A Cognitive Theory of Reasoning and Choice

Pedro Bordalo, Nicola Gennaioli, Giacomo Lanzani, Andrei Shleifer*

October 21, 2024

Abstract

We offer a theory of decisions in which selective attention to the features of the current problem is determined by its categorization in a set of problems the DM solved in the past. Categorization depends on goal-relevant features of the current and past problems, as well as on normatively irrelevant context, such as other goods in the choice set. The model delivers systematic heterogeneity in attention and choice based on past experiences, which persists despite identical current conditions, and discontinuous shifts when a bottom-up salient change in problem features or in their description causes re-categorization. The model accounts for major puzzles and framing effects in riskless choice, statistical problems, and lottery choice based on heterogenous and unstable mental representations.

*The authors are from Oxford Said Business School, Universita Bocconi, Harvard University, and Harvard University, respectively. We thank Daniel Bartels, Ben Enke, Alex Imas, Lawrence Jin, Spencer Kwon, Shengwu Li, Sendhil Mullainathan, Joshua Schwartzstein, Jesse Shapiro, Tomasz Strzalecki, and Richard Thaler for helpful comments. Steven Ma provided outstanding research assistance. Gennaioli thanks the Italian Ministry of Education for Financial Support (PRIN Grant).

1 Introduction

How do people select which features to attend to when making decisions? Is Donald Trump a hardened criminal or the champion of ordinary Americans? Is the car rental agency Avis, which since 1962 advertised itself with the slogan "We are number two – we try harder" an underdog or a loser? Is buying a stock a long-term investment or a risk? For the last century, economists have traced choice to stable preferences and correct beliefs, but it feels intuitive that in these and other instances many people have similar payoffs or information, and yet make sharply different assessments because they attend to some features of the problem, neglecting others. When a neglected feature becomes prominent due to a salient event, political propaganda, advertising, or social influence, many people refocus on that feature and change decisions.

A similar intuition emerges from work on judgment and decision-making. People disagree even if there is one correct answer and change choices after irrelevant reframing (e.g., Tversky and Kahneman [118], Thaler [109], Enke and Zimmermann [35], Graeber [51]). The behavioral economics of stable biases, such as reference points, self-serving beliefs, social preferences, base rate neglect, or under-reaction explains neither such disagreement nor instability. To understand decisions as well as mechanisms of influence – private (advertising) and public (information campaigns) – we must engage with a cognitive mechanism: selective attention to features.

We offer a theory in which attention is driven by a key new step: problem recognition. Before evaluating choice options, people use attention and memory to categorize a problem in a class of similar problems they solved in the past. This step produces a "mental representation": the set of features deemed relevant for solving the problem. Because multiple representations are often plausible, choice heterogeneity and instability follow.

In the duck-rabbit illusion, people focus on the beak and recognize a duck, or on the mouth and recognize a rabbit. Both representations are possible, because each selects i) what to attend to and ii) a confirming category from memory. Representation-dilemmas also arise in economic problems. Choosing a jam to buy may be categorized as "giving myself a treat," which focuses on consumption pleasure, or as "buying a staple food," which focuses on price. Assessing a stock investment may be categorized as "evaluating returns", which focuses on welfare under different wealth states, or as "evaluating risks", which focuses on the events in which such states materialize.

In standard theory, these categories are integrated. Our key idea is that integration is difficult because different categories are segregated in memory.[1] We experience the pleasure of the jam at home and the pain of paying at the supermarket, not together. We experience wealth as we consume it, and think about possible risks when imagining the future, not together. Because these experiences are associated with different contexts, they produce competing categories that utilize subsets of features.

In our model, categorization follows the regularities of human memory: it is stochastic and more likely to retrieve problems that have been more frequently solved in the past or are more similar along features that are attended to. Some such features are goal relevant in the current problem, others are goal relevant in other problems, and still others are purely contextual. Whether the jam is categorized as a treat or a staple depends on the prices and qualities on offer, but also on whether the day of the week, other goods on the shelf, etc., match past experiences of treats versus staples, commanding attention.

We obtain three key results. First, experiences create systematic heterogeneity in how a problem is represented and solved. Critically, and unlike in

[1]See Treisman and Gelade [113] for the seminal study showing that integration of features is significantly more difficult and time-consuming than considering each of the features separately, and Tversky [122] for difficulties in aggregation of spatial knowledge.

Bayesian learning, different representations focus on different features, causing excess persistence in disagreement, as in the confirmation bias (Nickerson [86]) or cognitive dissonance (Festinger [37]). Being familiar with the pain of paying, a formerly poor person repeatedly focuses more on the price of the jam than a person with the same income but without the same experiences. Heterogeneity is in the representations of choice, not in its hedonic consequences.

Second, we obtain instability due to changes in context drawing attention to a subset of experiences. A special occasion such as a festivity can cue a poor person to exceptionally focus on consumption pleasure and neglect price. In mental accounting (Thaler [109]), a "rainy day" account can induce thrift by serving as a reminder of financial shocks. Unlike in models of noisy perception (Woodford [125], Enke and Graeber [34]), instability reflects re-categorization, which produces large shifts associated with context. Choice instability also comes from changes in representations, not from information.

Third, top-down attention interacts with bottom-up salience (e.g., Bordalo, Gennaioli, and Shleifer [17], Bushong, Camerer, and Rangel [21]). A bottom-up salient feature is more heavily weighted in decisions, but can also trigger a category for which this feature is relevant, causing neglect of previously attended to features. An advertisement showing the jam at a beautiful breakfast table prompts retrieval of the "treat" category, boosting quality focus and dampening price focus. This mechanism explains uninformative persuasion (e.g., Mullainathan, Schwartzstein, and Shleifer, [83] Mullainathan and Shleifer [84]).

We develop our model using famous puzzles in three domains: riskless choice, statistical problems, and risky choice. Category-driven attention delivers mental accounting (e.g., sunk cost fallacy and opportunity cost neglect) in riskless choice, the Gambler's Fallacy and over vs. under reaction in statistical problems, the common ratio effect, the certainty effect, and the fourfold pattern in risky choice. It also accounts for the dumbfounding link between the fourfold pattern in risky choice and riskless mirrors (Oprea [88]). This link cannot come from risk preferences. It comes from similarity of representations across domains.

Our theory of mental representations offers the first mechanism accounting for biases and framing effects within and across domains based on contextual similarity and experiences with different problems. It also offers a recipe for how to produce "bias" or "rationality": draw attention to, or away from, a feature that is relevant in a similar problem, not in the current one.

We contribute to an extentive literature in decision making. Feature selection connects to bounded rationality (Simon [103]), but not in the sense of computational complexity. Categorization can cause a confident oversimplification or overcomplication of computations, due to a focus on the wrong features (Hascher, Imas, Ungeheuer, and Weber [56]).

Traditional behavioral theories, which we discuss in the paper, typically assume stable and domain specific biases. These theories cannot explain framing and the often weak within-person correlation of choices in a domain, e.g., demand for different insurance products and lotteries in the lab (Barseghyan, Prince, and Teitelbaum [4]). Across domains, loss aversion should drive the endowment effect and risk aversion when facing mixed lotteries, but these behaviors are empirically disconnected (Chapman, Dean, Ortoleva, Snowberg, and Camerer [24]). Recent work detects correlations between choices, within and across domains, that cut across standard biases (Stango and Zinman [107], Chapman, Dean, Ortoleva, Snowberg, and Camerer [24], de Clippel, Oprea, and Rozen [30]). Our theory suggests that the driver of this architecture is similarity or differences in representations, within and across domains. In the conclusions we offer some reflections on possible strategies to unveil it.

We build on the psychology of attention-based similarity perceptions (Nosofsky [87], Tversky [115],[116]), and of top-down attention, which stresses the role of experiences (Itti and Baldi [62], Awh, Belopolsky, and Theeuwes [2]). Economic research studies top-down attention shaped by priors (e.g., Schwartzstein [99], Gagnon-Bartsch, Rabin, and Schwartzstein [43]) and by goal optimality (e.g., Sims [104], Gabaix [41]). In our paper, attention is driven by memory, so it changes with cues, consistent with the evidence in Conlon ([28]).

Our focus on categories follows a large literature in psychology (e.g., Mack and Palmeri [81], Reed [91], Rosch and Lloyd [93]). Mullainathan [82] offers an economic model in which the category prototype pins down beliefs.[2] The role of experiences follows case-based learning (Schank [98]), formalized in economics by Gilboa and Schmeidler ([47]), and habit formation (e.g., Laibson [74]). Compared to this work, our categories encode past attention to features, not a specific choice or belief, and are retrieved based on bottom-up salience. This allows for persistent errors and framing effects due to the use of inappropriate categories. In Salant and Rubinstein [95] instability is due to the use of different choice functions for different frames. We offer a cognitive foundation based on context and past experiences.

Recent work on memory studies selective information retrieval about probabilities (e.g., Bordalo, Conlon, Gennaioli, Kwon, and Shleifer [19] and Fudenberg, Lanzani, and Strack [40]) and hedonics of goods (Bordalo, Gennaioli, and Shleifer [16]). In our model probabilities and hedonics are known, but memory shapes attention to them. Both mechanisms naturally shape decisions, and future work may study them together.

Section 2 presents a model of top-down attention in riskless choice, statistical

[2] Ellis and Masatlioglou [32] axiomatize utility that is stable within but not across categories of choices. Evers, Imas, and Kang [36] present and test a model of hedonic editing based on loss and gain categories. In psychology, the ALCOVE model by Kruschke ([73], [116]), and ADDCOVE, by Verguts, Ameel, and Storms [123] formalize the assignment of multidimensional stimuli to categories. The key differences with our approach include, among others, that we endogenize similarity to categories shaped by both top-down and bottom-up forces, and that the result of the categorization is not the mere assignement to a group of similar instances, but a way to evaluate the problem that leads to its (possibly wrong) solution.

problems and risky choice. Section 3 derives its general predictions. Section 4 applies them to mental accounting and judgmend biases. Section 5 introduces bottom-up salience and studies choice under risk. Section 6 concludes.

2 The Model

Even a simple problem such as buying jam requires the evaluation of many aspects (taste, price, "is it spoiled?", etc.). Only some features, however, are attended to in choice. We offer a theory based on two blocks: 1) how attention causes choice, and 2) attention allocation. Our key innovation is that in 2) attention is shaped "top down" by a simplified representation based on a category of experiences. If the DM thinks about consumption, which often occurs removed from paying, she focuses on taste and neglects price. If she thinks about buying, when paying is prominent, she focuses on price. For many goods, risk is neglected: it is not mentioned and rarely materializes. For others, such as flying, it is attended to. So, categories also shape neglect vs alertness to risk.

For step 1) consider a lottery o delivering a good with hedonics (u_{1s}, u_{2s}), such as quality and price, in state $s = e_{1s} \cap e_{2s}$ defined by events e_{1s} and e_{2s} (e.g., "selection of urn A" and "extraction of a green ball from it"). A DM paying attention $\alpha_x \in (0, 1]$ to feature x values o as:

$$\sum_s \mathbb{P}(e_{1s})^{\alpha_{e_1}} \mathbb{P}(e_{2s}|e_{1s})^{\alpha_{e_2}} \cdot (\alpha_{u_1} u_{1s} + \alpha_{u_2} u_{2s}). \tag{1}$$

Lower attention α_x reduces weighting of probabilities or hedonics, distorting valuation.

Riskless choice entails a sure event s and valuation of monetary and non-monetary payoffs. With quality q and price p, Equation (1) yields weighted utility $\alpha_Q q - \alpha_P p$, as in Bordalo, Gennaioli, and Shleifer [15].

A statistical problem entails estimating the probability of event H, to which payoff 1 is attached, with zero payoff outside. If $H =$"green ball g from urn A," Equation (1) becomes $\mathbb{P}(A)^{\alpha_U}\,\mathbb{P}(g|A)^{\alpha_C}$, akin to Grether's [52] formula.

Risky choice entails hedonic and event features. A lottery paying x_g with probability π and $x_b < x_g$ otherwise is valued in Equation (1) as

$$\pi^{\alpha_{e_g}} \cdot \alpha_{u_g} x_g + (1-\pi)^{\alpha_{e_b}} \cdot \alpha_{u_b} x_b,$$

which combines payoff weights in Bordalo, Gennaioli, and Shleifer [14] with probability weights as in Prospect Theory (Kahneman and Tversky [68]).

Categorization allows attention to depend on non payoff-relevant context features (e.g., an irrelevant good in the choice set, location, etc.), which cue different categories. A beer advertisement showing a party with dancing youngsters can cue "consumption" experiences, causing price neglect and affecting choice even though the beer's attributes are unchanged.

Section 2.1 models step 1 in a general setup nesting Equation (1). It also lays out choice features, whose value is good-specific, and context features, which are constant across goods. Section 2.2 links features to categories and models step 2: how features cue a representation, shaping attention and choice.

2.1 How Attention Causes Valuation

A problem's primitives are: i) a menu of options, ii) a set of features, iii) an attention-driven valuation function as in Equation (1), and iv) a decision rule mapping valuations into choice.

Menu of Options. There is a nonempty finite menu of lotteries O. Each lottery $o \in O$ is a finite set of event-payoff combinations, which we call atoms. As in Equation (1), the value of an atom is an attention-based combination of its hedonic and event features and the value of o is the sum of the values of its atoms. Riskless choice and statistical hypotheses are special cases. The valuation of atoms depends on the DM's attention to their features.

Features of Atoms. The features of atom y are collected in the vector

$$y = (u, e).$$

Subvector u reports *hedonic* features, such as a dollar payoff or the jam's quality and price. The value u_i of hedonic $i \in M_H$ is a real number. There is a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ known to the DM. Subvector e reports *event* features: delivery states for hedonics, e.g., $e_i \in \{urn\ A, urn\ B\}$ or $e_i \in \{jam\ spoiled,\ not\}$, but also events like the jam's brand (e.g. Bonne Maman versus others). Each feature $i \in M_E$ identifies a partition of the state space Ω. Event e_i, $i \in M_E$, reports the atom's equivalence class in the partition corresponding to i.

Attention and Valuation. Hedonic and event features can vary across options, so attention to them can shape choice. We thus call them *choice features*, $i \in M_O = M_H \sqcup M_E$. Attention to them is a vector of weights $\alpha_O \in [0,1]^{M_O}$, that do not necessarily add up to 1. Feature $i \in M_O$ is fully weighted if $\alpha_i = 1$, underweighted if $0 < \alpha_i < 1$, edited out if $\alpha_i = 0$. An edited out event is set to $e_i = \Omega$ (any of its realizations is allowed), and an edited out hedonic feature takes the feature's average value across atoms, i.e., $u_i = \overline{u}_i = \frac{\sum_{o \in O} \sum_{y \in O} u_i}{\sum_{o \in O} |o|}$. Let $y(\alpha_O) = (u(\alpha_O), e(\alpha_O))$ be the representation of atom y. It contains the edited version of events $e(\alpha_O)$ and the attention-based perception of hedonics $u(\alpha_O)$, where:

$$u_i(\alpha_O) = \alpha_{O,i} \cdot u_i + (1 - \alpha_{O,i}) \cdot \overline{u}_i. \tag{2}$$

Limited attention to hedonic $i \in M_H$ shrinks perception toward its average value $\overline{u}_i$ (fully so for $\alpha_{O,i} = 0$, in which case i will not impact choice).

The value of $y(\alpha_O)$ multiplies its perceived probability and hedonics:

$$v(y(\alpha_O)) = \left[\prod_{r \in M_E} \mathbb{P}\left(y_r(\alpha_O) \mid \cap_{j<r} y_j(\alpha_O)\right)^{\alpha_r}\right] \cdot \sum_{i \in M_H} u_i(\alpha_O). \tag{3}$$

The first bracket is the perceived probability of $y(\alpha_O)$: the attention-weighted chain product of the probabilities of its events. The product follows a linear order $<$ over event features set by the true sampling process, e.g., first select an urn and then extract a ball from it.[3] The second term is the perceived hedonics in $y(\alpha_O)$. The value of payoff-states in Equation (1) is a special case with two events and zero average hedonics, $\overline{u}_i = 0$.

As in expected utility, the value of lottery $o(\alpha_O) = \{y(\alpha_O) : y \in o\}$ is the sum of the values of its atoms:

$$v(o(\alpha_O)) = \sum_{y(\alpha_O) \in o(\alpha_O)} v(y(\alpha_O)) \qquad \forall o \in O. \tag{4}$$

Full attention to choice features, $\alpha_O = 1$, yields the rational benchmark. A feature is "relevant" if it affects $v(o(\alpha_O))$ when $\alpha_O = 1$. Whether a feature is relevant or not depends on the problem.

Decision Rule. Given a vector of valuations $v \in \mathbb{R}^O$, choice in the admissible set A is set by a decision map $D : \mathbb{R}^O \to A$. For goods or lotteries D picks the highest valued option, in some statistical problems it computes the relative value of two lotteries-hypotheses (yielding their relative probability).

Valuation is shaped by attention to choice features α_O, but the problem is categorized based on context features. These are common across goods and allow the DM to compare the whole problem to problems faced in the past.

[3] The DM can compute the probabilities of the factors in the product. If the sampling process is not specified, $>$ reflects the DM's subjective beliefs. The order is irrelevant if $\alpha_i \in \{0, 1\}$ for $i \in M_E \cup M_L$.

Features of Context. Context is partly derived from choice features. For instance the choice set describes whether the DM must choose a good or estimate different hypotheses, or whether the randomization device is known or unknown. Such context can cue past problems with similar choices and attribute values. But context also includes aspects that are wholly irrelevant for valuation such as time, location, recent events, etc. These features may be spuriously associated with experiences, cueing different problems.

Context features $i \in M_K$ are summarized by a vector $\kappa = (\kappa_u, \kappa_e, z)$. Hedonic context κ_u reports, for each hedonic, its possible values in the choice set (the prices of available jams, their qualities, etc.). Event context κ_e reports, for each event, its realizations in the choice set (safe versus spoiled jam, the brands of the jams, etc.). The *situation* $z \in Z$ reports irrelevant non-good specific dimensions such as time, location, etc. (e.g. whether the jam is bought during a festivity).[4] Each feature $i \in M_K$ has an associated distance d_i that measures the perceived dissimilarity between two possible assumed values.

We denote current context by κ_t and its generic entry by the dated name of the corresponding feature, i_t for $i \in M_K$. A context feature affects categorization only if the DM attends to it. Vector $\alpha_K \in [0,1]^{M_K}$ reports attention to each such feature. Overall attention is $\alpha = (\alpha_O, \alpha_K)$ and the representation of the current problem is (α_t, κ_t).

2.2 Categorization and Attention

The choice features and situation in κ_t cue categorization. The DM's database at time $t \in \mathbb{N}$ is partitioned into a set of categories C. Each category c is summarized by the prototypical attention and context vector (α_c, κ_c) experienced

[4]Some situation features are also derived from hedonics and events, reporting for instance the average price level (expensive versus cheap goods problem) or the average probabilities of specific events (high versus low risk problem).

in the problems belonging to c, and by their temporally discounted frequency $F_c \in \mathbb{R}_+$.[5] Attention to context is binary $\alpha_{c,i} \in \{0,1\}$ for $i \in M_K$ and identifies which features are diagnostic of the category (Rosch and Lloyd [93]). To ease notation, κ_c reports only diagnostic features.

We illustrate this formalization using four categories. The first two capture "riskless" experiences, in which the DM evaluated a good's hedonics in a specific state. The second two categories capture "statistical" experiences, in which she estimated event probabilities. We study our model using only these categories, but in reality many more categories are possible.

Illustrative Categories. The riskless categories are "consuming" and "buying". In the consuming category con, the DM's evaluation of goods focuses on qualities, not prices, as encoded in the attention subvector $\alpha_{con,O} = (\alpha_{con,i})_{i \in M_O}$. The context κ_{con} of these experiences specifies a set of experienced qualities $q \in Q_{con}$ and situations $z \in Z_{con}$ (e.g. being at home, where price is not prominent). Diagnosticity of quality and situation context is reported in $\alpha_{K,con} = (\alpha_{con,i})_{i \in M_K}$ by setting $\alpha_{con,i} = 1$ for $i = Q, Z$, and zero otherwise.

In buying, $c = buy$, attention to choice features $\alpha_{O,buy}$ focuses on the pain of paying, but partly also on the goods' typical quality, else we would not buy. Context reports the experienced qualities $q \in Q_{buy}$, prices $p \in P_{buy}$, and situations $z \in Z_{buy}$ (e.g. being in a shop). These diagnostic features take value 1 in attention to context $\alpha_{buy,K}$.

There are two statistical categories: "simple sampling" and "agnostic inference". Simple sampling, $c = ss$, refers to experiences of estimating the prob-

[5] The category "prototype" can be formalized as having the average attention

$$\alpha_c = \sum_{\tau \in c} \alpha_\tau / |c|$$

and the best compromise context κ_c where for every $i \in M_K$, κ_{ic} minimizes some distance from past contexts $(\kappa_\tau)_{\tau \in c}$. Recency-weighted frequency is $F_c = \sum_{\tau \in c} \delta^{t-\tau}$, $\delta \in (0,1)$.

ability of a single draw from a known process, e.g., that a fair coin lands h or t. Attention $\alpha_{ss,O}$ focuses on the single event corresponding to the name of the hypothesis, h or t. Diagnostic context includes there being a single draw and hypotheses coinciding with its realizations.

Agnostic inference, $c = ai$, refers to experiences of judging a data generating process (DGP) based on one or more i.i.d. signals without having prior information about it, e.g. assessing the quality of a restaurant based on one bad meal. Attention $\alpha_{ai,O}$ focuses on the share of positive signals (a sufficient statistic for the DGP), not on the prior. Diagnostic context includes having at least two draws (selection of the DGP and signals) and hypotheses coinciding with the possible DGPs (e.g. good or bad restaurant).

These categories are building blocks in standard choice theory. They thus yield comparative statics based on routinely measured parameters such as the number of draws, the price paid and its spatio/temporal distance from consumption (what Thaler calls "decoupling").[6] Like in Mullainathan [82], categories are given. In the conclusions we discuss how categories may be measured and also how they may be endogenized.

Categorization and attention are jointly determined based on similarity. A real valued function $S\left[\left(\alpha_t, \kappa_t\right), \left(\alpha_c, \kappa_c\right)\right]$ measures the similarity between the problem (α_t, κ_t) and the prototype (α_c, κ_c) of category c. It decreases if the two disagree in: i) attention (i.e. α_t vs. α_c) and ii) context (κ_t vs. κ_c). We use

[6] One could also consider a category in which the DM "has no clue" and pays attention to nothing, and as a result is indifferent across options. By construction, similarity to this category is constant, so it sets a minimum similarity threshold below which no other category is adopted. Reliance on this category can be traced to reluctance to choose or, when choice is forced, to measured low confidence.

the separable form:

$$S\left[\left(\alpha_t,\kappa_t\right),\left(\alpha_c,\kappa_c\right)\right]=\frac{\sum_{i\in M}\left[1-d\left(\left|\alpha_{t,i}-\alpha_{c,i}\right|\right)\right]+\sum_{i\in M_K}\left[1-\alpha_{t,i}\alpha_{c,i}d_i\left(\kappa_{it},\kappa_{ic}\right)\right]}{|M|+|M_K|}, \tag{5}$$

where $d:\mathbb{R}_+\rightarrow\mathbb{R}$ is increasing, strictly convex and twice continuously differentiable with $d'(0)=0$, $d(1)=1$. The first sum in (5) measures disagreement in attention $\alpha_{t,i}\neq\alpha_{c,i}$, the second in category-diagnostic context.

The DM represents problem κ_t by setting an attention profile that maximizes total (perturbed) similarity with a category in C. This process is formalized in two steps: matching and categorization.

Matching. The DM fits the problem into each $c\in C$ by picking an attention vector $\alpha_t(c)$ that maximizes the total similarity between the problem (α_t,κ_t) and the the members of c, summarized by the prototype (α_c,κ_c) and frequency F_c. The maximum total similarity with $c\in C$ is given by:[7]

$$S(t,c)=\max_{\alpha_t\in[0,1]^{M_O\cup M_K}} F_c\cdot S\left[\left(\alpha_t,\kappa_t\right),\left(\alpha_c,\kappa_c\right)\right]. \tag{6}$$

Categorization. Following model assignment tasks in psychology (Mack and Palmeri [81]), the DM chooses the category $c\in C$ maximizing

$$S(t,c)+\epsilon_c,$$

where ϵ_c is an type I extreme-value random shift with scale parameter λ, reflecting random attention to context.

We are often unaware of this top-down process, as when we first see duck-rabbit and unconsciously select one feature to attend (e.g., the beak) and one category from experience (ducks). Sometimes we are aware of it. In duck-

[7]This yields exactly F_c if $\delta^{t-\tau}$ multiplicatively reduces similarity in Equation ([84]).

rabbit, after being told of the ambiguity of the figure, we see different animals on different occasions, but we never fail to see one, nor do we see both at the same time. In real world choices, we often entertain different perspectives, but then select one, which determines our choice.

3 Top-Down Attention and Choice

We characterize the model's general implications for categorization and choice.

3.1 Attention and Categorization

Let $d_{tc}(i) = d_i(\kappa_{t,i}, \kappa_{c,i})$ be the distance between the problem and the category along context feature i. When matching c, the DM tunes attention to this feature to satisfy, in an interior equilibrium, the first order condition for the maximization of Equation (6):

$$\frac{\partial}{\partial \alpha_{t,i}} d\left(|\alpha_{t,i} - \alpha_{c,i}|\right) + d_{tc}(i) \cdot \mathbb{I}_c(i) = 0, \tag{7}$$

where indicator $\mathbb{I}_c(i)$ is equal to one if i is a diagnostic feature for c and zero otherwise. Adapting attention to the category, reducing $|\alpha_{t,i} - \alpha_{c,i}|$, tends to increase similarity but may backfire along a discrepant diagnostic feature, $d_{tc}(i) \cdot \mathbb{I}_c(i) > 0$. This yields the following result.

Proposition 1 *When matching the problem with category c, attention to a discrepant diagnostic feature $d_{tc}(i) \cdot \mathbb{I}_c(i) > 0$ is shrunk toward zero, while attention fully adapts to the category otherwise, $\alpha_{t,i}(c) = \alpha_{c,i}$.*

To best match the problem with c, the DM's attention aligns as much as possible with α_c but neglects discrepant diagnostic context. When choosing a jam, a DM adapts to "consuming" by focusing on taste and neglecting price.

The DM also neglects the current "supermarket" location, which is diagnostic of "buying". Editing, i.e. full neglect, entails similarity cost $d(|0-1|) = d(1) = 1$.

The problem's similarity to c is thus endogenous, equal to $S(t,c) = F_c(1 - d(t,c))$ where $d(t,c)$ is the minimized distance from c. Categorization works as follows.

Proposition 2 *Categorization in context κ_t satisfies*

$$\Pr(c|t) = \frac{\exp(\lambda \cdot F_c \cdot (1 - d(t,c)))}{\sum_{c' \in C} \exp(\lambda \cdot F_{c'} \cdot (1 - d(t,c')))}. \tag{8}$$

Problem t is more likely to be categorized in c when:

i) category c was used more frequently (and recently), $\partial \Pr(c|t)/\partial F_c > 0$.

ii) category c is less dissimilar to current context κ_t, $\partial \Pr(c|t)/\partial d(t,c) < 0$.

The DM is more likely to rely on a category that she more frequently used in the past, higher F_c. This entails a form of confirmation bias, as we discuss later. Higher context discrepancy $d(t,c)$ reduces similarity for each DM, producing instability. This is a reasonable approach, as relying on frequency and context often promote fitting categories and good decisions. When the DM is highly trained on a specific problem such as the probability that a fair coin lands h, the description context of a fair coin and its h or t outcome fosters perfect recognition, prompting a focus on, and a correct answer of, 50%.

In other cases, frequent use of c or spurious context may block a more fitting c', causing attention distortions and error. Repeated discussions of political corruption in Washington may prompt a honesty judgment category, causing voters to neglect competence. A sporting event or a referendum, contexts where national pride is cued, may temporarily increase purchases of goods associated with the country's flag, causing neglect of typically attended to price or quality (Nardotto and Sequeira, [85]).[8]

[8]Psychologists have documented extensively both excessive reliance on a decision logic/model, which is called "overgeneralization", and the failure to apply a known model

3.2 Choice

As the DM categorizes the problem in c, she adopts the category's attention to choice features $\alpha_{O,c}$. This yields valuation v_{tc} and choice $a_{tc} = D(v_{tc}) \in A$ (the Appendix relaxes injectivity of a_{tc}), giving three properties.

Proposition 3 *Attention and choice are stochastic due to categorization,*

$$\Pr(a_{tc}) = \Pr(\alpha_t = \alpha_t(c)) = \Pr(c|t)$$

for all $a_{tc} \in A$. Furthermore:

i) Higher F_c increases $\Pr(a_{tc})$ and weakly decreases $\Pr(a_{tc'})$ for $c' \neq c$.

ii) Higher $d(t,c)$ decreases $\Pr(a_{tc})$ and weakly increases $\Pr(a_{tc'})$ for $c' \neq c$.

iii) Increasing similarity to c, decreasing $d(t,c)$, boosts $\Pr(a_{tc})$ more at higher F_c if and only if c is not dominant, i.e. $\Pr(c|t) \leq \pi^*$ for $\pi^* > 1/2$.

Attention and choice are stochastic due to shock ϵ_c. Stochastic choice is often explained by noisy perception of hedonics or probabilities (Woodford [125], Enke and Graeber [34]). In that approach, noisy signals are used in a Bayesian way, so the DM's valuation for an action is distributed around a single mode. Categorization can instead yield multi-modal valuation, with discontinuous shifts across categories. Multi-modality is observed in statistical problems, where the correct answer is continuous in the problem's parameters. Indeed, in inference problems the same person sometimes uses only the base rate, other times only the likelihood (Bordalo, Conlon, Gennaioli, Kwon, and Shleifer [19]). Categorization can produce such systematic and large valuation shifts, but noise cannot.

Different experiences create systematic heterogeneity (Point i). A DM who has more frequently or recently used a category c, higher F_c, is more likely to

- even when correct - when context changes, which is called "limited portability of knowledge"; see Bassok [6] for a compelling illustration in math problems. In our model, these forces emerge due to similarity and frequency based retrieval of categories.

focus on its relevant features and choose a_{tc}. Familiarity with different categories may explain persistent interpersonal differences in attention and judgments in statistical problems despite common information and incentives (Bordalo, Conlon, Gennaioli, Kwon, and Shleifer [19]).

Changes in context deliver instability (Point *ii*). A higher context discrepancy $d(t, c)$ reduces categorization in c, reducing the probability of choosing a_{tc}. Even spurious changes in context can yield this effect. Describing the same inference problem as taxicabs versus balls and urns changes problem representation, namely the focus across available statistics, not information. This yields a theory of framing effects whose consequences we explore in detail in Section 4.

Crucially, experiences and context interact (Point *iii*). If category c is not a dominant representation of the current problem, $\Pr(c|t) \leq \pi^*$, context and frequency are complements: familiarity with c boosts $\Pr(a_{tc})$ especially if the context is more similar. This explains mental simulation: non-domain specific experiences affect choice in similar contexts, especially if the current problem is "new".[9] It also means that changes in irrelevant context produce strong framing effects when the DM is "indifferent" between different representations.[10]

If instead category c is dominant, then similarity and frequency are substitutes: changing context has a small or no effect for DMs who often use c. Framing effects are thus small in conditions that include those of the rational model (frequent use of a correct representation), but extend to strong familiarity with a category despite its low fit, leading to overgeneralization and neglect of

[9] Bordalo, Burro, Coffman, Gennaioli, and Shleifer ([12]) show that past personal financial losses make it easier to imagine a severe cyberattack by helping the DM to imagine how these events may lead to losses. See also Taubinsky, Butera, Saccarola, and Lian [108] on inflation.

[10] Framing effects should also be heterogeneous. There is evidence of this: a risky lottery with a stock market label is less likely to be chosen by people who think stock market participants are greedy (Henkel and Zimpelmann [58]). Describing default on a loan as contrary to sacred texts increases repayment (Bursztyn, Fiorin, Gottlieb, and Kanz [20]), but possibly more so for more religious people. In the original problems the primed category (stocks, religion) is not dominant, but evoking it sways people for whom it is familiar.

informative data, as in cognitive dissonance (Festinger [37]) or the confirmation bias (Nickerson [86]). Highly religious people may see many choices from the perspective of their values and neglect other features, even if important.

4 Top-Down Attention in Famous Puzzles

We illustrate top-down attention in consumer choice and statistical problems. Categorization of choice as consuming versus buying explains the high - sometimes excessive - price elasticity of people with poverty experiences and various forms of mental accounting. Categorization of statistical problems into simple sampling versus inference reconciles the Gambler's Fallacy, over- and underreaction to data, and their instability.

4.1 Consumer Choice

A consumer values goods g and b with expected qualities q_g and q_b, $q_g \geq q_b$, and prices p_g and p_b, $p_g \geq p_b$. The net utility of these hedonics is $q_l - \eta \cdot p_l$ for $l = g, b$, where $\eta > 0$ maps dollars into utils. Before consumption, quality and market price can be hit by shocks Δq_l and $-\eta \cdot \Delta p_l$, respectively. We do not formalize the shocks' event features because we study how the consumer represents choice with riskless categories, which neglect events. We study risky choice in Section 5 with bottom-up attention.

The consuming category, $c = con$, has diagnostic context reporting the expected qualities Q_{con} and shocks ΔQ_{con} evaluated during consumption, and the situations Z_{con} in which consumption occurred, so $\kappa_{con} = (Q_{con}, \Delta Q_{con}, Z_{con})$. For choice features, the DM attends to realized qualities and neglects price features, setting $\alpha_{Q,con} = \alpha_{\Delta Q,con} = \overline{\alpha} > \alpha_{P,con} = \alpha_{\Delta P,con} = 0$.

Buying, $c = buy$, has diagnostic context $\kappa_{buy} = (Q_{buy}, P_{buy}, Z_{buy})$ consisting of the expected qualities Q_{buy} and prices paid P_{buy} in buying experiences and their situations Z_{buy}. The DM attends to paying, to a lesser extent to expected quality, and neglects (low probability) shocks, consistent with *buy* being a riskless category, $\alpha_{P,buy} = \overline{\alpha} > \alpha_{Q,buy} = \underline{\alpha} > \alpha_{\Delta Q,buy} = \alpha_{\Delta P,buy} = 0$.

Current context lists all potentially accessible features: qualities $Q_t = \{q_g, q_b\}$, prices $P_t = \{p_g, p_b\}$, shocks to quality $\Delta Q_t = \{\Delta q_g, \Delta q_b\}$ and to price $\Delta P_t = \{\Delta p_g, \Delta p_b\}$, and situation z_t. The distances for diagnostic features of categories *con* and *buy* is $d_Q(Q_t, Q_c)$ for quality, $d_P(P_t, P_{buy})$ for price and $d_Z(z_t, Z_c)$ for the situation. We only study changes of situation distance.[11]

4.1.1 Poor Experiences and Price Sensitivity

Shah, Zhao, Mullainathan, and Shafir [102] show that poor people are more likely to deem price as very relevant, and hence to exhibit high price elasticities, across many situations. They argue that such price focus is the consequence of living on a tight budget and can sometimes leads to mistakes, such as cutting down on important medication whose price increased (Chandra, Flack, Obermeyer [23]). Our model yields this intuition, with new predictions.

Consider a consumer deciding whether to buy g, so $q_b = p_b = 0$. She may think about consuming and focus on quality, or about buying and focus also on price, $C = \{con, buy\}$. In this example, we suppose for simplicity that there are no shocks, $\Delta q_g = \Delta p_g = 0$, so choice is truly riskless. Such shocks play a role in mental accounting, Section 4.1.2. Across situations the consumer retrieves buying with probability π_{buy} and consuming otherwise, so average quality and price weights are $\overline{\alpha}_Q = \overline{\alpha} - (\overline{\alpha} - \underline{\alpha}) \cdot \pi_{buy}$ and $\overline{\alpha}_P = \overline{\alpha} \cdot \pi_{buy}$. Using Propositions 2 and 3 we characterize the effect of poverty on valuation $v(g)$ as a comparative

[11] We can realistically set zero distance along quality and price context, capturing the intuition that a problem can be seen as consumption if there are qualities and as buying if there is also a positive price. More interesting metrics arise along finer categories such as buying "high vs. low price goods" or consuming "great vs. normal qualities."

static on frequency: having often been on a tight budget, poor people more often attended to prices, so their F_{buy} is higher. In the Appendix we show:

$$\begin{aligned} \frac{\partial v(g)}{\partial F_{buy}} &\propto \left[(\alpha_{Q,buy}-\overline{\alpha}_Q)\cdot q_g-(\alpha_{P,buy}-\overline{\alpha}_P)\cdot\eta\cdot p_g\right]\cdot\left[1-d\left(t,buy\right)\right] \qquad (9)\\ &= -\left[(\overline{\alpha}-\underline{\alpha})\cdot q_b+(1-\underline{\alpha})\cdot\eta\cdot p_b\right]\cdot\pi_{buy}\cdot\left[1-d\left(t,buy\right)\right]<0. \end{aligned}$$

Due to a "mental set" of price-benefit evaluations, a poor consumer is more price sensitive and values a price-quality combination less compared to a richer one. This focus on price does not reflect a higher price distaste η nor optimal attention allocation. It reflects overgeneralization of experiences. This leads to:

1. Mistakes. A poor consumer may forsake valuable expenditures such as health copayments due to her focus on cutting costs and neglect of future benefits, which arises because $\underline{\alpha} < \overline{\alpha} \leq 1$ in (9). A formerly poor consumer may exhibit high price elasticity even if she is no longer poor, namely even though η is small in (9), inconsistent with neoclassical and rational inattention models.[12]

2. Instability. Price focus depends on spurious context. By Proposition 1, $d\left(t,buy\right)$ in (9) increases in situation distance $d_Z\left(z_t,Z_{buy}\right)$, which boosts quality focus, reducing price elasticity. The poor can "splurge" on festivals or "treat" goods such as cigarettes (Banerjee and Duflo [3]). These situations are associated with consumption pleasure, which increases $d_Z\left(z_t,Z_{buy}\right)$. Conversely, the poor are more price elastic if costs are monetary rather than in kind: out of pocket costs increase similarity to buying, reducing $d_Z\left(z_t,Z_{buy}\right)$. This is in line with the compatibility principle (Tversky, Sattath, and Slovic [120] and Slovic, Griffin, and Tversky [105]). A no-longer-poor consumer may be price elastic on items she used to buy (e.g., clothes), but not on new goods (e.g., i-Phones). The

[12] Hoch, Kim, Montgomery, and Rossi [59] show that consumers' price elasticity can be predicted from a range of characteristics beyond current income and wealth.

former goods match buying experiences, having lower $d_Z(z_t, Z_{buy})$.

Rick, Cryder, and Loewenstein ([92]) develop a survey measure of thriftiness and show systematic (e.g. age based) heterogeneity in consumers' focus on paying. Wakefield and Inman ([124]) show that consumers' price elasticity is strongly situation-dependent and correlated with the extent to which a good is categorized as "hedonic" (low elasticity) versus "functional" (high elasticity). We offer a mechanism for these findings, with new implications. Systematic heterogeneity in similar conditions between people with different experiences and context-driven instability are the key new predictions of our model compared to stable preferences or biases.

4.1.2 Mental Accounting

People use different accounts to track costs and benefits in different situations, leading to opportunity cost neglect, sunk cost fallacy, non-fungibility of money, etc. Consider the examples below.

Opportunity Cost Neglect. Many years ago a person bought for 20$ a bottle of wine worth 75$ today. The person drinks the wine today. What is the cost she feels? Many answers to this question are zero or 20$ (Thaler [109]). They neglect the opportunity cost of drinking, the 75$ market price.

Sunk Cost Fallacy. A person bought a 20$ ticket to a football game to be played a month later. On the day of the game, there is a severe blizzard. 1) Does the person drive to the game? 2) Would she drive if she was given the ticket for free? Frequent answers are: "yes" to 1) and "no" to 2), which violate revealed preference: if the blizzard is severe enough, it should discourage driving regardless of whether a price had been paid.

Explanations of mental accounting invoke different forces. Opportunity cost neglect is due to temporal remoteness of the wine's purchase price (Shafir and

Thaler [101]), the sunk cost fallacy to diminishing sensitivity (Thaler [109]) or distaste for "waste" (Shafir and Thaler [101]). We offer a unifying explanation, with new predictions, based on the use of the consuming versus buying categories to determine which features are relevant.[13]

In both problems, there is utility q_g - from drinking or seeing the game - and the price p_g paid for it. Price and quality "shocks" are also described: the wine capital gain $\Delta p_g > 0$ in one case, the blizzard $\Delta q_g < 0$ in the other. The dollar cost m felt after drinking, and the value v of going to the game, depend on attention $\alpha_O = (\alpha_Q, \alpha_{\Delta Q}, \alpha_P, \alpha_{\Delta P})$:

$$m(drinking(\alpha)) - m(not\ drinking(\alpha)) = \alpha_P \cdot p_g + \alpha_{\Delta P} \cdot \Delta p_g, \quad (10)$$

$$v(driving(\alpha)) - v(not\ driving(\alpha)) = \alpha_Q \cdot q_g + \alpha_{\Delta Q} \cdot \Delta q_g. \quad (11)$$

Full attention $\alpha_O = (1, 1, 1, 1)$ yields rationality. In wine, $\alpha_P = \alpha_{\Delta P} = 1$ recovers in (10) the market price, $p + \Delta p = 75\$$. For the purpose of estimating monetary costs, all that matters is attention to prices, utility q is not relevant. In the football problem, $\alpha_Q = \alpha_{\Delta Q} = 1$ also recovers the rational rule: go to the game if and only if $q + \Delta q > 0$. Now all that matters is attention to quality, the sunk price p is not relevant.

Each vignette creates a context κ_t matching both the diagnostic features of "consuming" (it reports qualities), and of "buying" (it also reports price). These competing categories prompt evaluations:

$$m(drinking(\alpha)) - m(not\ drinking(\alpha)) = \begin{cases} 0 & if\ \alpha = \alpha_{con} \\ p_g & if\ \alpha = \alpha_{buy} \end{cases}, \quad (12)$$

$$v(driving(\alpha)) - v(not\ driving(\alpha)) = \begin{cases} \overline{\alpha} \cdot (q_g + \Delta q_g) & if\ \alpha = \alpha_{con} \\ \underline{\alpha} \cdot q_g & if\ \alpha = \alpha_{buy} \end{cases} \quad (13)$$

[13] Kőszegi and Matějka [70] offer a rational inattention theory for category-budgets and naive diversification, but cannot explain the sunk cost fallacy or the wine example.

In the wine example, a DM thinking about consuming, α_{con}, focuses on the pleasure of drinking the wine and feels no opportunity cost. A DM thinking about buying, α_{buy}, focuses on the pain of paying and reports p_g, neglecting the capital gain, which is not a standard feature of most purchase decisions.

In the football example, a DM thinking about consuming, α_{con}, focuses on current hedonics, the game q_g and the blizzard Δq_g, making the rational evaluation. A DM thinking about buying, α_{buy}, instead focuses on the pain of paying p_g and the benefit q_g it had secured. This consumer represents her choice as "enjoying a game I paid for". The blizzard shock is neglected because it is not a standard feature of purchase decisions.

Most people adopt the *con* category in the wine problem, and *buy* in the football one. In both cases, mistakes are due to "top down" focus on an irrelevant feature: the pleasure of drinking in wine and the sunk price in football. These features draw attention because they are relevant in frequent past problems, but also cause neglect of relevant features of the current one. Proposition 3 yields several comparative statics:

1. Frequency. People who have recently bought lots of wine but have not yet drunk it have high F_{buy}, which favors $c = buy$ and hence the \$20 mode. In football, people who bought season tickets face only one buying decision but many consuming experiences; thus, they have a high F_{con}, which hinders a *buy* representation, reducing the sunk cost fallacy. A wine trader has more frequent buying experiences, so should exhibit less opportunity cost neglect, as in List [78] (and will have a "sell wine" category, prompting attention to the capital gain). Having hgher F_{buy}, poor people should exhibit less opportunity cost neglect and more sunk cost fallacy.

2. Instability. The prevailing modes in the two examples can be explained by context. Describing the wine situation as $z_t =$ "drinking the wine" prompts

similarity to *con* and dissimilarity to *buy*, fostering full price neglect.[14] Describing the football situation as z_t = "going versus not going to the game" evokes a buying decision we already made: paying to see a game, fostering neglect of the blizzard. Making the blizzard more salient in the description or making an ex ante plan for bad weather should increase reliance on *cons* and reduce that on *buy*, reducing the fallacy. We study the role of description in categorization in Section 5.

The same mechanism can explain non-fungibility: transferring money into a category cues buying experiences in that category, promoting in-category spending.[15] Consider also account-based commitment (Thaler [109]): setting up a "rainy day" account creates a category of decisions focused on future financial risks, which is retrieved when we consider withdrawing from this account. Such strategies are however on a slippery slope: exceptions can destroy the categories on which commitments are based.

4.2 Statistical Problems

Probability judgments in i.i.d. draws and inference exhibit systematic biases (Benjamin [9]). Bordalo, Conlon, Gennaioli, Kwon, and Shleifer [19] show that multi-modality and instability of attention can account for them. Their theory does not however explain why different people systematically attend to different features, making different errors in the same problem, and why changes in irrelevant context create instability. Our theory does.

[14] During many consumption experiences, with prices not explicitly mentioned, we often feel no opportunity costs. Frederick, Novemsky, Wang, Dhar, and Nowlis [39] show experimentally that describing the option "not buy" in terms of keeping the money for other purchases substantially decreases the probability of purchase.

[15] A 5$ bonus for drinks at a restaurant is similar to past "drink discount" experiences, in which the consumer focused on whether to buy en extra beer or a higher quality one. This focus on drinks reduces attention to the food spending category (Abeler and Marklein [1]).

A coin with probability $\theta \in [0,1]$ of heads is selected according to a prior $\mathbb{P}_0$ in a set of coins Θ, and generates i.i.d. sequences of up to $D \in \mathbb{N}$ binary draws. Hypotheses $o \in O$ are events in the sampling space Ω, whose probability can be estimated in $(\Omega, \mathcal{F}, \mathbb{P})$.[16] A hypothesis is an elementary event $\omega \in \Omega$, with feature vector:

$$y(\omega) = \left(e_\theta, e_{1|\theta}, ..., e_{D|\theta}, s_{|\theta}\right), \tag{14}$$

where e_θ is the event of selecting coin θ, $e_{j|\theta}$ is the realization of the j-th i.i.d. draw given coin θ, and $s_{|\theta}$ is the event corresponding to the share of positive i.i.d. draws (e.g. heads) given coin θ. Statistical categories emerge from experiences with these features.

The context κ_{ss} of simple sampling, category ss, is "the DGP produces a single event" whose probability is given, and "hypotheses' names are possible realizations of the event". The diagnostic features of ss are thus the number of events N, the event realizations R (a fair coin has $R = \left\{h_{|0.5}, t_{|0.5}\right\}$), and the hypotheses' names V. Category ss has context:

$$\kappa_{ss} = (N_{ss} = 1, R_{ss}, V_{ss} \subseteq R_{ss}). \tag{15}$$

When evaluating $y(\omega)$ the DM focuses on the names of hypotheses, setting $\alpha_{ss,R} = 1$ for $R \subseteq V$ and neglects the rest $\alpha_{ss,i} = 0$ for $i \in M_O \backslash R$. Within ss, this process is both intuitive and correct.[17]

The context of agnostic inference, category ai, is "the DGP has more than one event", "one event is associated with coin selection", "other events are i.i.d. draws", and "the names of hypotheses are realizations of the coin selection

[16] The option corresponding to $H \subset \Omega$ is a lottery whose atoms specify, for each elementary $\omega \in \Omega$, a payoff of 1\$ if $\omega \in H$ and zero otherwise. Consistent with Savage [97], Section 9, we could allow for an incentive compatible elicitation for multiple events.

[17] In the described DGP, ss can entail estimating the probability of the iid draw given the coin, e.g. $V = D_{1|\theta}$ or that of selecting a coin type θ, $V = \Theta$.

event". The diagnostic features of ai are thus the total number of draws N, the DGP selection event Θ, the i.i.d. draws R, and hypotheses' names V. Thus, category ai has context:

$$\kappa_{ai} = \left(N_{ai} \neq \{1\}, \Theta_{ai} \neq \{\theta\}, R_{ai}, V_{ai} \subseteq \Theta_{ai}\right), \tag{16}$$

so there is more than one draw and Θ_{ai} is not a singleton. When evaluating $y(\omega)$ the DM focuses on its share of positive signals $s_{|\theta}$ and neglects the rest, including coin selection, setting $\alpha_{ai,s_{|\theta}} = 1$ and $\alpha_{ai,i} = 0$ for $i \in M_O \backslash S_{|\theta}$. With an uninformative prior this process is also intuitive and correct: the share of positive signals is a sufficient statistic for θ. Since we often have little prior information, $c = ai$ generally works well for inference.[18]

Current context κ_t concerns the DGP and hypotheses. DGP context comprises the number of realizations N_t, the set of possible coin types Θ_t, a sequence of i.i.d. draws $D_{1|\theta}$,..., $D_{N_t-1|\theta}$ and a set of the possible shares $S_{|\theta,t}$ for each coin. Hypotheses V_t reflect events based on the DGP features Θ_t or $D_{j|\theta}$. Statistical context consists of potentially relevant events. Mistakes are due to the DM's focus on the wrong events based on past problems in which they were relevant.

4.2.1 The Gambler's Fallacy

The DM estimates the relative probability of obtaining sequences $H_1 = hhhhhh$ versus $H_2 = htthht$ from a fair coin. The diagnostic features of categories ss, ai take values $N_t = 6$, $\Theta_t = \{0.5\}$, $R_t = \left\{h_{|0.5}, t_{|0.5}\right\}$, $V_t = \{H_1, H_2\}$.

Comparing these features to those of simple sampling $c = ss$ in (15), there is a discrepancy along the number of flips $d_N(6, 1) > 0$ and along the names of hypotheses, which do not correspond to any individual realizations in the

[18] A category of lopsided inference, in which the prior is highly skewed and so it receives attention while the signal is neglected, is not necessary in our model. It endogenously emerges from the simple sampling category when statistical contrast is introduced in Section 5.

described DGP, $d_V(V_t, R_t)$, $d_V(V_t, \Theta_t) > 0$.

Compared to agnostic inference in (16), there is a discrepancy along the set of coin types (the current one is a singleton, $d_\Theta(\Theta_t, \Theta) > 0$), and with hypotheses' names, which do not correspond to coin types, $d_V(V_t, \Theta_t) > 0$.

Because the problem is not a perfect match to either category, yet both categories are often encountered, some people focus on the fairness of the coin and represent the problem as ss, others focus on the length of sequences and represent the problem as ai.

A DM relying on ss deems an individual flip relevant, say R_1. Thus, her attention is $\alpha_1 = 1$ and $\alpha_i = 0$ otherwise. By Equation (3):

$$v(H_1(\alpha)) = v(H_2(\alpha)) = \mathbb{P}(h) = 0.5.$$

Hypotheses are deemed equally likely. The DM does not commit the GF nor does she estimate the probability of H_1 and H_2 correctly. She is not "rational". She uses a sampling intuition "with a fair coin any draw is equally likely!"

A DM relying on ai deems only the share of heads to be relevant, $\alpha_{s_{|0.5}} = 1$ and $\alpha_i = 0$ otherwise. By Equation (3),

$$v(H_1(\alpha)) = \mathbb{P}\left(s_{|0.5} = 1\right) = (0.5)^6 \; \textit{and} \; v(H_2(\alpha)) = \mathbb{P}\left(s_{|0.5} = 0.5\right) = 5 \cdot (0.5)^4.$$

The DM commits the Gambler's Fallacy, using a "comparative" inference intuition: balanced sequences are much more likely with a fair coin! Consistent with the evidence, her estimated probability for H increases in the size of its share of heads equivalence class.[19]

Heterogeneity in representations thus explains disagreement despite common

[19] We could allow for richer sampling categories entailing $r > 1$ iid flips. [19] allow for them in reduced form. These categories would not change our basic results, which rely on the familiarity of inference experiences. r-sampling categories would help produce the well known insensitivity to sample size, in a less crude form than with $r = 1$.

incentives and information. The GF is due to categorization into inference, driven by familiary and by having long sequences. Accordingly, as shown in [19] there is less GF when sequences get shorter, $d_N(N_t, 1)$ falls, and when the name of hypotheses coincides with individual flips, $d_V(V_t, R_t) = 0$, which fosters ss.[20] More generally, our model implies that adding irrelevant context diagnostic of a similar problem should boost attention to a feature that is relevant in that problem (and vice versa if such context is removed).

Rabin and Vayanos [89] show that a DM who incorrectly believes in the GF may exhibit a nonmonotonic pattern of under/overreaction depending on whether the DGP is independent or autocorrelated. Their model cannot explain why the same person exhibits the GF on one occasion but not on a normatively equivalent occasion based on changing context.

4.2.2 Biases in Inference

Competition between $c = ss$ and $c = ai$ also yields multimodality and instability in inference. In balls and urns problems, urn A is selected with probability 25% and the likelihood of drawing a green (versus blue) ball from it is 80%. The composition of urn B is symmetric. People are told: "a drawn ball is green. What is the probability it comes from A versus B?" Current context has $N_t = 2$, $\Theta_t = \{A, B\}$, $R_t = \left\{g_{|A}, b_{|A}, g_{|B}, b_{|B}\right\}$, $V_t = \{A, B\}$, where g stands for green, b for blue.

By (14), the urn U-hypothesis is $y_U = \left(U, g_{|U}, s_U = 1\right)$. By (3), its value is:

$$v(y_U(\alpha_O)) = \mathbb{P}(U(\alpha_O))^{\alpha_U}\,\mathbb{P}((1|U)\,|U(\alpha_O))^{\alpha_S\cdot\mathbb{I}_S+\alpha_D\cdot(1-\mathbb{I}_S)}, \tag{17}$$

where α_U is attention to urn selection, α_D to the i.i.d. draw, and α_S to the

[20] Specifically, when asked to judge *hhhhht* vs. *hhhhhh*, the incidence of the GF is reduced if subjects are told "the first five flips are *hhhhh*" is the last flip h or t?

green share. $\mathbb{I}_S = 1$ if $\alpha_{\mathcal{S}} > 0$, zero otherwise.[21] Equation (17) is Grether's [52] formula with coefficients shaped by categorization.

Balls and urns are more similar to inference ai than to ss because there is more than one draw, $d_N(2, N_{ai}) < d_N(2,1)$, the i.i.d. process is not known, $d_\Theta(\Theta_t, \neq \{\theta\}) < d_\Theta(\Theta_t, \{\theta\})$. Balls and urns are however still quite similar to ss because hypotheses' names coincide with the urn selection event, $d_V(V_t, \Theta_t) = 0$, whose probability is given.

Familiarity with simple sampling, high F_{ss}, prompts some people to rely on it, despite its lower fit compared to ai. These people attend only to urn selection, answering by $v(y_U(\alpha_{ss})) = \mathbb{P}(U)$. They reason "the ball has to come from A or B!" They neglect color and report the base rate. People who rely on inference instead attend only to the green share, answering by $v(y_U(\alpha_{ai})) = \mathbb{P}(1|U) = \mathbb{P}(g_{|U})$. They reason: "$A$ is more green than B!". They neglect the low base rate of A and answer by the likelihood. Categorization yields multimodality in [19].

Experiments show strong instability of estimates with the taxicab frame: There are two companies, called Green (25% of all cabs) and Blue (75%) according to the color of their cabs. After a hit and run accident, a witness, whose accuracy is 80% for each cab color, reports the cab to be green. What is the probability that the cab is indeed green? In this frame, most people answer "80%" and almost no one answers with the base rate.

Our model explains this instability because in taxicabs DGP context is the same as under balls and urns but now hypotheses' names are the events "accurate" versus "inaccurate" witness reports, which correspond to i.i.d. draw events in $D_{1|G}$ and $D_{1|B}$. One hypothesis, "the reported color is correct", corresponds to $g_{|G} \in D_{1|G}$, while the other corresponds to $g_{|B} \in D_{1|B}$. By setting

[21] Consistent with the green share being a sufficient summary of iid draw events, the DM's causal model first conditions on it and then on individividual iid draws from U. This assumption is not material for our results.

names $V_t = \{g_{|G}, g_{|B}\}$ taxicabs increase similarity to and hence categorization into simple sampling ss. This focuses the DM on the i.i.d. draws, setting $\alpha_D = 1$ and $\alpha_U = \alpha_S = 0$ in Equation (17), yielding the "80%" answer. Bordalo, Conlon, Gennaioli, Kwon, Shleifer [19] indeed show that instability in the taxicab frame obtains almost entirely due to switchers from the base rate to the likelihood.[22]

When the question corresponds to a described feature of the DGP, people anchor to the corresponding statistics based on a simple sampling representation. They then under-react if the statistic is the base rate, over-react if it is the likelihood. This compatibility effect (Tversky, Sattath, and Slovic [120] and Slovic, Griffin, and Tversky [105]) is naturally delivered by our model.

5 Top Down and Bottom up Attention

Bottom-up salience is a driver of choice. Consumers are more sensitive to taxes when these are shown on the price tag (Chetty, Looney, and Kroft [25]) and prefer goods that are physically present (Bushong, Camerer, and Rangel [21]). Prices, payoffs, or statistics draw attention if they are contrasting (Bordalo, Gennaioli, and Shleifer [14],[15], Koszegi and Szeidl [71]).[23] We show how two bottom-up forces, sensory prominence and contrast, interact with top-down attention.

Sensory prominence depends on the problem's description. A description is a vector $\alpha_\delta \in [0,1]^{M_O \cup M_K}$ paired with context κ_δ. It is partly set by nature (e.g., "sun" is exogenously part of $\kappa_{\delta,i}$ and visually prominent, high $\alpha_{\delta,i}$) and partly designed. A shrouded feature has $\alpha_{\delta,i} = 0$ (Gabaix and Laibson [42]). A

[22] This explanation also accounts for why the same effect is observed when the balls and urns problem is "cabified", namely urn compositions are described in terms of the color match between the urn and the ball, and the hypotheses themselves concern accuracy. The effect is so strong also because the DGP is described in terms of the "match" feature, a result that naturally emerges from bottom up attention.

[23] See Lanzani [75] for the axioms underpinning this model.

fully described feature has $\alpha_{\delta,i} = 1$. Described context is all the DM perceives, so $\kappa_t = \kappa_\delta$. Our prior analysis captures adding or shrouding context features in κ_t. We now consider the role of α_δ for choice features.

Contrast increases in the variability of a choice feature across atoms. High variability draws attention. Let Y be the set of atoms. Contrast of $i \in M_O$ is a real valued function $\sigma_i = \sigma\left[(y_i)_{y\in Y}\right] \geq 1$. A special case is:

$$\sigma_i = 1 + \frac{\sum_{y,y'\in Y:y\neq y'} d_i\left(y_i, y_i'\right)/|Y||Y-1|}{\sum_{y\in Y} d_i\left(y_i, \widetilde{y}\right)/|Y| + \epsilon}, \tag{18}$$

where $\epsilon > 0$ and $\widetilde{y}$ is a reference. For $y_i \in \mathbb{R}$, it nests the exemplar salience function in Bordalo, Gennaioli, and Shleifer [15], with $d\left(y_i, y_i'\right) = |y_i - y_i'|$ and $\widetilde{y} = 0$. For instance, large price variability increases price contrast via the numerator of (18), high average price reduces it via the denominator (by Weber-Fechner's Law). Event contrast depends on event probabilities, as in Bordalo, Conlon, Gennaioli, Kwon, and Shleifer [19].

Described contrast of a feature, $\sigma_{i\delta}$, is computed using the described values of i. If the feature is shrouded, $\alpha_{\delta,i} = 0$, its bottom-up contrast is minimal, $\sigma_{i\delta} = 1$. Category contrast of the same feature, σ_{ic}, is computed using the feature's experienced values in c. If i was neglected ($\alpha_{c,i} = 0$), $\sigma_{ic} = 1$. Experiences create "top down" contrast. Before flying, we often think about crashes: even if not described, these events are contrasting in $c =$"flying".[24]

Matching. The DM matches description $(\alpha_\delta, \kappa_t)$ with each category $c \in C$.

[24] The feature values used to compute contrast are encoded in context, κ_t and κ_c, which report the possible values of choice features. Top down contrast can be extended to heterogeneous categories. For each $\tau \in c$, feature κ_{τ,i_K} reports the experienced values of i at τ. If the feature was not attended to, $\alpha_{ic\tau} = 0$, then κ_{τ,i_K} is empty so its contrast in $\tau \in c$ is zero. Else, κ_{τ,i_K} reports the feature's values in Y_τ. The DM computes σ_{ic} as the average $\sum_{\tau\in c}\sigma_{\tau ic}/|c|$ across problems in c.

Equation (6) becomes:

$$S(t, c, \delta|\sigma) = \max_{\alpha_t} \; F_c \cdot S[(\alpha_t, \kappa_t), (\alpha_c, \kappa_c)|\sigma_c] + S[(\alpha_t, \kappa_t), (\alpha_\delta, \kappa_t)|\sigma_\delta]. \quad (19)$$

The DM trades off similarity to c, the first term, with similarity to the description, the second term. Similarities to the category and the description depend on contrast σ_c and σ_δ, with more contrasting features receiving higher weight in similarity (see Nosofsky [87] and Reed [91]). Under a simple multiplicative specification,[25] the interior optimal attention satisfies:

$$\frac{\partial}{\partial \alpha_{t,i}} d(|\alpha_{t,i} - \alpha_{c,i}|) + d_{tc}(i) \cdot \mathbb{I}_c(i) + \frac{\sigma_{i\delta}}{\sigma_{ic}} \cdot \frac{1}{F_c} \cdot \frac{\partial}{\partial \alpha_{t,i}} d(|\alpha_{t,i} - \alpha_{\delta,i}|) = 0, \quad (20)$$

where $\sigma_{i\delta} = \sigma_{ic} = 1$ for context features (which do not vary across options). Compared to (7), bottom-up salience directly affects attention via $\alpha_{\delta,i}$ in the third term above. The "top down" analysis of Section 2, in which description only affects the context vector κ_t, is a special case for $F_c \to \infty$.

Proposition 4 *When matching c and $(\alpha_\delta, \kappa_t)$, attention $\alpha_{t,i}(c, \delta, \sigma)$ to feature $i \in M_O$ increases in sensory prominence $\alpha_{\delta,i}$ and attention $\alpha_{c,i}$ of category c. The description is more influential, $|\alpha_{t,i} - \alpha_{\delta,i}|$ is lower, when $\frac{\sigma_{i\delta}}{\sigma_{ic}} \cdot \frac{1}{F_c}$ is higher.*

If a feature such as price is visually prominent, high $\alpha_{P,\delta}$, it is more attended to for any matched category c, as in Li and Camerer [76]. The effect is especially strong if current prices are striking, high $\sigma_{P\delta}$, as in Bordalo, Gennaioli, and

[25] Contrast weighted similarity is given by:

$$S[(\alpha_t, \kappa_t), (\alpha_x, \kappa_x)] = \frac{\sum_{i \in M} \sigma_{ix} \cdot [1 - d(|\alpha_{it} - \alpha_{ix}|)] + \sum_{i \in M_K} \sigma_{ix} \cdot [1 - \alpha_{it}\alpha_{ic} d_i(\kappa_{it}, \kappa_{ix})]}{|M| + |M_K|},$$

where $x = \delta, c$, and where we use the convention $\sigma_{ix} = 1$ for $i \in M_K$.

Shleifer [15],[14]. But top-down factors also matter.[26] If the DM is accustomed to using a quality focused category (e.g. F_{con} is high) price is neglected even if prominently described. Conversely, a poor consumer who often worries about prices, σ_{Pc} is high, attends to them even when shrouded. As shown before, top-down attention can produce stability/description invariance.

Description shapes similarity and categorization.

Proposition 5 *Increasing sensory prominence $\alpha_{\delta,i}$ or described contrast, higher $\sigma_{i\delta}$ for $\alpha_{\delta,i} = 1$, increases similarity more for categories in which that feature is more relevant, $\alpha_{c,i}$ is higher. That is*

$$\frac{\partial^2 S\left(t, c, \delta|\sigma\right)}{\partial \alpha_{\delta,i} \partial \alpha_{c,i}} \geq 0 \textit{ and } \left.\frac{\partial^2 S\left(t, c, \delta|\sigma\right)}{\partial \sigma_{\delta,i} \partial \alpha_{c,i}}\right|_{\alpha_{\delta,i}=1} \geq 0.$$

Categories that attend to features that are prominent and contrasting are more likely to be selected. A "buying" representation is more consistent with prominent or contrasting described prices. A price-focused advertisement can favor retrieval of this category, hinder retrieval of "consuming", and lead to neglect of quality.

Bottom up attention explains why people tend to focus on what is prominently described, neglecting the rest (What You See is All There Is, Kahneman [63], Graeber [51]). The model entails a top-down limit to this force. A shrouded feature is not neglected if the DM frequently attends to it (high F_c) or if it is highly contrasting from experience (high $\sigma_{c,i}$). An ad showing a comfortable hotel room can boost our quality focus but does not cause price neglect: hotel prices are frequently attended to and often contrasting.

Contrast introduces goal relevance a contrasting hedonic is relevant but not goal optimality in attention (Sims [103], Woodford [125], Gabaix [41]). Indeed,

[26] See Kay and Ross [69] for evidence that the sensory stimuli are mediated by categories.

one goal - saving money - can distract us from another goal - buying a good product.

The interaction between bottom up and top-down attention can throw light on choice under risk, statistical problems, and measured similarity judgments.

5.1 Implications for Lottery Choice

Riskless choice and statistical categories shed light on biases in lottery choice, and unifies them with biases in riskless choice. Experiences with risk can create additional categories, for instance on elation for gains and disappointment for losses, but these are best left for a separate treatment.

The DM chooses between monetary lotteries

$$x = (x_g, x_b; \pi) \textit{ and } w = (w_g, w_b; \beta),$$

which pay their highest prizes x_g and w_g with probabilities π and β, respectively, and pay x_b and w_b otherwise. The lotteries have the same expected value but x is riskier than w, $x_g > w_g$ and $\pi < \beta$. Each lottery has four features: its maximum and minimum payoffs (*hedonic*) and the events in which each is delivered (*event*). The generic atom is $y = (u_g, u_b, e_g, e_b)$. Event e_s takes the value s when this state realizes and the neutral value Ω otherwise. Hedonic u_s is equal to the delivered payoff in state $s \in \{g, b\}$, and takes the neutral value 0 otherwise.[27]

When evaluating a lottery, the DM reasons about payoffs and events. Reasoning about payoffs cues experiences of "consuming" similar amounts, and hence category *con* (Section 4.1) Reasoning about events cues experiences of esti-

[27] In a more complete formalization, the atom of lottery x also includes the two event features of the alternative lottery w, so that the probability of the atom is computed using the joint probability distribution of payoffs in $(\Omega, \mathcal{F}, \mathbb{P})$. These features are however redundant in our case because the lotteries are independent.

mating probabilities, the "simple sampling", category ss (Section 4.2).[28] Rather than being integrated, these categories compete, distorting attention to payoffs and probabilities.

In typical experiments features are prominently described, $\alpha_{\delta,i} = 1$ for all $i \in M_O$. Proposition 4 then implies that, when matching to $c \in \{con, ss\}$, under quadratic distance $d(.)$ attention satisfies:

$$\alpha_{t,u_s}(con) = \frac{\overline{\alpha} + \sigma_{\delta,u_s}/F_{con}}{1 + \sigma_{\delta,u_s}/F_{con}}, \qquad \alpha_{t,e_s}(con) = \frac{\sigma_{\delta,e}/F_{con}}{1 + \sigma_{\delta,e}/F_{con}}, \tag{21}$$

$$\alpha_{t,u_s}(ss) = \frac{\sigma_{\delta,u_s}/F_{ss}}{1 + \sigma_{\delta,u_s}/F_{ss}}, \qquad \alpha_{t,e_s}(ss) = 1, \quad s = g, b. \tag{22}$$

Two properties are noteworthy. Ceteris paribus, the "consuming" category boosts attention to payoffs but dampens it to probabilities compared to "simple sampling", $\alpha_{t,u_s}(con) > \alpha_{t,u_s}(ss)$ and $\alpha_{t,e_s}(con) < \alpha_{t,e_s}(ss)$. Second, more payoff or event contrast increases attention to the payoff in state s or to its probability, but less so in more frequent categories (see $\sigma_{u_s\delta}/F_c$ and $\sigma_{e\delta}/F_c$). To see the implications note that, by (18), payoff contrast increases in the payoff difference $|x_s - w_s|$, event contrast increases in the probability difference $|\pi - \omega|$.[29]

By Proposition 5, bottom up attention also shapes categorization. Higher payoff contrast $\sigma_{u_s\delta}$ fosters matching of "consuming" with the description, because in *con* payoffs are relevant. Higher event contrast $\sigma_{e\delta}$ fosters matching of "simple sampling" with the description, because in ss probabilities are relevant.

These forces account for a range of puzzles in risky choice.

Common Ratio Effect. Suppose that $x_b = w_b = 0$. It is well known that if $(100, 0; 0.2) \sim (25, 0; 0.8)$, then for many people $(100, 0; 0.02) \succ (25, 0; 0.08)$.

[28] The other categories are less relevant here: there is no price paid as in *buy* and there are neither an unknwon data generating process nor iid draws as in inference.

[29] We implicitly assume that contrast in state s is only computed for features that take a proper value in such state, and not using features that take values in a diferent state.

This pattern violates expected utility, in which preference rankings are invariant to uniform scaling of probabilities. In our model this effect arises due to contrast. The DM prefers x to w if and only if:

$$v\left(x\left(\alpha_O\right)\right)-v\left(w\left(\alpha_O\right)\right)=\pi^{\alpha_{t,e}}\cdot\alpha_{t,u_g}\cdot(x_g-w_g)+(\pi^{\alpha_{t,e}}-\beta^{\alpha_{t,e}})\cdot w_g\left(\alpha_O\right)\geq 0. \quad (23)$$

Given that $\pi \cdot x_g = \beta \cdot w_b$, the left hand side increases in α_{t,u_g} and decreases in $\alpha_{t,e}$. Thus, higher attention to payoffs α_{t,u_g} boosts risk taking, higher attention to probabilities $\alpha_{t,e}$ boosts risk aversion.

The reduction in probabilities lowers event contrast from $|0.8-0.2|=0.6$ to $|0.08-0.02|=0.06$. This reduces attention to probabilities $\alpha_{t,e}$ in any category, but crucially it also fosters matching to the consuming category, which strongly boosts attention to payoffs α_{t,u_g}. The DM's attention is then drawn away from "peanuts" differences in probabilities; it is drawn instead to the 100$ versus 25$ payoff difference, fostering risk taking.[30]

Rubinstein [94] offers an account of the common ratio whereby similarity among (small) probabilities prompts people to choose based on payoffs. We offer a cognitive foundation based on bottom up contrast and top down similarity to past payoff evaluations versus probability estimations. Compared to Bordalo, Gennaioli, and Shleifer [15], we obtain unstable weighting of probabilities versus payoffs and experienced-based heterogeneity.

Suppose that w is a sure thing $\widetilde{w} \in (0, x_g)$. Hedonic features are whether the lottery pays more or less than $\widetilde{w}$, with corresponding event features. Contrast σ_{δ,u_s} of payoff state $s = g, b$ increases in $|x_s - \widetilde{w}|$. Event contrast is minimal,

[30] This may explain why we neglect many risks in everyday life (Gennaioli, Shleifer, and Vishny [44],[45]). Additional context can include the DM's gain/loss state, her wealth, insider/outsider view of risk (Kahneman and Lovallo [65]). For example, when people are prompted to think as a trader, their loss aversion declines (Sokol-Hessner, Hsu, Curley, Delgado, Camerer, and Phelps [106]).

$\sigma_{\delta,e} = 1$, because events are now perfectly correlated across x and w. Provided $\alpha_{t,u_g} + \alpha_{t,u_b} > 0$, the DM prefers x to $\widetilde{w}$ if and only if:

$$\widetilde{w} \leq \frac{\alpha_{t,u_g} \cdot \pi^{\alpha_{t,e}}}{\alpha_{t,u_g} \cdot \pi^{\alpha_{t,e}} + \alpha_{t,u_b} \cdot (1-\pi)^{\alpha_{t,e}}} \cdot x_g. \tag{24}$$

Given that $\pi \cdot x_g = \widetilde{w}$, the right hand side increases (boosting risk seeking) when relative attention to the lottery upside payoff $\left(\alpha_{t,u_g}/\alpha_{t,u_b}\right)$ is higher. Risk seeking increases with attention to probability $\alpha_{t,e}$ if and only if the lottery is left skewed, $\pi > 0.5$. The following puzzles arise form these effects.

Framing. Risk attitudes change when different payoffs are more or less visually prominent. Consider two normatively equivalent descriptions of x.

$$Full\ Prominence \ : \ win\ x_g\ with\ probability\ \pi\ and\ 0\ otherwise, \tag{25}$$

$$Shrouded\ Downside \ : \ win\ x_g\ with\ probability\ \pi. \tag{26}$$

In (25) both the lottery upside and downside are prominently described, as in (21) and (22). In (26) the lottery downside is described less prominently, i.e., $\alpha_{\delta,u_b} < 1$. This affects equilibrium attention to the downside $\alpha_{t,u_b}(c)$ because, for any matched category c we have:

$$\alpha_{t,u_b}(c) = \frac{\sigma_{\delta,u_b} \cdot \alpha_{\delta,u_b} + F_c \cdot \alpha_{c,u_b}}{\sigma_{\delta,u_b} + F_c}.$$

Although payoffs are constant, lower α_{δ,u_b} reduces attention to the downside α_{t,u_b}, promoting risk taking. This or similar effects are familiar to experimentalists and often treated as unimportant aberrations. In our model they arise from bottom up attention, which can shed light on important real world phenomena. Advertising high returns can foster investment and neglect of risk during good times, when the crash event is shrouded, so investors adopt a "consuming" category, Mullainathan and Shleifer [84] and Célérier and Vallée [22].[31]

Discontinuities. Categories can yield aversion to minuscule risks, as in Kahneman and Tversky's [68] "certainty effect". When $\pi = 1$, x is better than w because $x_g > \widetilde{w}$. Categorization in category *con* and choice of x are straightforward: probabilities are not involved.[32]

Adding a small risk ε, so that $\pi = 1 - \varepsilon$, has two effects: i) it creates a downside payoff feature, whose salience increases in contrast $|\widetilde{w} - 0|$, and ii) it creates an event feature, increasing similarity with "simple sampling".

High payoff contrast and small probability contrast favor a payoff evaluation representation.[33] Many DMs thus focus on the highly contrasting downside, $\alpha_{t,u_b} > 0$, leading to insensitivity to probabilities, fully so $\alpha_{t,e} = 0$ if "consuming" is very frequent $F_{con} \to \infty$. High aversion to the small risk follows.

Barseghyan, Molinari, O'Donoghue, and Teitelbaum [5] show that discontinuous probability weighting at 0 (as suggested in Kahneman and Tversky [68] but subsequently abandoned in favor of a continuous model) is important to understand insurance demand. Haigh and List [54] document discontinuity also with professional traders.[34] In our model this behavior arises because a strik-

[31] A related treatment varies the description of the sure thing by explicitly mentioning that it never yields a downside (unlike the risky alternative), or by keeping it implicit. We thank Alex Imas for making this point as a discussant.

[32] Arguably, in this case there is neither stochasticity nor heterogeneity in choice. In our model, this occurs provided λ is sufficiently high, or if the attention shock is only relevant when all highly frequent categories exhibit an imperfect match, which seems plausible.

[33] Our model predicts that some people may edit out the small risk, but a few people sticking to *con* and neglecting numerical probabilities are enough to produce a discontinuity.

[34] Another important example of discontinuity arises in situations involving social norms. Gneezy and Rustichini [49] show that a small payment reduce effort in the collection of donations, presumably because the payment is now categorized as a low-salary job. Social norms are also at play in purely strategic situations. The majority of players in the dictator game share some of their endowment, consistent with a categorization in terms of the social norm of sharing (Krupka and Weber [72]). However, adding the possibility of taking money leads to a discontinuous change in behavior towards no sharing (List [79]).

ing downside prompts a payoff evaluation mode. This behavior illustrates that choice discontinuities can "reveal" categories: either the continued use of a payoff focused category with an added downside as in the current example, or switches between different categories (e.g., describing an irrelevant "coin toss" context can favor the adoption of ss and thus sensitivity to probabilities).

The Fourfold Pattern and "Simplicity Equivalents" Equation (24) yields the so-called "fourfold pattern" in the gain domain, based on payoff contrast. Given that $x_g = \pi \cdot \widetilde{w}$, upside payoff contrast is $|\widetilde{w}/\pi - \widetilde{w}| = \widetilde{w} \cdot \left(\frac{1-\pi}{\pi}\right)$, downside contrast is $|\widetilde{w} - 0| = \widetilde{w}$. If the lottery is right skewed, $\pi < 0.5$, upside contrast is higher than downside contrast, and vice-versa if $\pi > 0.5$. By (21) and (22), then, in every category the DM focuses more on the upside iff $\pi < 0.5$. The DM is risk seeking for $\pi < 0.5$ and risk averse for $\pi > 0.5$, as in the fourfold pattern.

Contrast of described payoffs reproduces the Bordalo, Gennaioli, and Shleifer [14] payoff-salience mechanism. Our model adds new explanatory power. By Proposition 5, payoffs contrast favors categorization into con, leading to probability neglect, $\alpha_e = 0$, which has three implications.

First, the fourfold pattern is now stronger: it can arise even if the lottery upside and downside are equally attended to $\alpha_{t,u_g} = \alpha_{t,u_b}$. Second, context matters. For instance, eliciting certainty equivalents should strengthen the fourfold pattern compared to, say, making binary choices or choosing a "probability equivalent π" to $\widetilde{w}$. Being in the same dollar units as hedonics, certainty equivalents favor payoff evaluation, in line with the compatibility principle.

Third, and key, "payoff evaluation" is based on a riskless category, so it is a fortiori relevant for riskless domains in which the DM must aggregate payoffs with other features. In Oprea's [88] experiment, option A consists of 90 boxes with \$25 and 10 boxes with \$0, and option B consists of 100 boxes with \$2.5. The subject is paid the total value of the chosen option divided by 100. The two options pay the same, but subjects exhibit a preference for B.

Options have two features: the payoff in boxes $s = g, b$ and their numerosities n_g and n_b. When evaluating A versus B, the sharp contrast between A's payoff in its g and b boxes with B's \$2.5 payoff prompts the DM to categorize the problem as "payoff evaluation", $c = con$. This representation prompts a separate evaluation of different boxes, interfering with attention to the "box type" event and their precise numerosities. From Equation (3), the value of the generic atom of A is

$$v\left(\left(s, x_s\right)\left(\alpha_O\right)\right) = \frac{n_s^{\alpha_{e,t}} \cdot \alpha_{u_s,t} \cdot x_s}{100}, \tag{27}$$

so that option A is preferred to B if an only if:

$$v\left(A\left(\alpha_O\right)\right) \geq v\left(B\left(\alpha_O\right)\right) \Leftrightarrow x \leq \frac{\left(\frac{n_g}{100}\right)^{\alpha_{e,t}} \cdot \alpha_{t,u_g}}{\left(\frac{n_g}{100}\right)^{\alpha_{e,t}} \cdot \alpha_{t,u_g} + \left(\frac{n_b}{100}\right)^{\alpha_{e,t}} \cdot \alpha_{t,u_b}} \cdot x_g. \tag{28}$$

Equation (28) yields, in relative frequencies, the preferences obtained in Equation (24) with respect to probabilities. Behavior in risky lottery and riskless mirrors is the same because mental representations are the same, even though the choice domains are different (so the fourfold pattern cannot come from risk preferences). Striking payoff differences favor a "payoff evaluation" category, causing neglect of probabilities or frequencies.

When choosing between riskless mirrors, the retrieval of the *con* category interferes with the retrieval of the correct "adding up" category. We suspect that many subjects would be able to perform the addtion if explicitly asked to do so, and that they would then be indifferent between A and B. Complexity is in representation, not in computation.

5.2 Bottom up Contrast in Statistical Problems

In Bordalo, Conlon, Gennaioli, Kwon, Shleifer [19], statistical contrast yields: i) stronger GF for longer sequences and ii) stronger base rate neglect in inference when likelihoods are more extreme. We now show that categorization throws new light on these phenomena.

A DM judges the likelihood that n draws of a fair coin produce a balanced sequence H_1 versus a full heads sequence H_2. Among the problem's event features, only the contrast of the share of heads varies with n. Consider as a proxy for it the largest probability difference between any two shares of heads in $(\Omega, \mathcal{F}, \mathbb{P})$:

$$\sigma_{\mathcal{S}} = \frac{\left|\binom{n}{n/2} - 1\right| (0.5)^n}{\binom{n}{n/2}(0.5)^n + (0.5)^n + \epsilon},$$

which indeed increases in n. By Proposition 4, when matching the problem to any category, the DM pays more attention to the share of heads. It feels very striking, and thus attention grabbing, to obtain zero tails in 6 flips. By Proposition 5, this fosters categorization in agnostic inference ai compared to simple sampling ss, causing the GF.

In inference, a DM evaluates the likelihood that a green ball comes from urn A whose base rate is $\pi_A < 0.5$ and whose likelihood of green is $q > 1/2$ or from the symmetric urn B. Here, contrast of the feature "urn selection", $i = U$, increases in $|\pi_A - \pi_B| = 1 - 2\pi_A$. Contrast of the feature "share of green given U" increases in $|q - (1-q)| = 2q - 1$. The more extreme the likelihood, the higher is q, the higher is the contrast of the share of heads. As for the GF, categorization in inference ai is more likely, triggering focus on the green share and base rate neglect. This is in line with the evidence [19].

Instability in GF and inference are due to the same force: bottom-up salience of the "share of heads", triggered by strong statistical contrast, which causes greater reliance on the inference category.

5.3 Top Down Contrast and Unstable Similarity

Tversky [116] famously showed that when people rate similarities between countries on a list, they judge Austria and Sweden as more similar when the list includes Hungary and Poland than when it includes Hungary and Norway. He explained this finding by the contrast principle. When Poland is on the list, political differences are contrasting, so Sweden and Austria are deemed similar. When Norway is on the list, geographic differences are contrasting, so Sweden and Austria are deemed dissimilar.

Here contrast arises top down: the only information people are given is country names, but these prompt focus on a feature contrasting among them (similarly to when seeing the "flight" label we think of the "crash" feature).

When assessing the similarity between Austria (A), Sweden (S), Hungary (H) and either Norway or Poland ($X = N, P$), each atom $y \in Y$ lists the features of a country pair. The Austria-Sweden atom (A, S) reports two "hedonic" features: geographical distance $u_G(A, S)$, political distance $u_P(A, S)$. It also reports the country names (event feature). Attention $(\alpha_{G,t}, \alpha_{P,t})$ to hedonics entails estimated distance

$$v_{AS}(\alpha_{G,t}, \alpha_{P,t}) = \alpha_{G,t} \cdot u_G(A, S) + \alpha_{P,t} \cdot u_P(A, S),$$

which is used as an inverse measure of similarity (embedded in the decision rule D). Because Austria and Sweden are intuitively more distant geographically than politically, $u_P(A, S) < u_G(A, S)$, they are judged more similar when attention to politics $\alpha_{t,P}$ is higher compared to geography $\alpha_{t,G}$.

The DM has experienced two categories: problems $c = G$ in which geographic features are learned or judged, and problems $c = P$ in which political features are learned or judged. The former category attends to geography features while neglecting politics, $\alpha_{G,G} = 1 > \alpha_{P,G} = 0$, the latter does the reverse, $\alpha_{G,P} =$

$0 < \alpha_{P,P} = 1$.

Top down contrast in $c = G$ occurs only along geography, $\sigma(\kappa_G)$ where $\kappa_G = \{u_G(y)\}_{y \in Y}$ are the distances between the four countries. In $c = P$, on the other hand, it only occurs along politics, $\sigma(\kappa_P)$, with κ_P accordingly defined. The description of the problem makes the country names fully prominent while it shrouds the hedonics.

Critically, top down contrast of hedonic features changes with the described country names. When the DM is presented with A,S,H,N, variability along geography is high (N, S versus A, H) while variability along politics is low (N,S, A versus H), so $\sigma(\kappa_{G1}) > \sigma(\kappa_{P1})$, where 1 captures the list A,S,H,N. When the DM is presented with A,S,H,P, variability along geography is low (S versus A, H, P) while variability along politics is high (S, A versus H, P), so $\sigma(\kappa_{G2}) < \sigma(\kappa_{P2})$, where 2 refers to the list A,S,H,P.

Instability in similarity judgments arises because, by point i) in Proposition 6, when the most contrasting category is $c = G$ (country list 1), the DM retrieves this category and focuses on geography, holding A and S dissimilar. When instead the most contrasting category is $c = P$ (country set 2), the DM retrieves this category and focuses on politics, holding A and S similar. The similarity judgment is unstable as documented by Tversky. As in the GF, upon seeing A, S, H, N the DM thinks "what a striking North-East separation between N, S and A, H!". This spontaneous association between the current task and geography does not just increase attention to this feature. It causes neglect of politics, which causes an unstable similarity judgment between A and S, shaped by irrelevant countries in the list.

6 Conclusion

Heterogeneous and unstable mental representations, based on experience and contextual cues, unify puzzling decisions within and across domains. This mechanism remedies a key pitfall of neoclassical and behavioral approaches: the assumption of stable utilities or biases, which is at odds with the strong heterogeneity and instability of choice we see in the data. Our paper characterizes cognitive determinants of mental representations and their effect on choice. More than providing definitive answers, we opens several empirical and theoretical avenues. We discuss four key challenges ahead.

First, a major task for future work is to measure or experimentally control: i) mental representations/categories, ii) contextual similarity, and iii) experiences, and use them in conjunction with choice data. In this respect, our approach is highly complementary to the use of AI methods. On the one hand, the latter can help recover representations from self-reported reasons for choice (Haaland, Roth, Stantcheva, and Wohlfart [53], Link, Peschl, Roth, and Wohlfart, [77]), from which similarity within and across domains can be extracted, as well as to unveil subtle context features as in Ludwig and Mullainathan ([80]). These subtle features can complement domain-inspired context parameters such as the choice set, the DGP, etc. On the other hand, our model may help achieve cognitive interpretability of machine findings, connecting them to precise out of sample predictions, increasing testability and external relevance. This is again possible because our theory, rather than specifying stable choice parameters, allows to link representations and choice with measurable context features.

A second set of implications concerns experiments. Current practice favors the use of abstract protocols so as to better identify "universal" choice biases and minimize experimenter demand. In our model abstraction is desirable for studying general cognitive mechanics, for it affords control of lab-context, bottom-up salience and experiences, with minimal influence from events outside the lab. Abstraction is however problematic for studying real-world choices (e.g. demand for insurance), because removal of naturalistic context can also remove consumers' spontaneous representations, reducing external validity. Our theory suggests that "naturalistic immersion" is a significant benefit of field experiments but also that lab methods can strongly benefit from engineering controlled variation in naturalistic context.

A third major avenue is to study the real world implications of our mechanism. When thinking about redistribution, some voters may think about fairness, others about zero-sum transfers from taxpayers (Chinoy, Nunn, Sequeira, and Stantcheva [26]) based on different experiences, but changes in context such as the specific name or domain of the tax may change problem similarity, representations, and voter preferences. Similar considerations apply to fairness judgments (Kahneman, Knetsch, and Thaler [64]), to cooperation with strangers versus in groups (Enke [33]), etc. Our approach can also help explain strategic behavior. Goke, Weintraub, Mastromonaco, and Seljan ([50]) find that bids in a first-price auction neglect the number of bidders after repeated exposure to a second price auction, in which the number of bidders is irrelevant.

Fourth, and finally, one open question is where categories come from. One possibility is that they are formed through experiences in which bottom-up salient dimensions become diagnostic markers for future classification and storage. The salience of sensory pleasure makes the consumption experience immediately different from the choice whether to buy. Culturally-coded festivities or inappropriate behaviors also create diagnostic markers of experiences, which become salient during socialization. The bottom up forces formalized in Section 5 may thus be helpful in thinking about category formation, possibly shedding light on the persistent role of childhood experiences, of culture, but also on the instability in beliefs and preferences caused by exposure to novel experiences that create new categories, such as moving to a new country or sharp technological or social change.

References

[1] Abeler, J. and Marklein, F., 2017. Fungibility, labels, and consumption, Journal of the European Economic Association, 15, pp. 99-127.

[2] Awh, E. Belopolsky A., and Theeuwes, J., 2012, Top-down versus bottom-up attentional control: a failed theoretical dichotomy, Trends in Cognitive Science, 16, 437-443.

[3] Banerjee, A.V. and Duflo, E., 2007. The economic lives of the poor. Journal of Economic Perspectives, 21(1), pp.141-167.

[4] Barseghyan, L., Prince, J., and Teitelbaum, J.C., 2011. Are Risk Preferences Stable across Contexts? Evidence from Insurance Data. American Economic Review, 101(2), 591-631.

[5] Barseghyan, L., Molinari, F., O'Donoghue, T. and Teitelbaum, J.C., 2013. The nature of risk preferences: Evidence from insurance choices. American Economic Review, 103(6), pp.2499-2529.

[6] Bassok, M., 2003. Analogical transfer in problem solving. The Psychology of Problem Solving, pp.343-369.

[7] Bateman, I., Dent, S., Peters, E., Slovic, P. and Starmer, C., 2007. The affect heuristic and the attractiveness of simple gambles. Journal of Behavioral Decision Making, 20(4), pp.365-380.

[8] Bénabou, R. and Tirole, J., 2011. Identity, morals, and taboos: Beliefs as assets. The Quarterly Journal of Economics, 126(2), pp.805-855.

[9] Benjamin, D.J., 2019. Errors in probabilistic reasoning and judgment biases. Handbook of Behavioral Economics: Applications and Foundations 1, 2, pp.69-186.

[10] Bhui, R. and Gershman, S.J., 2020. Paradoxical effects of persuasive messages. Decision, 7(4), p.239 - 258.

[11] Birnbaum, M.H., Coffey, G., Mellers, B.A. and Weiss, R., 1992. Utility measurement: Configural-weight theory and the judge's point of view. Journal of Experimental Psychology: Human Perception and Performance, 18(2), p.331 - 346.

[12] Bordalo, P., Burro, G., Coffman, K.B., Gennaioli, N. and Shleifer, A., 2024. Imagining the future: memory, simulation and beliefs about COVID. Review of Economic Studies, forthcoming.

[13] Bordalo, P., Coffman, K., Gennaioli, N., Schwerter, F. and Shleifer, A., 2021. Memory and representativeness. Psychological Review, 128(1), p.71.

[14] Bordalo, P., Gennaioli, N. and Shleifer, A., 2012. Salience theory of choice under risk. The Quarterly Journal of Economics, 127(3), pp.1243-1285.

[15] Bordalo, P., Gennaioli, N. and Shleifer, A., 2013. Salience and consumer choice. Journal of Political Economy, 121(5), pp.803-843.

[16] Bordalo, P., Gennaioli, N. and Shleifer, A., 2020. Memory, attention, and choice. The Quarterly Journal of Economics, 135(3), pp.1399-1442.

[17] Bordalo, P., Gennaioli, N. and Shleifer, A., 2022. Salience. Annual Review of Economics, 14, pp.521-544.

[18] Bordalo, P., Conlon, J., Gennaioli, N., Kwon, S. and Shleifer, A., 2023. Memory and probability. The Quarterly Journal of Economics, 138(1), pp.265-311.

[19] Bordalo. P., Conlon, J., Gennaioli, N., Kwon S., and Shleifer A., 2024. How People Use Statistics. Review of Economic Studies, forthcoming.

[20] Bursztyn, L., Fiorin, S., Gottlieb, D. and Kanz, M., 2019. Moral incentives in credit card debt repayment: Evidence from a field experiment. Journal of Political Economy, 127(4), pp.1641-1683.

[21] Bushong, B., King, L.M., Camerer, C.F. and Rangel, A., 2010. Pavlovian processes in consumer choice: The physical presence of a good increases willingness-to-pay. American Economic Review, 100(4), pp.1556-1571.

[22] Célérier, C. and Vallée, B., 2017. Catering to investors through security design: Headline rate and complexity. The Quarterly Journal of Economics, 132(3), pp.1469-1508.

[23] Chandra, A., Flack, E., Obermeyer, Z., 2024. The health costs of cost sharing. The Quarterly Journal of Economics, 139 (4), 2037 - 2082.

[24] Chapman, J., Dean, M., Ortoleva, P., Snowberg, E. and Camerer, C., 2023. Willingness to accept, willingness to pay, and loss aversion (No. w30836). National Bureau of Economic Research.

[25] Chetty, R., Looney, A. and Kroft, K., 2009. Salience and taxation: Theory and evidence. American Economic Review, 99(4), pp.1145-1177.

[26] Chinoy, S., Nunn, N., Sequeira, S. and Stantcheva, S., 2023. Zero-sum thinking and the roots of US political divides (No. w31688). National Bureau of Economic Research.

[27] Colonnelli, E., Gormsen, N.J. and McQuade, T., 2024. Selfish corporations. Review of Economics Studies, 91, pp. 1498-1536.

[28] Conlon, J., 2024, Attention, Information, and Persuasion, mimeo.

[29] Conlon, J., and Kwon, S.,2024, Persuasion through Cues, mimeo.

[30] De Clippel, G. Oprea, R., and Rozen, K., 2024, As if, mimeo.

[31] Einhorn, H.J. and Hogarth, R.M., 1986. Decision making under ambiguity. Journal of business, pp.S225-S250.

[32] Ellis, A. and Masatlioglu, Y., 2022. Choice with endogenous categorization. The Review of Economic Studies, 89(1), pp.240-278.

[33] Enke, B., 2019. Kinship, cooperation, and the evolution of moral systems. The Quarterly Journal of Economics, 134(2), pp.953-1019.

[34] Enke, B. and Graeber, T., 2023. Cognitive uncertainty. The Quarterly Journal of Economics, 138(4), pp.2021-2067.

[35] Enke, B. and Zimmermann, F., 2019. Correlation neglect in belief formation. The Review of Economic Studies, 86(1), pp.313-332.

[36] Evers, E.R., Imas, A. and Kang, C., 2022. On the role of similarity in mental accounting and hedonic editing. Psychological Review, 129(4), p.777 - 789.

[37] Festinger, L., 1957. A Theory of Cognitive Dissonance. Evanston, IL: Row and Peterson.

[38] Finkelstein, A. and McGarry, K., 2006. Multiple dimensions of private information: evidence from the long-term care insurance market. American Economic Review, 96(4), pp.938-958.

[39] Frederick, S., Novemsky, N., Wang, J., Dhar, R. and Nowlis, S., 2009. Opportunity cost neglect. Journal of Consumer Research, 36(4), pp.553-561.

[40] Fudenberg, D., Lanzani, G. and Strack, P., 2024. Selective memory equilibrium. Journal of Political Economy, Forthcoming.

[41] Gabaix, X., 2019. Behavioral inattention. In Handbook of behavioral economics: Applications and foundations 1 (Vol. 2, pp. 261-343). North-Holland.

[42] Gabaix, X. and Laibson, D., 2006. Shrouded attributes, consumer myopia, and information suppression in competitive markets. The Quarterly Journal of Economics, 121(2), pp.505-540.

[43] Gagnon-Bartsch, T., Rabin, M. and Schwartzstein, J., 2023. Channeled Attention and Stable Errors.

[44] Gennaioli, N., Shleifer, A. and Vishny, R., 2015. Neglected risks: The psychology of financial crises. American Economic Review, 105(5), pp.310-314.

[45] Gennaioli, N., Shleifer, A. and Vishny, R., 2012. Neglected risks, financial innovation, and financial fragility. Journal of Financial Economics, 104(3), pp.452-468.

[46] Gennaioli, N. and Tabellini, G., 2023. Identity Politics.

[47] Gilboa, I. and Schmeidler, D., 1995. Case-based decision theory. The Quarterly Journal of Economics, 110(3), pp.605-639.

[48] Gilboa, I. and Schmeidler, D., 2001. A Theory of Case-based Decisions. Cambridge University Press.

[49] Gneezy, U. and Rustichini, A., 2000. Pay enough or don't pay at all. The Quarterly Journal of Economics, 115(3), pp.791-810.

[50] Goke, S., Weintraub, G.Y., Mastromonaco, R. and Seljan, S., 2021. Learning new auction format by bidders in internet display ad auctions. arXiv preprint arXiv:2110.13814.

[51] Graeber, T., 2023. Inattentive inference. Journal of the European Economic Association, 21(2), pp.560-592.

[52] Grether, D.M., 1980. Bayes rule as a descriptive model: The representativeness heuristic. The Quarterly Journal of Economics, 95(3), pp.537-557.

[53] Haaland, I.K., Roth, C., Stantcheva, S. and Wohlfart, J., 2024. Measuring what is top of mind (No. w32421). National Bureau of Economic Research.

[54] Haigh, M.S. and List, J.A., 2005. Do professional traders exhibit myopic loss aversion? An experimental analysis. The Journal of Finance, 60(1), pp.523-534.

[55] Halevy, Y., 2007. Ellsberg revisited: An experimental study. Econometrica, 75(2), pp.503-536.

[56] Hascher, J., Imas, A., Ungeheuer, M. and Weber, M., 2024. A Cognitive Foundation for Perceiving Uncertainty. Available at SSRN.

[57] Hastings, J.S. and Shapiro, J.M., 2013. Fungibility and consumer choice: Evidence from commodity price shocks. The Quarterly Journal of Economics, 128(4), pp.1449-1498.

[58] Henkel, L. and Zimpelmann, C., 2023. Proud to not own stocks: How identity shapes financial decisions.

[59] Hoch, S.J., Kim, B.D., Montgomery, A.L. and Rossi, P.E., 1995. Determinants of store-level price elasticity. Journal of Marketing Research, 32(1), pp.17-29.

[60] Hsee, C.K., 1998. Less is better: When low-value options are valued more highly than high-value options. Journal of Behavioral Decision Making, 11(2), pp.107-121.

[61] Huber, J., Payne, J.W. and Puto, C., 1982. Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. Journal of Consumer Research, 9(1), pp.90-98.

[62] Itti, L. and Baldi, P., 2009. Bayesian surprise attracts human attention. Vision Research, 49(10), pp.1295-1306.

[63] Kahneman, D., 2011. Thinking, Fast and Slow. Macmillan.

[64] Kahneman, D., Knetsch, J.L. and Thaler, R., 1986. Fairness as a constraint on profit seeking: Entitlements in the market. American Economic Review, 76(4), pp.728-741.

[65] Kahneman, D. and Lovallo, D., 1993. Timid choices and bold forecasts: A cognitive perspective on risk taking. Management Science, 39(1), pp.17-31.

[66] Kahneman, D. and Miller, D.T., 1986. Norm theory: Comparing reality to its alternatives. Psychological Review, 93(2), p.136 - 153.

[67] Kahneman, D. and Tversky, A., 1982. The simulation heuristic (pp. 201-208). Judgment under Uncertainty.

[68] Kahneman, D. and Tversky, A., 1979. Prospect Theory: An Analysis of Decision under Risk. Econometrica, 47(2), pp.263-292.

[69] Kay, A.C. and Ross, L., 2003. The perceptual push: The interplay of implicit cues and explicit situational construals on behavioral intentions in the Prisoner's Dilemma. Journal of Experimental Social Psychology, 39(6), pp.634-643.

[70] Kőszegi, B. and Matějka, F., 2020. Choice simplification: A theory of mental budgeting and naive diversification. The Quarterly Journal of Economics, 135(2), pp.1153-1207.

[71] Kőszegi, B. and Szeidl, A., 2013. A model of focusing in economic choice. The Quarterly Journal of Economics, 128(1), pp.53-104.

[72] Krupka, E.L. and Weber, R.A., 2013. Identifying social norms using coordination games: Why does dictator game sharing vary? Journal of the European Economic Association, 11(3), pp.495-524.

[73] Kruschke, J.K., 2020. ALCOVE: An exemplar-based connectionist model of category learning. In Connectionist Psychology (pp. 107-138). Psychology Press.

[74] Laibson, D., 2001. A cue-theory of consumption. The Quarterly Journal of Economics, 116(1), pp.81-119.

[75] Lanzani, G., 2022. Correlation made simple: Applications to salience and regret theory. The Quarterly Journal of Economics, 137(2), pp.959-987.

[76] Li, X. and Camerer, C.F., 2022. Predictable effects of visual salience in experimental decisions and games. The Quarterly Journal of Economics, 137(3), pp.1849-1900.

[77] Link, S., Peichl, A., Roth, C. and Wohlfart, J., 2024. Attention to the Macroeconomy. Available at SSRN 4697814.

[78] List, J.A., 2007. Neoclassical Theory Versus Prospect Theory: Evidence from the Marketplace. Journal of Political economy, 72(2), pp.615-625.

[79] List, J.A., 2007. On the interpretation of giving in dictator games. Journal of Political economy, 115(3), pp.482-493.

[80] Ludwig, J. and Mullainathan, S., 2024. Machine learning as a tool for hypothesis generation. The Quarterly Journal of Economics, 139(2), pp.751-827.

[81] Mack, M.L. and Palmeri, T.J., 2020. Discrimination, Recognition, and Classification, mimeo.

[82] Mullainathan, S., 2002. Thinking through categories, Working Paper, Harvard University.

[83] Mullainathan, S., Schwartzstein, J. and Shleifer, A., 2008. Coarse thinking and persuasion. The Quarterly Journal of Economics, 123(2), pp.577-619.

[84] Mullainathan, S. and Shleifer, A., 2005. Persuasion in Finance, NBER Working Paper 11838.

[85] Nardotto, M. and Sequeira, S., 2021. Identity, media and consumer behavior.

[86] Nickerson, R.S., 1998. Confirmation bias: A ubiquitous phenomenon in many guises. Review of General Psychology, 2(2), pp.175-220.

[87] Nosofsky, R.M., 1986. Attention, similarity, and the identification–categorization relationship. Journal of Experimental Psychology: General, 115(1), p.39.

[88] Oprea, R., 2024. Decisions Under Risk are Decisions Under Complexity. American Economic Review, forthcoming.

[89] Rabin, M. and Vayanos, D., 2010. The gambler's and hot-hand fallacies: Theory and applications. The Review of Economic Studies, 77(2), pp.730-778.

[90] Rabin, M. and Weizsäcker, G., 2009. Narrow bracketing and dominated choices. American Economic Review, 99(4), pp.1508-1543.

[91] Reed, S.K., 1972. Pattern recognition and categorization. Cognitive Psychology, 3(3), pp.382-407.

[92] Rick, S.I., Cryder, C.E. and Loewenstein, G., 2008. Tightwads and spendthrifts. Journal of Consumer Research, 34(6), pp.767-782.

[93] Rosch, E. and Lloyd, B.B., 1978. Principles of Categorization, Routledge.

[94] Rubinstein, A., 1988. Similarity and decision-making under risk (Is there a utility theory resolution to the Allais paradox?). Journal of Economic Theory, 46(1), pp.145-153.

[95] Salant, Y. and Rubinstein, A., 2008. (A, f): choice with frames. The Review of Economic Studies, 75(4), pp.1287-1296.

[96] Samuelson, P., 1963. Risk and uncertainty: A fallacy of large numbers. Scientia, 57(98), 108-113.

[97] Savage, L.J., 1971. Elicitation of personal probabilities and expectations. Journal of the American Statistical Association, 66(336), pp.783-801.

[98] Schank, R.C., 1982. Dynamic memory: A theory of learning in people and computers. Cambridge: Cambridge University Press.

[99] Schwartzstein, J., 2014. Selective attention and learning. Journal of the European Economic Association, 12(6), pp.1423-1452.

[100] Schwartzstein, J. and Sunderam, A., 2021. Using models to persuade. American Economic Review, 111(1), pp.276-323.

[101] Shafir, E. and Thaler, R.H., 2006. Invest now, drink later, spend never: On the mental accounting of delayed consumption. Journal of Economic Psychology, 27(5), pp.694-712.

[102] Shah, A.K., Zhao, J., Mullainathan, S. and Shafir, E., 2018. Money in the mental lives of the poor. Social Cognition, 36(1), pp.4-19.

[103] Simon, H.A., 1955. A behavioral model of rational choice. The Quarterly Journal of Economics, pp.99-118.

[104] Sims, C.A., 2003. Implications of rational inattention. Journal of Monetary Economics, 50(3), pp.665-690.

[105] Slovic, P., Griffin, D. and Tversky, A., 1990. Compatibility effects in judgment and choice. R.M. Hogarth (Ed.) Insights into decision making: Theory and applications.

[106] Sokol-Hessner, P., Hsu, M., Curley, N.G., Delgado, M.R., Camerer, C.F. and Phelps, E.A., 2009. Thinking like a trader selectively reduces individuals' loss aversion. Proceedings of the National Academy of Sciences, 106(13), pp.5035-5040.

[107] Stango, V. and Zinman, J., 2023. We are all behavioural, more, or less: A taxonomy of consumer decision-making. The Review of Economic Studies, 90(3), pp.1470-1498.

[108] Taubinsky, D., Butera, L., Saccarola, M. and Lian, C., 2024. Beliefs About the Economy are Excessively Sensitive to Household-Level Shocks: Evidence from Linked Survey and Administrative Data (No. w32664). National Bureau of Economic Research.

[109] Thaler, R., 1980. Toward a positive theory of consumer choice. Journal of Economic Behavior & Organization, 1(1), pp.39-60.

[110] Thaler, R., 1985. Mental accounting and consumer choice. Marketing Science, 4(3), pp.199-214.

[111] Thaler, R., 1999. Mental accounting matters. Journal of Behavioral Decision Making, 12(3), pp.183-206.

[112] Thaler, R. and Johnson, E.J., 1990. Gambling with the house money and trying to break even: The effects of prior outcomes on risky choice. Management Science, 36(6), pp.643-660.

[113] Treisman, A.M. and Gelade, G., 1980. A feature-integration theory of attention. Cognitive psychology, 12(1), pp.97-136.

[114] Tulving, E., 1972. Episodic and semantic memory, Organization of memory, Academic Press.

[115] Tversky, A., 1972. Elimination by aspects: A theory of choice. Psychological Review, 79(4), p.281 - 299.

[116] Tversky, A., 1977. Features of similarity. Psychological Review, 84(4), p.327 - 352.

[117] Tversky, A. and Gati, I., 1982. Similarity, separability, and the triangle inequality. Psychological Review, 89(2), p.123 - 154.

[118] Tversky, A. and Kahneman, D., 1981. The framing of decisions and the psychology of choice. Science, 211(4481), pp.453-458.

[119] Tversky, A. and Kahneman, D., 1983. Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. Psychological Review, 90(4), p.293 - 315.

[120] Tversky, A., Sattath, S. and Slovic, P., 1988. Contingent weighting in judgment and choice. Psychological review, 95(3), p.371.

[121] Tversky, A. and Shafir, E., 1992. The disjunction effect in choice under uncertainty. Psychological Science, 3(5), pp.305-310.

[122] Tversky, B., 1993. Cognitive maps, cognitive collages, and spatial mental models. In European conference on spatial information theory (pp. 14-24). Berlin, Heidelberg: Springer Berlin Heidelberg.

[123] Verguts, T., Ameel, E. and Storms, G., 2004. Measures of similarity in models of categorization. Memory & Cognition, 32, pp.379-389.

[124] Wakefield, K.L. and Inman, J.J., 2003. Situational price sensitivity: the role of consumption occasion, social context and income. Journal of Retailing, 79(4), pp.199-212.

[125] Woodford, M., 2012. Prospect theory as efficient perceptual distortion. American Economic Review, 102(3), pp.41-46.

[126] Yechiam, E. and Hochman, G., 2013. Loss-aversion or loss-attention: The impact of losses on cognitive performance. Cognitive Psychology, 66(2), pp.212-231.

A Appendix

A.1 Proofs

Proof of Proposition 1. First, we will consider the case when $i \notin M_{Kc}$. By Equation (5), we have that $\alpha_{t,i}$ only affects $S\left[(\alpha_t, \kappa_t), (\alpha_c, \kappa_c)\right]$ additively through the term $-d(|\alpha_{t,i} - \alpha_{c,i}|)$ in the numerator. This is clearly maximized when $\alpha_{t,i} = \alpha_{c,i}$, so DM's attention $\alpha_{t,i}(c) = \alpha_{c,i}$ follows the category.

Next, consider the case where $i \in M_{Kc}$. Because d is strictly convex, $S\left[(\alpha_t, \kappa_t), (\alpha_c, \kappa_c)\right]$ is a strictly concave function of $\alpha_{t,i}$. So it follows that the attention $\alpha_{t,i} \in [0, 1]$ that maximizes similarity must satisfy the following first order condition:

$$\frac{\partial S\left[(\alpha_t, \kappa_t), (\alpha_c, \kappa_c)\right]}{\partial \alpha_{t,i}} \begin{cases} = 0 \ \textit{and} \ \alpha_{t,i} \in [0, 1] \\ > 0 \ \textit{and} \ \alpha_{t,i} = 1 \\ < 0 \ \textit{and} \ \alpha_{t,i} = 0 \end{cases} . \tag{29}$$

Plugging in $\alpha_{c,i} = 1$, and defining

$$G(\alpha_{t,i}, d_{c,i}) = \frac{\partial}{\partial \alpha_{t,i}} d\left(|\alpha_{t,i} - 1|\right) + d_{c,i}$$

the first order condition simplifies to

$$G(\alpha_{t,i}, d_{c,i}) \begin{cases} = 0 \ \textit{and} \ \alpha_{t,i} \in [0, 1] \\ < 0 \ \textit{and} \ \alpha_{t,i} = 1 \\ > 0 \ \textit{and} \ \alpha_{t,i} = 0 \end{cases} .$$

Note that $G(1, d_{c,i}) = \frac{\partial}{\partial \alpha_{t,i}} d\left(0\right) + d_{c,i} \geq d_{c,i} \geq 0$, so the second case for the first order condition can never be satisfied. So if $\alpha(d_{c,i})$ is implicitly defined by

$$\alpha(d_{c,i}) = \begin{cases} 0 & if\ G(0,d_{c,i}) > 0 \\ solution\ to\ G(\alpha(d_{c,i}),d_{c,i}) = 0 & if\ G(0,d_{c,i}) \le 0 \end{cases}, \tag{30}$$

then this also characterizes the optimal attention $\alpha_{t,i}(c) = \alpha(d_{c,i})$. (Note that if $G(0,d_{c,i}) \le 0$, then since $G(1,d_{c,i}) \ge 0$, it follows that $G(\alpha(d_{c,i}),d_{c,i}) = 0$ must have some solution for $\alpha(d_{c,i})$ on $[0,1]$.)

Next, we compute $\frac{\partial\alpha(d_{c,i})}{\partial d_{c,i}}$. If $G(0,d_{c,i}) > 0$, then for all $d'_{c,i}$ in a neighborhood of $d_{c,i}$, we still have $G(0,d'_{c,i}) > 0$ and $\alpha(d'_{c,i}) = 0$. So in this case, $\frac{\partial\alpha(d_{c,i})}{\partial d_{c,i}} = 0$. If $G(0,d_{c,i}) \le 0$, then we can implicitly differentiate $G(\alpha(d_{c,i}),d_{c,i}) = 0$ with respect to $d_{c,i}$ to obtain

$$\begin{aligned} \frac{\partial^2 d(1-\alpha(d_{c,i}))}{\partial\alpha_{t,i}^2}\frac{\partial\alpha(d_{c,i})}{\partial d_{c,i}} + 1 &= 0 \\ \implies \quad \frac{\partial\alpha(d_{c,i})}{\partial d_{c,i}} &= -\frac{1}{\frac{\partial^2 d(1-\alpha(d_{c,i}))}{\partial\alpha_{t,i}^2}} < 0, \end{aligned}$$

due to the strict convexity of d. Therefore, $\frac{\partial\alpha(d_{c,i})}{\partial d_{c,i}} \le 0$ in general.

Observe that

$$\begin{aligned} d\left(t,c\right) &= \min_{\alpha_t\in[0,1]^M} \frac{\sum_{i\in M} d\left(\left|\alpha_{t,i}-\alpha_{c,i}\right|\right) + \sum_{i\in M_K} \alpha_{t,i}\alpha_{c,i}d_i\left(\kappa_{it},\kappa_{ic}\right)}{|M|+|M_K|} \\ &= 1 - \max_{\alpha_t\in[0,1]^M} S\left[\left(\alpha_t,\kappa_t\right),\left(\alpha_c,\kappa_c\right)\right]. \end{aligned}$$

Then it follows that

$$S(t,c) = F_c \cdot \left(1 - d\left(t,c\right)\right)$$

and (**??**) follows immediately.

Finally,

$$\begin{aligned}
\frac{\partial S(t,c)}{\partial d_{c,i}} &= F_c \cdot \frac{\partial \max_{\alpha_t \in [0,1]^M} \;\; S\left[\left(\alpha_t, \kappa_t\right), \left(\alpha_c, \kappa_c\right)\right]}{\partial d_{c,i}} \\
&= F_c \cdot \left[\frac{-\alpha_{t,i}(c)\alpha_{c,i}}{|M| + |M_K|}\right] \\
&\leq 0
\end{aligned}$$

where the second step follows from the envelope theorem (we can ignore the effect of $d_{c,i}$ on $\alpha_t(c)$). ■

Proof of Proposition 2. First, by Equation (8) we calculate

$$\begin{aligned}
&\frac{\partial \Pr\left(c|t\right)}{\partial S(t,c)} \\
&= \frac{\begin{array}{c}\left(\sum_{c' \in C} \exp\left[\lambda \cdot S\left(t,c'\right)\right]\right)\left(\lambda \cdot \exp\left[\lambda \cdot S\left(t,c\right)\right]\right) \\ -\left(\exp\left[\lambda \cdot S\left(t,c\right)\right]\right)\left(\lambda \cdot \exp\left[\lambda \cdot S\left(t,c\right)\right]\right)\end{array}}{\left(\sum_{c' \in C} \exp\left[\lambda \cdot S\left(t,c'\right)\right]\right)^2}, \\
&= \lambda \cdot \left[\frac{\exp\left[\lambda \cdot S\left(t,c\right)\right]}{\sum_{c' \in C} \exp\left[\lambda \cdot S\left(t,c'\right)\right]} - \left(\frac{\exp\left[\lambda \cdot S\left(t,c\right)\right]}{\sum_{c' \in C} \exp\left[\lambda \cdot S\left(t,c'\right)\right]}\right)^2\right] \\
&= \lambda \cdot \left[\Pr\left(c|t\right) - \left(\Pr\left(c|t\right)\right)^2\right] \\
&= \lambda \cdot \Pr\left(c|t\right) \cdot \left[1 - \Pr\left(c|t\right)\right].
\end{aligned}$$

Combining this with Equations (**??**) and (**??**) using the chain rule immediately yields the desired result. ■

Proof of Proposition 3. For the proof of the first part of the proposition, we omit the index j. First, note that following Proposition 2,

$$\frac{\partial \Pr(\textit{choose } a_{tc})}{\partial F_c} = \frac{\partial \Pr(c|t)}{\partial F_c} > 0,$$

so choice of a_{tc} is more likely when F_c is higher.

Next, recalling from the proof of Proposition 2 that

$$\frac{\partial \Pr(c|t)}{\partial S(t,c)} = \lambda \cdot \Pr(c|t) \cdot [1 - \Pr(c|t)],$$

and also the fact that $\frac{\partial S(t,c)}{\partial \hat{d}(c)} = -F_c$, we have

$$\frac{\partial \Pr(c|t)}{\partial \hat{d}(c)} = -\lambda \cdot F_c \cdot \Pr(c|t) \cdot [1 - \Pr(c|t)].$$

Then applying Proposition 2 again, we get

$$\begin{aligned}
\frac{\partial^2 \Pr(c|t)}{\partial d(c)\,\partial F_c} &= -\lambda \cdot \Pr(c|t) \cdot [1 - \Pr(c|t)] \\
&+ (-\lambda) \cdot F_c \cdot \frac{\partial \Pr(c|t)}{\partial F_c} \cdot [1 - \Pr(c|t)] \\
&+ (-\lambda) \cdot F_c \cdot \Pr(c|t) \cdot \left[-\frac{\partial \Pr(c|t)}{\partial F_c} \right] \\
&\propto - \Pr(c|t) \cdot [1 - \Pr(c|t)] \\
&+ (-\lambda) \cdot F_c \cdot (1 - d_{tc}) [1 - \Pr(c|t)] \Pr(c|t) \cdot [1 - \Pr(c|t)] \\
&+ \lambda \cdot F_c \cdot (1 - d_{tc}) \cdot \Pr(c|t) \Pr(c|t) \cdot [1 - \Pr(c|t)] \\
&= (-1 + \lambda S(t,c) [2\Pr(c|t) - 1]) \Pr(c|t) \cdot [1 - \Pr(c|t)].
\end{aligned}$$

So

$$\begin{aligned}
&\frac{\partial^2 \Pr(c|t)}{\partial d_{tc} \partial F_c} \leq 0 \\
\iff &\lambda S(t,c) [2\Pr(c|t) - 1] \leq 1 \\
\iff &\Pr(c|t) \leq \frac{1}{2} + \frac{1}{2\lambda S(t,c)}.
\end{aligned}$$

For the second part of the proposition, we first recall that

$$\Pr\left(c\left|t,j\right.\right)=\frac{\exp\left(\lambda\cdot F_{c}\left(j\right)\cdot\left(1-\hat{d}\left(c\right)\right)\right)}{\sum_{c'\in C}\exp\left(\lambda\cdot F_{c'}\left(j\right)\cdot\left(1-\hat{d}\left(c'\right)\right)\right)}.$$

Observe that $\hat{d}\left(c\right)$ is constant across individuals, and only depends on the category c. It is immediately clear that if $F_{c}\left(j\right)=F_{c}\left(j'\right)$ for all $c\in C$, then $\Pr\left(c\left|t,j\right.\right)=\Pr\left(c\left|t,j'\right.\right)$ for all $c\in C$. So we just need to prove the converse, which is that if $\sum_{c'\in C}F_{c'}\left(j\right)=\sum_{c'\in C}F_{c'}\left(j'\right)$ and $\Pr\left(c\left|t,j\right.\right)=\Pr\left(c\left|t,j'\right.\right)$ for all $c\in C$, then $F_{c}\left(j\right)=F_{c}\left(j'\right)$ for all $c\in C$.

For sake of contradiction, assume that we can select some $c^{*}\in C$ with $F_{c^{*}}^{j}\neq F_{c^{*}}^{j'}$. Without loss of generality, let $F_{c^{*}}^{j}>F_{c^{*}}^{j'}$. Now select some arbitrary category $c'\in C\setminus\{c^{*}\}$, and we have

$$\begin{aligned}\frac{\Pr\left(c^{*}\left|t,j\right.\right)}{\Pr\left(c'\left|t,j\right.\right)} &= \frac{\exp\left(\lambda\cdot F_{c^{*}}\left(j\right)\cdot\left(1-\hat{d}\left(c^{*}\right)\right)\right)}{\exp\left(\lambda\cdot F_{c'}\left(j\right)\cdot\left(1-\hat{d}\left(c'\right)\right)\right)}\\ &= \exp\left(\lambda\left[F_{c^{*}}\left(j\right)\cdot\left(1-\hat{d}\left(c^{*}\right)\right)-F_{c'}\left(j\right)\cdot\left(1-\hat{d}\left(c'\right)\right)\right]\right).\end{aligned}$$

Therefore

$$\begin{aligned}&\frac{\Pr\left(c^{*}\left|t,j\right.\right)}{\Pr\left(c'\left|t,j\right.\right)}=\frac{\Pr\left(c^{*}\left|t,j'\right.\right)}{\Pr\left(c'\left|t,j'\right.\right)}\\ &\implies F_{c^{*}}\left(j\right)\cdot\left(1-\hat{d}\left(c^{*}\right)\right)-F_{c'}\left(j\right)\cdot\left(1-\hat{d}\left(c'\right)\right)\\ &=F_{c^{*}}\left(j'\right)\cdot\left(1-\hat{d}\left(c^{*}\right)\right)-F_{c'}\left(j'\right)\cdot\left(1-\hat{d}\left(c'\right)\right)\\ &\implies\left(F_{c^{*}}\left(j\right)-F_{c^{*}}\left(j'\right)\right)\left(1-\hat{d}\left(c^{*}\right)\right)=\left(F_{c'}\left(j\right)-F_{c'}\left(j'\right)\right)\left(1-\hat{d}\left(c'\right)\right)\end{aligned}$$

so $F_{c^*}(j) > F_{c^*}(j')$ means that $F_{c'}(j) > F_{c'}(j')$ for any arbitrary category

$$c' \in C \setminus \{c^*\}.$$

But this means that it is impossible for $\sum_{c' \in C} F_{c'}(j) = \sum_{c' \in C} F_{c'}(j')$ to hold, so we have a contradiction. ■

Proof of Proposition 4. Because d is strictly convex, $S(t, c, \delta|\sigma)$ is a strictly concave function of $\alpha_{t,i}$. So the optimal attention $\alpha_{t,i}$ satisfies

$$\frac{\partial S(t, c, \delta|\sigma)}{\partial \alpha_{t,i}} \begin{cases} = 0 \textit{ and } \alpha_{t,i} \in [0,1] \\ > 0 \textit{ and } \alpha_{t,i} = 1 \\ < 0 \textit{ and } \alpha_{t,i} = 0 \end{cases}.$$

First let $\sigma^* = \frac{\sigma_{i\delta}}{\sigma_{ic}} \cdot \frac{1}{F_c}$. Then if (based on (20)) we define

$$\begin{aligned} & G(\alpha_{t,i}, \alpha_{c,i}, \alpha_{\delta,i}, \sigma^*) \\ = \; & \frac{\partial}{\partial \alpha_{t,i}} d(|\alpha_{t,i} - \alpha_{c,i}|) + d_{tc}(i) \cdot \alpha_{c,i} + \sigma^* \cdot \frac{\partial}{\partial \alpha_{t,i}} d(|\alpha_{t,i} - \alpha_{\delta,i}|) \\ = \; & d'(|\alpha_{t,i} - \alpha_{c,i}|) \, sign(\alpha_{t,i} - \alpha_{c,i}) + d_{tc}(i) \cdot \alpha_{c,i} \\ & + \sigma^* \cdot d'(|\alpha_{t,i} - \alpha_{\delta,i}|) \, sign(\alpha_{t,i} - \alpha_{\delta,i}), \end{aligned}$$

(where $sign(x)$ is equal to 1 if $x > 0$, -1 if $x < 0$, and 0 if $x = 0$) then the first order condition simplifies to

$$G(\alpha_{t,i}, \alpha_{c,i}, \alpha_{\delta,i}, \sigma^*) \begin{cases} = 0 \textit{ and } \alpha_{t,i} \in [0,1] \\ < 0 \textit{ and } \alpha_{t,i} = 1 \\ > 0 \textit{ and } \alpha_{t,i} = 0 \end{cases}.$$

One can easily see that $G(1, \alpha_{c,i}, \alpha_{\delta,i}, \sigma^*) \geq 0$ because

$$sign\left(1-\alpha_{c,i}\right), sign\left(1-\alpha_{\delta,i}\right) \geq 0.$$

So if $\alpha(\alpha_{c,i}, \alpha_{\delta,i}, \sigma^*)$ is implicitly defined by

$$\alpha\left(\alpha_{c,i}, \alpha_{\delta,i}, \sigma^*\right) = \\ \begin{cases} 0 & if\ G\left(0, \alpha_{c,i}, \alpha_{\delta,i}, \sigma^*\right) > 0 \\ solution\ to\ G\left(\alpha\left(\alpha_{c,i}, \alpha_{\delta,i}, \sigma^*\right), \alpha_{c,i}, \alpha_{\delta,i}, \sigma^*\right) = 0 & if\ G\left(0, \alpha_{c,i}, \alpha_{\delta,i}, \sigma^*\right) \leq 0 \end{cases},$$

then this also characterizes the optimal attention $\alpha_{t,i}(c, \delta, \sigma)$.

Now, we can compute $\partial \alpha_{t,i}(c, \delta, \sigma)/\partial \alpha_{c,i} \geq 0$. First, note that the derivative with respect to $\alpha_{c,i}$ only makes sense if i is a choice feature (because context attention is either 0 or 1). In this case, $\mathbb{I}_{Kc}(i) = 0$, and so the continuous differentiability of d implies that $G(\alpha_{t,i}, \alpha_{c,i}, \alpha_{\delta,i}, \sigma^*)$ is continuous with respect to $\alpha_{c,i}$. If $G(0, \alpha_{c,i}, \alpha_{\delta,i}, \sigma^*) > 0$, then for all $\alpha_{c,i}$ in a neighborhood of $\alpha_{c,i}$, we still have $G(0, \alpha_{c,i}, \alpha_{\delta,i}, \sigma^*) > 0$ and $\alpha(\alpha_{c,i}, \alpha_{\delta,i}, \sigma^*) = 0$. Therefore $\frac{\partial \alpha\left(\alpha_{c,i}, \alpha_{\delta,i}, \sigma^*\right)}{\partial \alpha_{c,i}} = 0$. In the case where $G(0, \alpha_{c,i}, \alpha_{\delta,i}, \sigma^*) \leq 0$, we can implicitly differentiate $G(\alpha(\alpha_{c,i}, \alpha_{\delta,i}, \sigma^*), \alpha_{c,i}, \alpha_{\delta,i}, \sigma^*) = 0$ with respect to $\alpha_{c,i}$ to get

$$\begin{aligned} & \frac{\partial \alpha\left(\alpha_{c,i}, \alpha_{\delta,i}, \sigma^*\right)}{\partial \alpha_{c,i}} \frac{\partial G\left(\alpha\left(\alpha_{c,i}, \alpha_{\delta,i}, \sigma^*\right), \alpha_{c,i}, \alpha_{\delta,i}, \sigma^*\right)}{\partial \alpha_{t,i}} \\ + & \frac{\partial G\left(\alpha\left(\alpha_{c,i}, \alpha_{\delta,i}, \sigma^*\right), \alpha_{c,i}, \alpha_{\delta,i}, \sigma^*\right)}{\partial \alpha_{c,i}} \\ = & \ 0. \end{aligned}$$

We can first compute

$$\begin{aligned}&\frac{\partial G\left(\alpha\left(\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right),\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right)}{\partial\alpha_{c,i}}\\=\;&\frac{\partial^2}{\partial\alpha_{t,i}\partial\alpha_{c,i}}d\left(\left|\alpha\left(\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right)-\alpha_{c,i}\right|\right)\\=\;&-d''\left(\left|\alpha\left(\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right)-\alpha_{c,i}\right|\right)<0,\end{aligned}$$

since d has strictly positive second derivative.

Next, we compute $\frac{\partial\alpha\left(\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right)}{\partial\alpha_{\delta,i}}$. For the same reason as before, if

$$G\left(0,\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right)>0,$$

we immediately get $\frac{\partial\alpha\left(\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right)}{\partial\alpha_{\delta,i}}=0$. If $G\left(0,\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right)\leq 0$, then we can just implicitly differentiate the equation $G\left(\alpha\left(\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right),\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right)=0$ with respect to $\alpha_{\delta,i}$ to get

$$\begin{aligned}&\frac{\partial\alpha\left(\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right)}{\partial\alpha_{\delta,i}}\frac{\partial G\left(\alpha\left(\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right),\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right)}{\partial\alpha_{t,i}}\\&+\frac{\partial G\left(\alpha\left(\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right),\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right)}{\partial\alpha_{\delta,i}}\\=\;&0.\end{aligned}$$

We already know $\frac{\partial G\left(\alpha\left(\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right),\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right)}{\partial\alpha_{t,i}}>0$, and we can compute

$$\begin{aligned}&\frac{\partial G\left(\alpha\left(\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right),\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right)}{\partial\alpha_{\delta,i}}\\=\;&\sigma^*\cdot\frac{\partial^2}{\partial\alpha_{t,i}\partial\alpha_{\delta,i}}d\left(\left|\alpha\left(\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right)-\alpha_{\delta,i}\right|\right)\\=\;&-\sigma^*\cdot d''\left(\left|\alpha\left(\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right)-\alpha_{\delta,i}\right|\right)<0,\end{aligned}$$

so we have $\frac{\partial\alpha\left(\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right)}{\partial\alpha_{\delta,i}} > 0$, as desired.

For the second part of the proposition, we'll want to compute $\frac{\partial\alpha\left(\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right)}{\partial\sigma^*}$. For the same reason as the previous two parts, if $G\left(0,\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right) > 0$, we immediately get $\frac{\partial\alpha_{t,i}(c,\delta,\sigma)}{\partial\sigma^*} = \frac{\partial\alpha\left(\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right)}{\partial\sigma^*} = 0$, which means that

$$|\alpha_{t,i}(c,\delta,\sigma) - \alpha_{\delta,i}|$$

is (weakly) getting smaller. If $G\left(0,\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right) \leq 0$, then we can just implicitly differentiate the equation $G\left(\alpha\left(\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right),\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right) = 0$ with respect to σ^* to get

$$\begin{aligned}
& \frac{\partial\alpha\left(\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right)}{\partial\sigma^*}\frac{\partial G\left(\alpha\left(\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right),\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right)}{\partial\alpha_{t,i}} \\
& +\frac{\partial G\left(\alpha\left(\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right),\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right)}{\partial\sigma^*} \\
= \; & 0.
\end{aligned}$$

As before, we already know $\frac{\partial G\left(\alpha\left(\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right),\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right)}{\partial\alpha_{t,i}} > 0$, and we can compute

$$\begin{aligned}
& \frac{\partial G\left(\alpha\left(\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right),\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right)}{\partial\sigma^*} \\
= \; & d'\left(\left|\alpha\left(\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right)-\alpha_{\delta,i}\right|\right)\cdot sign\left(\alpha\left(\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right)-\alpha_{\delta,i}\right) \\
\propto \; & sign\left(\alpha\left(\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right)-\alpha_{\delta,i}\right),
\end{aligned}$$

so we finally get that

$$\begin{aligned}
\frac{\partial\alpha_{t,i}(c,\delta,\sigma)}{\partial\sigma^*} &= \frac{\partial\alpha\left(\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right)}{\partial\sigma^*} \propto -sign\left(\alpha\left(\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right)-\alpha_{\delta,i}\right) \\
&= -sign\left(\alpha_{t,i}(c,\delta,\sigma)-\alpha_{\delta,i}\right).
\end{aligned}$$

This is equivalent to $|\alpha_{t,i}(c,\delta,\sigma) - \alpha_{\delta,i}|$ becoming smaller, as desired. ■

Proof of Proposition 5. First, we'll compute $\frac{\partial S(t,c,\delta|\sigma)}{\partial \alpha_{c,i}}$ where i is a choice feature. The envelope theorem says that we only need to evaluate the direct effect of $\alpha_{c,i}$ on $S(t,c,\delta|\sigma)$, so we get

$$\frac{\partial S(t,c,\delta|\sigma)}{\partial \alpha_{c,i}} = -\frac{\frac{\partial}{\partial \alpha_{c,i}} d(|\alpha_{t,i}(c,\delta,\sigma) - \alpha_{c,i}|)}{|M| + |M_K|}.$$

It therefore follows that

$$\begin{aligned}
\frac{\partial^2 S(t,c,\delta|\sigma)}{\partial \alpha_{\delta,i} \partial \alpha_{c,i}} &\propto -\frac{\partial^2 d(|\alpha_{t,i}(c,\delta,\sigma) - \alpha_{c,i}|)}{\partial \alpha_{c,i} \partial \alpha_{t,i}} \frac{\partial \alpha_{t,i}(c,\delta,\sigma)}{\partial \alpha_{\delta,i}} \\
&\propto d''(|\alpha_{t,i}(c,\delta,\sigma) - \alpha_{c,i}|) \frac{\partial \alpha_{t,i}(c,\delta,\sigma)}{\partial \alpha_{\delta,i}} \\
&\propto \frac{\partial \alpha_{t,i}(c,\delta,\sigma)}{\partial \alpha_{\delta,i}} \\
&\geq 0
\end{aligned}$$

where the last step follows from the first part of Proposition 4.

Next, if $\alpha_{\delta,i} = 1$, then

$$\begin{aligned}
\frac{\partial^2 S(t,c,\delta|\sigma)}{\partial \sigma_{\delta,i} \partial \alpha_{c,i}} &\propto -\frac{\partial^2 d(|\alpha_{t,i}(c,\delta,\sigma) - \alpha_{c,i}|)}{\partial \alpha_{c,i} \partial \alpha_{t,i}} \frac{\partial \alpha_{t,i}(c,\delta,\sigma)}{\partial \sigma_{\delta,i}} \\
&\propto \frac{\partial \alpha_{t,i}(c,\delta,\sigma)}{\partial \sigma_{\delta,i}} \\
&\propto -sign\left(\alpha_{t,i}(c,\delta,\sigma) - \alpha_{\delta,i}\right) \\
&\propto -sign\left(\alpha_{t,i}(c,\delta,\sigma) - 1\right) \\
&\geq 0
\end{aligned}$$

where the third step follows from the second part of Proposition 4. ■