gives $g(n-1) < m \log V$, while $u_i < mV$ gives $q \leq \mathbb{P}(T \leq mV - 1) = 1 - V^{-m}$, hence $L(q) \geq -\log(1 - V^{-m}) > 0$. Therefore, uniformly over $u_i < mV$,

$$\Psi_n \ < \ 1 - \frac{(n-1)\big(-\log(1 - V^{-m})\big)}{m \log V} \ \longrightarrow \ -\infty,$$

and $\Psi_n < \alpha$ for every $n \ \geq \ \bar{n} \equiv \max\Big\{2, \ 1 + \Big\lceil \frac{(1-\alpha)\, m \log V}{-\log(1-V^{-m})} \Big\rceil \Big\}$, which depends only on (α, V, m): for $n \geq \bar{n}$, $(n-1)\big(-\log(1 - V^{-m})\big) \geq (1-\alpha)\, m \log V$, so the strict display gives $\Psi_n < \alpha$. (At $\alpha = 1$ this is immediate: $L(q) > 0$ and $g(n-1) > 0$ give $\Psi_n < 1$ for every $n \geq 2$.)

Part 2 (product route: attending a rival). Proposition 3, Part 2(ii), allows a second product-route extrapolation: attend a rival $j \neq i$, of value u_j, and conclude "product i is weakly preferred." Conditional on the attended ratings, f_i holds iff product i's value T is at least u_j and at least each of the other $n-2$ rivals' values, so, as in Lemma 2,

$$\mathbb{P}(f_i \mid D) = \mathbb{E}\big[\mathbf{1}\{T \geq u_j\}\, F(T)^{n-2}\big] \ \leq \ \mathbb{E}\big[X^{n-2}\big], \qquad X = F(T),$$

while $\mathbb{P}(f_i) = \mathbb{E}[X^{n-1}]$ as before. Hence, uniformly over the rival's value u_j,

$$\Psi \ \leq \ 1 - \frac{g(n-2)}{g(n-1)} \ = \ \frac{g(n-1) - g(n-2)}{g(n-1)} \ =: \ \varepsilon_n,$$

with equality at $u_j = m$. Now g is strictly increasing (since $\mathbb{E}[X^k]$ is strictly decreasing) and strictly concave with $g(0) = 0$, so the increments $g(n-1) - g(n-2)$ strictly decrease in n while the denominator strictly increases: ε_n is strictly decreasing. Since g is bounded above by $m \log V$, its (positive, decreasing) increments tend to zero, and $\varepsilon_n \to 0$. In the other direction, $\mathbf{1}\{T \geq u_j\}\, F(T)^{n-2} \geq \mathbf{1}\{T = mV\}$ gives $\mathbb{P}(f_i \mid D) \geq V^{-m}$, so $L(\mathbb{P}(f_i \mid D)) \leq m \log V$; and $\mathbb{E}[X^{n-1}] \downarrow \mathbb{P}(X = 1) = V^{-m}$ by monotone convergence, so $g(n-1) \uparrow m \log V$ and

$$\Psi \ \geq \ 1 - \frac{m \log V}{g(n-1)} \ \longrightarrow \ 0.$$

Thus, uniformly over the rival's value u_j, the fit tends to zero: this route closes by convergence to zero, as on the attribute route, rather than by divergence, as on the own-product route—attending a rival leaves open that product i itself is maximal, so the conditional information gain stays bounded by $m \log V$. Finally, $\bar{n}' \equiv \min\{n \geq 2 : \varepsilon_n < \alpha\}$ is finite, depends only on (α, V, m), and for every $n \geq \bar{n}'$ no attended rival supports the conclusion, whatever the rival's value. (For $n = 2$ the upper bound is vacuous, $\varepsilon_2 = 1$, and rightly so: attending a rival of minimal value $u_j = m$ makes "product i is weakly preferred" deducible.)

Part 1 (attribute route: the limit). Fix the advantages $(\Delta_j)_{j \neq i}$, each positive by hypothesis (only $\Delta_j \geq 0$ is used in this part). We show $p_c(m) \to 1/n$ and $p_u(m) \to 1/n$; since L is continuous at $1/n$ with $L(1/n) = \log n > 0$, this gives $\Psi_m \to 0$.

For p_u: by exchangeability, $\mathbb{P}(T_i > T_j\ \forall j) = \frac{1}{n}\,\mathbb{P}(\text{the max is unique})$, and $0 \leq p_u(m) - \mathbb{P}(T_i > T_j\ \forall j) \leq \sum_{j \neq i} \mathbb{P}(T_i = T_j) \leq (n-1)\rho_m$, while $1 - \mathbb{P}(\text{max unique}) \leq \binom{n}{2}\rho_m$ (each tie event $T_x = T_{x'}$ has probability $\sup_z \mathbb{P}(W_m = z) \leq \rho_m$ by Lemma 3). Hence $|p_u(m) - 1/n| \leq 2n\rho_m \to 0$.

For p_c: with $S_j - S_i \overset{d}{=} W_{m-1}$,

$$0 \;\leq\; p_c(m) - \mathbb{P}\big(S_j \leq S_i\ \forall j\big) \;\leq\; \sum_{j \neq i} \mathbb{P}\big(0 < S_j - S_i \leq \Delta_j\big) \;\leq\; (n-1)(V-1)\,\rho_{m-1},$$

again by Lemma 3 (each of the $\Delta_j \leq V - 1$ lattice points carries mass at most ρ_{m-1}), and $|\mathbb{P}(S_j \leq S_i\,\forall j) - 1/n| \leq 2n\rho_{m-1}$ as above. Hence $|p_c(m) - 1/n| \leq \big[(n-1)(V-1) + 2n\big]\rho_{m-1} \to 0$.

Part 1 (attribute route: the uniform $\bar{m}$). By Proposition 3, the attribute-route fit is increasing in each advantage; hence for every configuration of advantages and every product i, the fit is at most the fit at the maximal configuration $\Delta_j = V - 1$ for all j, which we denote $\Psi_m^{\max}$. By the limit just proved, $\Psi_m^{\max} \to 0 < \alpha$, so $\bar{m} \equiv \min\{M : \sup_{m \geq M} \Psi_m^{\max} < \alpha\}$ is finite and depends only on (α, V, n). For $m \geq \bar{m}$, no attended attribute supports the conclusion for any product, whatever the advantages.

Part 1 (attribute route: strict decrease in m, case $n = 2$). Here the advantage Δ is a positive integer, so $\Delta \geq 1$; $p_c(m) = \frac{1}{2}(1 + \pi_{m-1}(\Delta))$, $p_u(m) = \frac{1}{2}(1 + \rho_m)$, and $\Psi_m = 1 - \zeta_m/\xi_m$ with $\zeta_m = \log 2 - \log(1 + \pi_{m-1}(\Delta))$, $\xi_m = \log 2 - \log(1 + \rho_m)$.

Step 1 (increment reduction). $\pi_{m-1}(\Delta) \geq \pi_{m-1}(0) = \rho_{m-1} > \rho_m$ gives $\zeta_m < \xi_m$; and $\zeta_m > 0$ for $m \geq 3$, since then W_{m-1} can exceed $\Delta \leq V - 1$. If additionally

$$\zeta_{m+1} - \zeta_m \;\geq\; \xi_{m+1} - \xi_m, \tag{$\star$}$$

then with $\delta = \xi_{m+1} - \xi_m > 0$, $\frac{\zeta_{m+1}}{\xi_{m+1}} \geq \frac{\zeta_m + \delta}{\xi_m + \delta} > \frac{\zeta_m}{\xi_m}$, i.e. $\Psi_{m+1} < \Psi_m$. Unwinding logarithms and using $(1 + \pi_m(\Delta))/(1 + \rho_{m+1}) \leq 2$, ($\star$) holds whenever

$$A_m \;\equiv\; \int \varphi^{m-1}(1 - \varphi)\,\big(K_\Delta - 2\varphi\big)\,d\mu \;\geq\; 0. \tag{$\star'$}$$

Step 2 (central region). Since $K_\Delta(t) \geq (2\Delta + 1) - \frac{t^2}{2}\sum_{|j| \leq \Delta} j^2 \geq (2\Delta + 1)\big(1 - \frac{(\Delta+1)^2 t^2}{6}\big)$ and $2\Delta + 1 \geq 3$, there is $t_\Delta \geq \sqrt{2}/(\Delta + 1) \geq \sqrt{2}/V$ with $K_\Delta(t) \geq 2$ for $|t| \leq t_\Delta$. On this region $K_\Delta - 2\varphi \geq 2(1-\varphi) \geq 0$, so

$$\int_{|t|\leq t_\Delta} \varphi^{m-1}(1-\varphi)(K_\Delta - 2\varphi)\,d\mu \;\geq\; 2\int_{|t|\leq t_\Delta} \varphi^{m-1}(1-\varphi)^2\,d\mu \;\geq\; c_\star(V)\,m^{-5/2}$$

for all $m \geq 2$, using $1-\varphi(t) \asymp t^2$ and $\varphi(t) \geq e^{-c_V'' t^2}$ near 0 (a Laplace-type lower bound with constants depending only on V).

Step 3 (outer region). Since $1-\cos\theta \;=\; 2\sin^2(\theta/2) \;\geq\; 2\theta^2/\pi^2$ on $[-\pi,\pi]$ (Jordan's inequality) and $\mathbb{P}(|D_\ell| = 1) = 2(V-1)/V^2$,

$$1-\varphi(t) \;=\; \mathbb{E}\big[1-\cos(tD_\ell)\big] \;\geq\; \frac{2(V-1)}{V^2}\cdot\frac{2t^2}{\pi^2}, \qquad |t| \leq \pi,$$

so on the outer region $t_\Delta \leq |t| \leq \pi$, using $t_\Delta \geq \sqrt{2}/V$,

$$\varphi(t) \;\leq\; 1-\frac{4(V-1)}{\pi^2V^2}\,t_\Delta^2 \;\leq\; 1-\frac{8(V-1)}{\pi^2V^4} \;=:\; X^*(V) \;<\; 1,$$

while $|K_\Delta(t) - 2\varphi(t)| \leq 1/|\sin(t/2)| + 2 \leq \pi/t_\Delta + 2 \leq C_3V$, since $|K_\Delta| \leq 1/|\sin(t/2)|$, $\varphi \in [0,1]$, and $|\sin(t/2)| = \sin(|t|/2) \geq |t|/\pi \geq t_\Delta/\pi$ on the outer region, by the same Jordan bound as above. Hence the outer contribution to A_m is bounded in absolute value by $C_3V\,(X^*)^{m-2}\int\varphi(1-\varphi)\,d\mu \leq C_4(V)\,(X^*)^m$.

Conclusion. $A_m \geq c_\star(V)m^{-5/2} - C_4(V)(X^*)^m > 0$ for all $m \geq M(V,\Delta)$, with M explicit from the constants. By Step 1, Ψ_m is strictly decreasing for $m \geq M$.

Part 1 (attribute route: strict decrease in m, general n). For general n we do not generalize the increment inequality $(\star)$; instead we expand the fit itself in powers of $m^{-1/2}$, to a precision that survives differencing in m. Conditioning on the attended product's unattended total reduces everything to one-dimensional lattice sums, for which the required expansions are the local expansions of Lemmas 4–6 above.

Expansion of the route probabilities. Conditioning on the attended product's total and using the independence of the rivals (Lemma 2),

$$p_u(m) = \sum_x q_m(x)\,G_m(x)^{\,n-1}, \qquad p_c(m) = \sum_x q_{m-1}(x)\prod_{j=1}^{n-1} G_{m-1}(x+\Delta_j).$$

Insert $G_k = \bar{G}_k + \frac{1}{2}q_k$ and expand the product over subsets $A \subseteq \{1,\ldots,n-1\}$ of "half-atom" factors:

$$p_u(m) = \sum_{\ell=0}^{n-1}\binom{n-1}{\ell}2^{-\ell}\sum_x q_m(x)^{\ell+1}\bar{G}_m(x)^{\,n-1-\ell},$$
$$p_c(m) = \sum_A 2^{-|A|}\sum_x q_{m-1}(x)\prod_{j\in A} q_{m-1}(x+\Delta_j)\prod_{j\notin A}\bar{G}_{m-1}(x+\Delta_j).$$

Every term is a lattice sum of the form in Lemma 6 with $r = \ell+1$ (resp. $r = 1+|A|$) atoms and prefactor h^{r-1}; terms with five or more atoms are $O(h^4)$ and can be discarded.

For $p_u(m)$ all shifts vanish. The $\ell = 0$ term is $I_{1,n-1} + \gamma_2 h_m^2 + \gamma_3 h_m^3 + o(h_m^3)$ with $I_{1,n-1} = \int \phi\, \Phi^{n-1} = 1/n$ and, crucially, $\gamma_1 = 0$; the $\ell = 1$ term is $\frac{n-1}{2} h_m \big[I_{2,n-2} + O(h_m)\big]$ with its own expansion through h_m^3; the $\ell = 2, 3$ terms contribute expansions beginning at h_m^2, h_m^3. Hence

$$p_u(m) \;=\; \frac{1}{n} + \frac{n-1}{2}\, I_{2,n-2}\, h_m + \beta_2^u h_m^2 + \beta_3^u h_m^3 + o(m^{-3/2}).$$

For $p_c(m)$, at $k = m-1$: the $A = \emptyset$ term is $r = 1$ with c.d.f. shifts $e_l = \Delta_l$, so its γ_1 equals $\sum_j \Delta_j \int \phi^2 \Phi^{n-2} = \sum_j \Delta_j\, I_{2,n-2}$—this is where the advantages enter; the $|A| = 1$ terms give $\frac{n-1}{2} I_{2,n-2} h_k$ plus higher orders as before, and $|A| \geq 2$ terms begin at h_k^2. Hence

$$p_c(m) \;=\; \frac{1}{n} + \Big(\sum_j \Delta_j + \frac{n-1}{2}\Big) I_{2,n-2}\, h_{m-1} + \beta_2^c h_{m-1}^2 + \beta_3^c h_{m-1}^3 + o(m^{-3/2}),$$

and since $h_{m-1} = h_m + \frac{\sigma^2}{2} h_m^3 + O(h_m^5)$, the same form holds in powers of h_m with an adjusted β_3^c.

Conclusion: differencing. Set $x_m := n\, p_c(m) - 1$ and $y_m := n\, p_u(m) - 1$, so that

$$x_m = a_1 h_m + a_2 h_m^2 + a_3 h_m^3 + o(m^{-3/2}), \qquad y_m = b_1 h_m + b_2 h_m^2 + b_3 h_m^3 + o(m^{-3/2}),$$

with

$$a_1 - b_1 \;=\; n\, I_{2,n-2} \sum_{j \neq i} \Delta_j \;>\; 0.$$

By the fit formula of the Preliminaries,

$$\Psi_m = \frac{\log(1+x_m) - \log(1+y_m)}{\log n - \log(1+y_m)}.$$

Substituting the expansions of x_m, y_m and expanding $\log(1+\cdot)$ and the quotient to third order in $h_m = 1/(\sigma\sqrt{m})$ yields constants c_2, c_3 with

$$\Psi_m \;=\; c_1 m^{-1/2} + c_2 m^{-1} + c_3 m^{-3/2} + e_m, \qquad c_1 = \frac{n\, I_{2,n-2} \sum_{j\neq i} \Delta_j}{\sigma \log n} \;>\; 0,$$

where the remainder satisfies $e_m = o(m^{-3/2})$. Differencing kills every term below the leading one: $m^{-1/2} - (m+1)^{-1/2} = \frac{1}{2} m^{-3/2} + O(m^{-5/2})$, while $m^{-1} - (m+1)^{-1} = O(m^{-2})$ and $m^{-3/2} - (m+1)^{-3/2} = O(m^{-5/2})$. No cancellation between e_m and e_{m+1} is needed: since $e_{m+1} = o\big((m+1)^{-3/2}\big) = o(m^{-3/2})$, the triangle inequality gives $|e_m - e_{m+1}| \leq |e_m| + |e_{m+1}| = o(m^{-3/2})$, which the leading difference $\frac{c_1}{2}\, m^{-3/2}$ dominates. Hence

$$\Psi_m - \Psi_{m+1} \;=\; \frac{c_1}{2}\, m^{-3/2} + o(m^{-3/2}) \;>\; 0 \qquad \text{for all } m \geq M(n, V, \Delta):$$

the fit is strictly decreasing in m once m is large enough. Since the positive advantage vectors lie in the finite set $\{1, \ldots, V-1\}^{n-1}$, M may be taken to depend only on (n, V).

Closing. Fix $\alpha \in (0, 1]$. Part 1 supplies $\bar{m}(\alpha, V, n)$ beyond which no attended attribute supports "product i is weakly preferred" for any i and any advantages. Part 2 and its rival variant supply $\bar{n}(\alpha, V, m)$ and $\bar{n}'(\alpha, V, m)$ beyond which no attended product supports a conclusion—neither choosing itself, at any value short of the maximum, nor choosing product i off a rival's ratings, at any value at all. Hence any problem with $m \geq \bar{m}(\alpha, V, n)$ and $n \geq \max\{\bar{n}, \bar{n}'\}(\alpha, V, m)$ closes both routes of Proposition 3 outright. □

Proof of Proposition 5. By hypothesis, the admissible thoughts are the routes of Proposition 3, whose fits—the F_m, G_m, H_m of the statement—are, by that proposition, strictly increasing (F_m, G_m) or strictly decreasing (H_m) in their single statistics.

Part (i). Since F_m is strictly increasing, the best attribute route for product k attends an attribute achieving the largest lead $\Delta_k = \max_y(v_{ky} - v_{-k,y})$, with fit $F_m(\Delta_k)$—independent of the products' values, which affect neither F_m nor Δ_k. The best route overall gives fit $B_k := \max\{F_m(\Delta_k), G_m(u_k), H_m(u_{-k})\}$.

Part (ii). The consumer can conclude "product 1 is weakly preferred" but not "product 2 is weakly preferred" iff $B_2 < \alpha \leq B_1$. It remains to show that on (B_2, ∞) the constraint $\alpha \leq B_1$ reduces to $\alpha \leq F_m(\Delta_1)$. Suppose $\alpha > B_2$. Since $u_1 < u_2$,

$$G_m(u_1) < G_m(u_2) \leq B_2 < \alpha, \qquad H_m(u_2) < H_m(u_1) \leq B_2 < \alpha :$$

product 1's value routes are unavailable, so $\alpha \leq B_1$ iff $\alpha \leq F_m(\Delta_1)$, and the interval is exactly $\big(\max\{F_m(\Delta_2), G_m(u_2), H_m(u_1)\},\ F_m(\Delta_1)\big]$, as stated. It is nonempty iff $F_m(\Delta_1)$ strictly exceeds each of $F_m(\Delta_2)$, $G_m(u_2)$, and $H_m(u_1)$; since F_m is strictly increasing on the admissible range (which contains $\Delta_1, \Delta_2 \geq 1$), the first condition is $\Delta_1 > \Delta_2$. This gives exactly the three stated conditions.

Finally, the claim in the footnote following the proposition: whenever the interval in (ii) is nonempty, the inferior product's best route is its attribute route, since its value routes $G_m(u_1)$ and $H_m(u_2)$ lie strictly below $B_2 < F_m(\Delta_1)$. □

Proof of Proposition 6. A proof by example is provided in the main text (pages 44-44). □

References

Abiteboul, Serge, Richard Hull, and Victor Vianu, *Foundations of databases*, Vol. 8, Addison-Wesley Reading, 1995.

Ackerman, Rakefet and Valerie A. Thompson, "Meta-Reasoning: Monitoring and Control of Thinking and Reasoning," *Trends in Cognitive Sciences*, 2017, *21* (8), 607–617.

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.

Anderson, Norman H, "Primacy effects in personality impression formation using a generalized order effect paradigm.," *Journal of personality and social psychology*, 1965, *2* (1), 1.

Aragones, Enriqueta, Itzhak Gilboa, Andrew Postlewaite, and David Schmeidler, "Fact-free learning," *American Economic Review*, 2005, *95* (5), 1355–1368.

Arpan, Laura M and David R Roskos-Ewoldsen, "Stealing thunder: Analysis of the effects of proactive disclosure of crisis information," *Public Relations Review*, 2005, *31* (3), 425–433.

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.

Barlow, Horace, "Redundancy Reduction Revisited," *Network: Computation in Neural Systems*, 2001, *12* (3), 241–253.

Baumeister, RF, 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.

Blake, Randolph and Nikos K. Logothetis, "Visual Competition," *Nature Reviews Neuroscience*, 2002, *3* (1), 13–21.

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.

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.

Corbetta, Maurizio and Gordon L. Shulman, "Control of Goal-Directed and Stimulus-Driven Attention in the Brain," *Nature Reviews Neuroscience*, 2002, *3* (3), 201–215.

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.

Feldman, Jacob, "Minimization of Boolean complexity in human concept learning," *Nature*, 2000, *407* (6804), 630–633.

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.

— **, David Laibson, Guillermo Moloche, and Steven Weinberg**, "Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model," *American Economic Review*, 2006, *96* (4), 1043–1068.

Gamma, Erich, Richard Helm, Ralph Johnson, and John Vlissides, *Design Patterns: Elements of Reusable Object-Oriented Software*, Addison-Wesley, 1994.

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.

Grünwald, Peter D, *The minimum description length principle*, MIT press, 2007.

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, 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.

Lakoff, George, *The all new don't think of an elephant!: Know your values and frame the debate*, Chelsea Green Publishing, 2014.

Leopold, David A and Nikos K Logothetis, "Multistable phenomena: changing views in perception," *Trends in cognitive sciences*, 1999, *3* (7), 254–264.

Luntz, Frank, *Words That Work: It's Not What You Say, It's What People Hear*, Hachette UK, 2007.

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.

Marr, David, *Vision: A computational investigation into the human representation and processing of visual information*, MIT press, 2010.

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.

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.

Oaksford, Mike and Nick Chater, *Bayesian rationality: The probabilistic approach to human reasoning*, Oxford University Press, 2007.

Olshausen, Bruno A. and David J. Field, "Emergence of Simple-Cell Receptive Field Properties by Learning a Sparse Code for Natural Images," *Nature*, 1996, *381* (6583), 607–609.

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.

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.

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 and Adi Sunderam, "Using models to persuade," *American Economic Review*, 2021, *111* (1), 276–323.

Shafir, Eldar, Itamar Simonson, and Amos Tversky, "Reason-based choice," *Cognition*, 1993, *49* (1-2), 11–36.

Shiller, Robert J, "Narrative economics," *American Economic Review*, 2017, *107* (4), 967–1004.

Simon, Herbert A, "A behavioral model of rational choice," *Quarterly Journal of Economics*, 1955, pp. 99–118.

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.

Tenenbaum, Joshua B, Thomas L Griffiths, and Charles Kemp, "Theory-based Bayesian models of inductive learning and reasoning," *Trends in cognitive sciences*, 2006, *10* (7), 309–318.

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 tomorrow™: Using behavioral economics to increase employee saving," *Journal of Political Economy*, 2004, *112* (S1), S164–S187.

Tversky, Amos, "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.

Wegener, Ingo, *The complexity of Boolean functions*, John Wiley & Sons, Inc., 1987.

Wertheimer, Max, "Experimentelle Studien über das Sehen von Bewegung," *Zeitschrift für Psychologie*, 1912, *61* (1), 161–265.

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.