A Cognitive Theory of Reasoning and Choice

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

February 2, 2025

Abstract

We present a theory of decisions in which attention to the features of choice options is determined by the decision maker's categorization of the current choice problem in a set of problems she solved in the past. Categorization depends on goal-relevant as well as contextual problem-level features. The model yields systematic heterogeneity in attention and choice in a given problem based on different past experiences, rigidity of choices when categorization does not change despite new data, and discontinuous shifts when changes in bottom-up salient features cause re-categorization. The model unifies major puzzles and framing effects in riskless, statistical, 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 dedicate this paper to the memory of Danny Kahneman. We thank Daniel Bartels, John Conlon, Ben Enke, Thomas Graeber, Alex Imas, Lawrence Jin, Spencer Kwon, Shengwu Li, Sendhil Mullainathan, Ryan Oprea, Joshua Schwartzstein, Jesse Shapiro, Tomasz Strzalecki, Richard Thaler, David Thesmar, Daniel Treisman, and Robert Waldmann for helpful comments. Steven Ma provided outstanding research assistance. Gennaioli thanks the European Commission for financial support though the ERC Advanced Grant (GA 101097578).

1 Introduction

People often represent and solve the same problem differently. Some see Donald Trump as a hardened criminal, others as the champion of ordinary Americans. Some view stocks as an opportunity, others as a gamble. Some view the car rental agency Avis as a loser to Hertz, others as a brand worth trying. These representations focus on different features. Trump's opponents focus on moral failures, his voters on "telling it as it is". Some investors focus on returns, others on safety. Hertz's customers focus on success, Avis's on effort.

Representations matter. Within a representation, choice is insensitive to changes in neglected features. Trump supporters are insensitive to his problems with the law. When representations change, choice becomes too sensitive to previously neglected features. Avis's advertising campaign "We are number two, we try harder" convinced many car renters to switch from Hertz to Avis not because of new facts but by changing representations. Where do representations come from? How do they affect choice? And why do they change, including based on irrelevant features?

We offer a general theory of choice in which a representation, defined as attention attached to particular features of choice options, is set in two stages. First, the DM fits the current problem into a category of frequent and similar past problems. Second, the DM attends to the features that matter in that category and neglects the others, which affects choice. Due to differences in past experiences, people can fit the same problem into different categories, causing choice heterogeneity. Increasing similarity to a given past problem causes a common category to be cued for many people, leading to choice instability. This mechanism unifies famous biases and framing effects across riskless, statistical, and risky choice problems, and yields many new predictions.

Categories play a key role in perception and recognition. In the duck-rabbit illusion, some people categorize the image as a duck and attend to the beak, others categorize it as a rabbit and attend to the mouth. Both representations are possible, because each i) recalls a frequently experienced category, and ii) focuses on a feature that is similar to that category. Likewise, the image "/-\ " is categorized as "A" in C/-\ T and as "H" in T/-\ E. The adjacent letters retrieve a familiar word and focus the DM's attention on features similar to that word. In both examples, categories are not integrated because they rely on different experiences, and re-categorization triggers jumps in attention and in choice.

Similar effects arise in economic problems. To value vertically differentiated jams, a consumer decides whether the choice is "a treat" or "a staple". Fancy packaging of jams cues the former, due to similarity to past treats. The consumer then focuses on quality, as she does with treats, and neglects price. Ordinary packaging instead promotes the "staple"category, and a focus on price. Also in this example, categories compete because they are segregated in memory. We feel the pain of paying at the supermarket but the pleasure of treats at home (or imagine them in fancy shops), not together. The context of different experiences cue different representations. Re-categorization again triggers a change in choice.

Section 2 presents a general model of this mechanism. A choice problem consists of options, which are vectors of choice features (price, quality, etc.), and of a vector κ_t of context features common to all options (the choice set, prices, location, etc.). The endogenous object is the problem's representation: a vector of attention weights α_t for choice and context features. There is a set of categories, each collecting a set of past problems and summarized by a typical context vector κ_c and representation α_c. In the first step the DM's selects the category that is most frequent or whose context κ_c is most similar to κ_t, based on attended-to features. In the second step, the selected category shapes the DM's attention to the features of choice options, and thus choice. Section 5 incorporates bottom-up salience in the model (e.g., Bordalo, Gennaioli, and Shleifer [19], Bushong, Camerer, and Rangel [23]), which is key for generating framing effects.

Section 3 analyzes the model. Consider a DM choosing between two riskless categories: "consuming", focused on pleasure, and "buying", focused on price. The model yields three properties. First, frequent past use of a category promotes its use even if current context κ_t differs along some entries from κ_c. This causes distortions: a person who has experienced poverty, and is familiar with the pain of paying, categorizes many choices as “buying”, and focuses more on price than a person with the same income but without the same experiences, who relies more on “consuming”, and focuses on pleasure. Unlike with rational inattention or Bayesian learning, price elasticity for “buying” is high even if the consumer is not poor anymore or the benefits of spending are huge, as with medical out-of-pocket costs (Baicker, Mullainathan, Schwartzstein [5], Chandra, Flack, Obermeyer [25]). A feature that was important in many past choices, in this example opportunity cost, is used for categorization and extrapolated to be important now, creating choice heterogeneity unrelated to hedonics (Handel and Schwartzstein [62]).

Second, changing context can change similarity, causing instability (Tversky and Kahneman [135], Enke and Zimmermann [38]). A festivity can cue a poor person to exceptionally categorize choice as “consuming”, focusing on pleasure and neglecting price. The previously poor person can switch from “buying” to “consuming” when choosing new goods (e.g., i-Phones), because these differ from their past “buying” problems. A fancy coffee shop leads consumers to categorize coffee as a daily treat, not as a staple. Instability occurs when a feature in κ_t, even if irrelevant, changes similarity to κ_c. The DM is insensitive to a feature neglected in the current category and highly sensitive to the same feature when category switches, producing under and over-reaction to information, respectively.

Third, a bottom-up salient feature favors a switch to a category in which this feature is relevant, reallocating attention across all features. This mechanism leads to framing effects: increasing the descriptive salience of a feature, without providing any new information, sharply changes representations and valuation. An advertisement showing the jam on a beautiful breakfast table prompts retrieval of the “consuming” category, enhancing focus on pleasure and reducing that on price.

Sections 4 and 5 show that these three properties explain famous puzzles in riskless, statistical, and risky choice, yielding new predictions. The analysis relies on four intuitive categories, two focused on payoffs ("consuming" and "buying"), and two focused on statistical features (inference and random draws). Consider opportunity cost neglect and non-fungibility in mental accounting (Thaler [126]). When thinking whether to splurge, some people retrieve frequent "consuming" experiences and neglect opportunity costs. If however spending comes from a "rainy day" account, salient opportunity costs trigger a "buying" category, and hence a price focus. When judging the relative probability that a fair coin produces *hthhtth* versus *hhhhhh* toss sequences, many people retrieve experiences of inferring the bias of a coin. They then focus on the share of heads, which is highly relevant in inference, and neglect individual flips, committing the Gambler's Fallacy. In both statistical and riskless choice, errors arise because the DM selects which features she considers relevant based on similarity with a familiar or similar past problem.

Lottery choice, which requires considering both payoffs and their probabilities, entails a competition between categories in *different domains*: a riskless category focused on payoffs and a statistical category focused on random draws. This insight yields many well-known puzzles, framing effects, as well as striking new predictions. For example, relying on a payoff-focused "consuming" category generates insensitivity to probabilities, leading to the certainty effect and the fourfold pattern of risky choice (Kahneman and Tversky [75]). Crucially, the mechanism also explains why the same fourfold pattern arises for riskless mirrors (Oprea [103]): indeed, both types of problems entail the common "consuming" representation that neglects probabilities in the first case and frequencies in the second.

We contribute to a large body of work. Classic behavioral theories trace heterogeneity to different biases and instability to changes in reference points. They do not explain, however, where biases come from or why reference points change with irrelevant context, as in framing effects.[1] A more recent approach studies top down

[1] These theories cannot likewise explain weak within-person correlation of choices both within a domain, e.g., low correlation between insurance demand and lottery choice (Barseghyan, Prince, and Teitelbaum [6]) and across domains, e.g. low correlation between the endowment effect and aversion to mixed lotteries (Chapman, Dean, Ortoleva, Snowberg, and Camerer [26]).

attention shaped by goal optimality (e.g., Sims [121], Gabaix [46]) or priors (e.g., Schwartzstein [116], Gagnon-Bartsch, Rabin, and Schwartzstein [48]). A related approach studies insensitivity due to noisy perception (Woodford [143]) or decision uncertainty (Enke and Graeber [37]). These theories deliver neither the heterogeneity nor the instability that we see in choice data.[2] Bordalo, Conlon, Gennaioli, Kwon and Shleifer [21] show coexistence of under and over-reaction in the same inference problem and common shifts to over-reaction after irrelevant changes. Ba, Bohren, and Imas [3], and Bohren, Imas, Ungeheuer, and Weber [13]) also show striking forms of instability that relate to shifts in representation.

In our theory, both insensitivity to data and instability are driven by the same forces of memory and attention. Relative to Bordalo, Conlon, Gennaioli, Kwon and Shleifer [21], who formalize attention to features in statistical problems, we offer a domain-general theory that combines key drivers of attention: a top down "problem recognition" stage and bottom up salience, consistent with the psychology of similarity perceptions (Nosofsky [101], Tversky [132],[133]) and of top down attention (Itti and Baldi [68], Awh, Belopolsky, and Theeuwes [2]). We explain heterogeneity of "biases" via differential familiarity with categories, and instability via changing similarity along bottom-up salient features. Our mechanism is linked to bounded rationality (Simon [120]) with the key difference that errors are not deliberate approximations to a complex problem. Categorization can in fact complicate computations rather than simplify them (as in the Gambler's Fallacy), and can cause over-confident errors due to reliance on a familiar but wrong category.

Our focus on categories and experiences follows much psychology (e.g., Mack and Palmeri [90], Reed [106], Rosch and Lloyd [108]). In case-based learning (Schank [115], Gilboa and Schmeidler ([52]) and habitual decisions (Laibson [82]), people choose actions that performed well in the past, including in similar contexts. Mullainathan [96] models categories as Bayesian updating with coarse types, so a person's probability judgment changes discretely only when categories are crossed. In

[2] Models of noisy perception (Woodford [143]) offer a cognitive theory for diminishing sensitivity in stimulus contrast, which is complementary to our approach.

our model, categories pin down not a belief or action but a representation: the features to which the DM is sensitive or insensitive, including in problems where the DM has all the data (e.g., both lottery payoffs and probabilities). Moreover, categorization is not Bayesian: it depends on attention and similarity which vary across people and change with bottom-up salience, yielding heterogeneity and framing effects.[3]

Recent work on memory studies selective retrieval of information about the value of the features of choice options (e.g., Bordalo, Conlon, Gennaioli, Kwon, and Shleifer [21], Fudenberg, Lanzani, and Strack [44]-[45], Wachter and Kahana [141], Bordalo, Gennaioli, and Shleifer [18]). In our model feature values are known, but memory shapes which features are deemed relevant to the current problem. Both mechanisms play a role in choice, and future work may study them together.

2 The Model

Even a simple problem such as buying jam has many features (taste, price, "is it spoiled?", etc.), only some of which are attended to in choice. We present a theory in which: 1) the DM selects a representation, i.e. what to attend to, and 2) attention drives choice. In stage 1 attention is shaped by categorization. If the DM thinks about consuming, which occurs removed from paying, she focuses on taste and underweights price compared to when she thinks about buying. For many goods, risk is shrouded or rarely materializes. For others, such as flying, it is attended to.

[3] In Salant and Rubinstein [111] instability is due to the use of different choice functions for different frames. Ellis and Masatlioglou [34] axiomatize utility that is stable within but not across categories of choices. Important economic work on categorization includes Mohlin [95] and Jehiel [69]. Evers, Imas, and Kang [40] present and test a model of editing based on loss and gain categories. In psychology, the ALCOVE model by Kruschke ([81], [133]), and ADDCOVE by Verguts, Ameel, and Storms [140] formalize the assignment of multidimensional stimuli to categories. Our key innovation, also relative to these papers, is to view categories as shaping representations and to endogenize similarity based on top-down and bottom-up attention.

Categories shape choice because valuation in stage 2 depends on attention to a range of payoffs and risks. To see this, consider a lottery o delivering a good with hedonics (u_{1s}, u_{2s}), e.g. 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:

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

Inattention $\alpha_x < 1$ causes insensitivity to a feature. Riskless choice entails a sure event s and the corresponding payoffs. With quality q and price p, Equation (1) yields weighted utility $\alpha_Q \cdot q - \alpha_P \cdot p$, as in Bordalo, Gennaioli, and Shleifer [17].

A statistical problem entails estimating the probability of event H, to which payoff 1 is attached, with zero payoff outside. If $H =$"green (g) ball from urn A," Equation (1) becomes $\mathbb{P}\left(A\right)^{\alpha_U} \mathbb{P}\left(g|A\right)^{\alpha_B}$, akin to Grether's [58] formula.

Risky choice combines 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 [16] with probability weights as in Prospect Theory (Kahneman and Tversky [75]).

Section 2.1 models stage 2 in a general setup nesting Equation (1) and lays out the features of choice options and context. Section 2.2 links features to categories and describes stage 1: how features shape categorization.

2.1 How Attention Causes Valuation

A problem's primitives are: i) a menu of options, ii) a set of features, and iii) an attention-based valuation function.

Menu of Options. There is a nonempty finite menu of lotteries O and a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ known to the DM. 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-weighted combination of its hedonics and event probabilities. In turn, the value of o is the sum of the values of its atoms. Riskless choice and statistical hypotheses are special cases. A case in which a single option must be evaluated (e.g. a politician) is also a special case.

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 $u_i \in \mathbb{R}$ of hedonic $i \in M_H$ reports the feature's utility impact. Subvector e reports *event* features: delivery states for hedonics, e.g., $e_i \in \{urn\ A, urn\ B\}$ or $e_i \in \{jam\ spoiled,\ not\}$. Each event feature $i \in M_E$ identifies a partition of the state space Ω. Its value e_i reports the cell of the partition to which the atom belongs. Event features have no direct utility impact.

Attention and Valuation. Hedonics and events can vary across options. We call them *choice features*, $M_O = M_H \sqcup M_E$. Attention to them is captured by a vector of weights $\alpha_O \in [0,1]^{M_O}$ that do not need to add up to 1. Feature $i \in M_O$ is fully weighted if $\alpha_i = 1$, underweighted if $\alpha_i \in (0,1)$, and edited out if $\alpha_i = 0$. An edited out event ($i \in M_E$ with $\alpha_i = 0$) is perceived as $e_i(\alpha) = \Omega$: any of its realizations is allowed. A non-edited out event is perceived correctly, $e_i(\alpha) = e_i$. The attention-based value of hedonic $i \in M_H$ is:

$$u_i(\alpha_O) = \alpha_i \cdot u_i + (1 - \alpha_i) \cdot \overline{u}_i, \quad \forall \alpha_O \in [0,1]^{M_O} \tag{2}$$

where $\overline{u}_i$ is average value across all atoms, $\overline{u}_i = \frac{\sum_{o \in O} \sum_{(u,e) \in o} u_i}{\sum_{o \in O} |o|}$. Inattention $\alpha_i < 1$ shrinks perception toward $\overline{u}_i$. A fully neglected feature, $\alpha_i = 0$, is perceived as $\overline{u}_i$ for all options, so that neglected features do not affect choice.

As in Equation (1), valuation of an option is based on the perceived features of its atoms, which depend on attention to choice features α_O. For each atom $y \in \cup_{o \in O} o$, let $y(\alpha_O) = (u(\alpha_O), e(\alpha_O))$ contain its perceived choice features. The value of $y(\alpha_O)$ multiplies its perceived probability and hedonics:

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

In the first bracket, the perceived probability of $y(\alpha_O)$ is the attention-weighted chain product of event probabilities, which follows a linear order $<$ over events based on the sampling process, e.g., first select an urn and then extract a ball from it.[4] The second term is the hedonic value in $y(\alpha_O)$. The value of an atom 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 $o(\alpha_O) = \{y(\alpha_O) : y \in o\}$ adds 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 normatively relevant if it affects $v(o(\alpha_O))$ when $\alpha_O = 1$. Whether or not a feature is normatively relevant depends on the problem.

Decision Rule. A vector of valuations $v \in \mathbb{R}^O$ fully determines choice in the admissible set A. For goods or lotteries the DM picks the highest valued option, in some statistical problems she computes the relative value of two lotteries-hypotheses, yielding their relative probability. In judging similarity between A and B, she aggregates their similarities in specific features (see Appendix A.3).

Valuation is shaped by choice features. The problem as a whole is categorized based on context features. Being common to options, they do not directly affect

[4] The DM can compute the probabilities in the product. If the sampling process is unspecified, $<$ reflects the DM's beliefs. The order is irrelevant if $\alpha_i \in \{0, 1\}$ for every $i \in M_E$.

valuation, but allow the DM to compare the current problem to past problems.

Context Features are partly derived from choice features. Context includes the choice set, describing available goods and their prices or the hypotheses to estimate and the probabilities of some events. These features cue past problems with similar parameters. But context also includes irrelevant aspects such as time, location, etc. If associated with specific past problems, these features may influence how the current problem is represented.

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 feature, the values it attains in the choice set (the prices of jams, their qualities, etc.). Event context κ_e reports, for each event feature, the values it attains in the choice set. The *situation* Z reports non-good specific dimensions such as time, location, etc. (e.g., today is a festivity).[5] Each feature $i \in M_K$ has an associated distance d_i measuring perceived dissimilarity between two possible realizations of context.

Problems faced over time are indexed by t. Current context at t is $\kappa_t = (\kappa_{t,u}, \kappa_{t,e}, \{z_t\})$. It affects categorization only if the DM attends to it, with $\alpha_K \in [0,1]^{M_K}$ reporting attention to context features. The current problem's representation is attention to its choice and context features, $\alpha_t = (\alpha_{O.t}, \alpha_{K,t})$. Overall, the problem at t is summarized by (α_t, κ_t).

2.2 Categories and Representations

The DM's database at time $t \in \mathbb{N}$ is partitioned into a set of categories C of past problems that have a similar representation. Formally, category c is summarized by a vector of attention and context, (α_c, κ_c), as for individual problems, and by their temporally discounted frequency in the database, $F_c \in \mathbb{R}_+$.[6] Vector α_c captures the

[5] 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).

[6] Recency-weighted frequency is $F_c = \sum_{\tau \in c} \delta^{t-\tau}$, $\delta \in (0,1)$. The "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

prototypical attention to features in problems belonging to c. 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 [108]). Attention to choice features $\alpha_{c,O}$ describes the sensitivity of valuation to choice features in c.

We cover basic decision problems using four categories. The first pair captures "riskless" experiences, in which the DM evaluated a good's hedonics in a specific state. The second pair captures "statistical" experiences, in which she estimated probabilities. These categories are "building blocks": choice is "rational" when the DM correctly uses them in the problems we consider.

Riskless Categories are "consuming" and "buying". In the consuming category, *con*, the DM evaluates goods focusing on qualities, not prices. This focus is encoded in the attention to hedonics subvector $\alpha_{con,O}$. Context specifies a set of experienced qualities Q_{con} and situations Z_{con} (e.g., being at home, where price is not prominent). The diagnosticity of these two context features is captured by the attention to context subvector $\alpha_{con,K}$.

In the buying category *buy*, attention to choice features $\alpha_{buy,O}$ focuses on price, but partly also on the goods' typical quality; otherwise we would not buy. Context reports experienced qualities Q_{buy}, prices P_{buy}, and situations Z_{buy} (e.g., being in a shop). The diagnosticity of these context features is captured by $\alpha_{buy,K}$.

Statistical Categories are "frequency estimation" and "agnostic inference". Frequency estimation, category $freq$, refers to experiences of estimating the probability of a single draw from a known process, e.g., the probability that a fair coin lands h or t. Attention $\alpha_{freq,O}$ focuses on the event corresponding to the hypothesis, h or t. Diagnostic category features include: i) there is a single draw, and ii) hypotheses coincide with the outcomes of that draw.

Agnostic inference, category inf, refers to experiences of judging a data generating process (DGP) based on i.i.d. signals without having prior information about it, e.g., assessing the quality of a restaurant based on a few reviews. Attention

every $i \in M_K$, $\kappa_{c,i}$ minimizes some discrepancy from past contexts $(\kappa_{\tau,i})_{\tau \in c}$.

$\alpha_{inf,O}$ focuses on the share of positive signals, which is a sufficient statistic for the DGP, and not on the prior. Diagnostic category features include: i) there are at least two draws (selection of the DGP and signals) and ii) hypotheses coincide with different DGPs (e.g., the restaurant is good or bad).

Like in Mullainathan's [96] categories, in our case prototypes (α_c, κ_c), are given. For our purposes, we assume that weights α_c reflect payoff relevance and sensory prominence in categorical experiences. For example, when "consuming" qualities are relevant and prominent, attention focuses on them. When "buying" prices are also relevant and prominent, so attention focuses on them also. Because categories focus on some relevant outcomes, decisions are good if categories are applied to the proper problems. Errors instead mostly arise due to the DM's: i) use of a wrong category, and ii) failure to integrate categories.[7] Because similarity depends on measurable features (e.g. the number of draws or knowledge of the DGP in a statistical problem, the price paid and the spatiotemporal distance between buying and consuming in choice among goods), our model yields testable predictions for when it is likely that a category is misused and errors arise.

There are many more categories than the four we use. In risky choice, natural categories entail "loss/regret" or "gain/elation" in which attention is focused on these different features of experience. Yechiam and Hochman [144] suggest an attention based explanation of loss aversion. In intertemporal choice, a natural category is one of "investment" decisions, in which attention is focused on long-term payoffs.[8] Our model also nests standard categories of objects as "recognition problems", in which the DM must determine whether a target is sufficiently similar to the category prototype (e.g. to the diagnostic features of the set of objects labeled

[7] Treisman and Gelade [130] famously showed that feature integration is more difficult and time-consuming than considering each feature separately. Barbara Tversky [139] documents misaggregation of spatial knowledge.

[8] We could add a "no clue" category in which the DM pays attention to nothing, being indifferent across options. This category sets a minimum similarity threshold below which no other category is used. Reliance on this category can cause, with forced choice, low confidence.

"chair" in the past).[9] We leave the analysis of these categories to future work. In the conclusion we discuss how categories are formed and can be measured.

Categorization and attention are jointly determined. A function S measures the similarity $S\left[\left(\alpha_t, \kappa_t\right), \left(\alpha_c, \kappa_c\right)\right]$ between the problem (α_t, κ_t) and the prototype (α_c, κ_c) of category c. Similarity decreases if: i) context κ_t is different from κ_c, so the problems are of different nature, or if ii) attention α_t differs from α_c, so the DM approaches the problems differently. We use the separable form:

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

where $d : \mathbb{R}_+ \to \mathbb{R}$ is strictly increasing, strictly convex, twice continuously differentiable, with $d'(0) = 0$, $d(1) = 1$. This functional form helps tractability, but our results rely on the idea – central in psychology (Nosofsky [101]) – that similarity falls in differences, and the more so when these are more attended to.

To represent κ_t the DM matches each category and selects the one with resulting maximal (perturbed) similarity.

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 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:

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

Endogenous attention weights in Equation (6) allow for a self-confirming representation as in duck-rabbit.

[9] Compared to existing models of categories of objects (Mullainathan [96]), key to our approach is not coarseness of categories, but rather heterogeneous and unstable categorization driven by experiences and spurious context, contrary to categories derived from first principles (as in Kant [76]) or through a possibly restricted Bayesian approach.

Categorization. Following in model assignment tasks in psychology (Mack and Palmeri [90]), the DM chooses the category $c \in C$ by maximizing similarity

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

where ϵ_c is a type I extreme-value random shift with scale parameter λ, reflecting random attention to context. The error structure can be more general, but this formulation allows for convenient closed forms.

Matching and categorization pin down the stochastic representation $\alpha_t(c)$ of the current problem, which entails stochastic choice. In Woodford ([143]), stochasticity is due to noise when perceiving hedonics. In Enke and Graeber ([37]), it is due to noise on the optimal action. Here, it reflects dilemmas as to which features are relevant.[10] The equilibrium representation then depends on two factors: the frequency F_c of a category in the database and the distance between current and category context (κ_t vs. κ_c). Representations of a given problem thus exhibit experience-driven heterogeneity and instability from changes in context.[11]

3 Representations, Attention and Choice

We next consider the model's implications for categorization and choice.

3.1 Equilibrium Representation and Attention

When matching c, the DM tunes attention to a feature $i \in M_K$ to satisfy, in an interior equilibrium, the first order condition for Equation (6):

[10] Often categorization is spontaneous, so we are unaware of the dilemma, such as when we first see the duck rabbit. If the DM is aware of the dilemma, the adherence to the selected category may be influenced by meta-cognitive factors, such as confidence as in Enke and Graeber [37].

[11] Psychologists have documented both excessive reliance on a decision model, which is called "overgeneralization", and the failure to apply a known correct model when context changes, which is called "limited portability of knowledge"; see Bassok [8]. In our model, these forces emerge due to similarity and frequency-based retrieval of categories.

$$\frac{\partial}{\partial \alpha_{t,i}} d\left(|\alpha_{t,i} - \alpha_{c,i}|\right) + d_i\left(\kappa_{t,i}, \kappa_{c,i}\right) \cdot \mathbb{I}_{\{1\}}(\alpha_{c,i}) = 0, \tag{7}$$

where $\mathbb{I}_{\{1\}} = 1$ if feature i is diagnostic of c, and zero otherwise. Adapting attention to the category, namely reducing $|\alpha_{t,i} - \alpha_{c,i}|$, tends to increase similarity but may backfire along a discrepant diagnostic feature, $d_i\left(\kappa_{t,i}, \kappa_{c,i}\right) \cdot \mathbb{I}_{\{1\}}(\alpha_{c,i}) > 0$. For choice features, $i \in M_O$, only the leftmost term in (7) is non-zero.

Proposition 1 *When matching the problem with category c, attention to a discrepant diagnostic feature $d_i\left(\kappa_{t,i}, \kappa_{c,i}\right) \cdot \mathbb{I}_{\{1\}}(\alpha_{c,i}) > 0$ is shrunk toward zero, $\alpha_{t,i}(c) \leq \alpha_{c,i}$, the more so the higher is $d_i\left(\kappa_{t,i}, \kappa_{c,i}\right)$. Attention fully adapts to the category otherwise, $\alpha_{t,i}(c) = \alpha_{c,i}$.*

To match the problem with c, the DM neglects discrepant diagnostic context. When choosing a jam, to match the problem with "consuming" the DM must neglect the current supermarket location, which is diagnostic of "buying". Full neglect entails similarity cost $d\left(|0 - 1|\right) = d(1) = 1$. Upon matching with c, attention to choice features then fully adapts to the category. When representing choice as "consuming", the DM attends to jam qualities and underweighs prices. Valuation is insensitive to features that are not relevant in the selected category.

By Proposition 1 endogenous similarity to c is $S(t, c) = F_c\left(1 - d(t, c)\right)$, where

$$d(t, c) = \min_{\alpha_t \in [0,1]^{M_O \cup M_K}} \frac{\sum_{i \in M} d\left(|\alpha_{t,i} - \alpha_{c,i}|\right) + \sum_{i \in M_K} \alpha_{t,i}\alpha_{c,i} d_i\left(\kappa_{t,i}, \kappa_{c,i}\right)}{|M| + |M_K|}$$

is the minimized distance from c. Lemma 6 in the Appendix shows that $S(t, c)$ increases in the DM's familiarity F_c with c and decreases in discrepancy between attended to context $d_i\left(\kappa_{t,i}, \kappa_{c,i}\right)$. These properties pin down the DM's representation, which is her equilibrium attention $\alpha_{t,O}(c)$ to choice features.

Proposition 2 *The DM's representation satisfies*

$$\Pr\left(\alpha_{t,O} = \alpha_{t,O}(c)\right) = \Pr\left(c\,|t\right) = \frac{\exp\left(\lambda \cdot F_c \cdot \left(1 - d\left(t,c\right)\right)\right)}{\sum_{c' \in C} \exp\left(\lambda \cdot F_{c'} \cdot \left(1 - d\left(t,c'\right)\right)\right)} \qquad \forall c \in C. \quad (8)$$

Problem t is more likely to be represented using category c when:

i) category c was used more frequently (and recently), $\partial \Pr\left(c\,|t\right)/\partial F_c = \lambda \cdot \Pr\left(c|t\right) \cdot \left[1 - \Pr\left(c|t\right)\right]\left[1 - d\left(t,c\right)\right] > 0$.

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

The representation is more likely to use a more frequently used category c, higher F_c, and less likely to use a discrepant ones, higher $d\left(t,c\right)$. Reliance on frequency and context often promotes good representations. A DM highly trained on a specific problem such as the probability that a fair coin lands h, perfectly recognizes it from the description and reaches the correct answer of 50%.

Sometimes, however, these forces block a more fitting category c', causing attention distortions and error. Repeated discussions of political corruption in Washington may render a "honesty assessment" category frequent, causing voters to neglect competence. A sporting event or a referendum, being similar to national pride contexts, may temporarily increase purchases of goods associated with the country's flag and cause neglect of typically attended to price or quality (Nardotto and Sequeira [99]).

3.2 Choice

As the DM represents the problem using c, she adopts the category's attention to choice features $\alpha_{c,O}$ and chooses $a_{tc} \in A$. Assuming injectivity of the map from categories to induced choices, three properties follow.

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

$$\Pr\left(a_{tc}\right) = \Pr\left(\alpha_{t,O} = \alpha_{t,O}(c)\right) = \Pr\left(c|t\right) \qquad \forall c \in C. \qquad (9)$$

Furthermore, for all $c \in C$:

i) Higher F_c increases $\Pr(a_{tc})$ and weakly decreases $\Pr(a_{tc'})$ for all $c' \neq c$. Take two DMs j and j' with $\sum_{c' \in C} F_{c'}^{j} = \sum_{c' \in C} F_{c'}^{j'}$. They choose $a_{tc} \in A$ with different probabilities for some $c \in C$ if and only if $F_c^j \neq F_c^{j'}$.

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

iii) Increasing similarity to c, i.e. decreasing $d(t, c)$, boosts $\Pr(a_{tc})$ more at higher F_c if and only if c is not dominant, i.e., $F_c < F_c^*$ where the threshold increases in the distance $d(t, c)$. In particular, this always holds if $\Pr(c|t) \leq 1/2$.

In (9) attention and choice are stochastic due to shock ϵ_c. Noise-driven stochastic choice (Woodford [143], Enke and Graeber ([37]) yields a valuation distributed around a single mode. By changing feature integration, categorization instead yield multi-modal and unstable valuations. Multi-modality and instability are observed in inference problems: some people anchor to the base rate, others to the likelihood, and a change in representation causes the same person to switch from one mode to another (Bordalo, Conlon, Gennaioli, Kwon, and Shleifer [21]). Different modes are often insensitive to one piece of information, but extreme sensitivity to previously neglected data causes instability. Such valuation shifts cannot be explained by noise.

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 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 [21]). This mechanism can be tested by measuring or experimentally manipulating experiences.

Changes in context cause instability (Point *ii*), even if spurious. Higher discrepancy $d(t, c)$ reduces the use of c and the probability of choosing a_{tc}. Describing the same inference problem as taxicabs versus balls and urns changes the focus on different statistics, not information. Combined with bottom-up salience in Section

5, this mechanism yields a theory of framing: different descriptions of the same problem alter choice. This prediction can be tested by measuring how changing context affects perceived similarity between problems.

Crucially, we endogenize the strength of framing effects, defined as the extent to which increasing contextual similarity with c fosters choice of a_{tc}. It is inverse-U shaped in the DM's baseline familiarity F_c with this category (Point *iii*). If F_c is not too high, framing effects are stronger for more familiar categories. This complementarity explains mental simulation, which is a key driver of probability estimates (Schacter, Addis, Hassabis, Martin, Spreng, and Szpunar [114], Bordalo, Burro, Coffman, Gennaioli, and Shleifer [14]): past experiences have a stronger impact on current estimates when they are more similar to the current problem.[12],[13]

If instead F_c is very high, the impact of priming c decreases with F_c, because c is already a dominant and hence stable representation for the current problem. Dominance and stability arise in "rational" conditions, when a DM has has often used a well-fitting category. But they may sometimes arise even if a category fits poorly, provided the DM is very familiar with it. In this case, stability is due to overgeneralization and neglect of informative data, as in cognitive dissonance (Festinger [41]) or the confirmation bias (Nickerson [100]). Highly religious people may see many choices from the perspective of their values and neglect potentially important features.

[12] Bordalo, Burro, Coffman, Gennaioli, and Shleifer ([14]) 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 [125] on inflation.

[13] 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 [64]). Describing default on a loan as contrary to sacred texts increases repayment (Bursztyn, Fiorin, Gottlieb, and Kanz [22]), 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.

4 Representations in Famous Puzzles

We show that categorization of choice as consuming versus buying explains systematic variation in price elasticity and various forms of mental accounting (Section 4.1). Categorization of statistical problems as frequency estimation versus inference explains the Gambler's Fallacy, its instability, and biases in inference (Section 4.2). These puzzles are unified by "overgeneralization": namely excessive attention paid to features that were highly relevant in a set of familiar past problems. Instability arises when changes in context causes a category switch, refocusing attention toward previously neglected features.

4.1 Consumer Choice

Two goods g and b have expected qualities $q_g \geq q_b$ and prices $p_g \geq p_b$. The consumer's utility from hedonics is $q_l - \eta \cdot p_l$ for $l \in \{g, b\}$, where $\eta > 0$ maps dollars into utils, reflecting opportunity cost. Consumer choice involves risk (is the jam spoiled? Will price increase in the future?), so we allow for low probability shocks to quality and price Δq_l and $-\eta \cdot \Delta p_l$ before consumption. The associated events are $e_Q \in \{e_{Qn}, e_{Qs}\}$ and $e_P \in \{e_{Pn}, e_{Ps}\}$, where e_{Q_n} and e_{P_n} reflect normal quality and price, while e_{Qs} and e_{Ps} reflect shocks. These features are all accessible to the consumer, Section 5 allows for some of them, e.g., shocks, to be shrouded.

Consider a consumer choosing ex-ante between g and b. Context includes qualities $Q_t = \{q_g, q_b\}$, prices $P_t = \{p_g, p_b\}$, shocks $\Delta Q_t = \{\Delta q_g, \Delta q_b\}$ and $\Delta P_t = \{\Delta p_g, \Delta p_b\}$, and events $e_{Q,t}$ and $e_{P,t}$. It also includes the situation z_t, e.g., whether choice is at the supermarket or in a fancy shop. Albeit multi-dimensional and risky, choice is represented using lower-dimensional categories: consuming or buying. These categories are "riskless": they fully neglect events/probabilities. For simplicity, and without changing the key results, we first assume that shocks are zero, $\Delta q_{g,b} = \Delta p_{g,b} = 0$. We allow for shocks in Section 4.1, to study "mental accounting", where choices are made after shocks are realized. In Section 5 we study lottery lottery choice and also consider attention to events.

The diagnostic features of the consuming category *con* include qualities Q_{con} and situations Z_{con}. In these experiences, the DM was focused on enjoying realized qualities, while paid prices were less salient. Thus, attention to choice features satisfies $\alpha_{con,Q} = \alpha_{con,\Delta Q} = \overline{\alpha} > \alpha_{con,P} = \alpha_{con,\Delta P} = 0$.[14]

The diagnostic features of the buying category *buy* include the qualities Q_{buy}, prices paid P_{buy}, and the situations Z_{buy}. In these experiences, the consumer imagines consumption pleasure but is focused especially on the salient price paid, to a lesser extent on expected quality, and neglects unlikely shocks, $\alpha_{buy,P} = \overline{\alpha} > \alpha_{buy,Q} = \underline{\alpha} > \alpha_{con,\Delta Q} = \alpha_{buy,\Delta P} = 0$.

Neither *con* nor *buy* alone attaches proper attention to relevant features, but integrating high attention to quality in *con*, with high attention to price in *buy* would retrieve rational valuation. These categories however compete, and are selected based on their familiarity and on the distances between the current problem and diagnostic category features. For many goods distances between quality and price hedonics are small, so discrepancies arise mostly based on the situation.[15]

In the first stage the DM tunes attention to match and then selects category *con* or *buy* based on the stochastic rule in Equation (8). With probability $\Pr(buy\,|t)$ the consumer represents the problem as buying, and with probability $\Pr(con\,|t) = 1 - \Pr(buy\,|t)$ as consuming. In the second stage, the DM values g and b using the attention weights in the selected category. The expected value difference between the two goods is then given by:

$$v(g) - v(b) = \Pr(buy\,|t)\left[\underline{\alpha}\,(q_g - q_b) - \eta\overline{\alpha}\,(p_g - p_b)\right] + \Pr(con\,|t)\,\overline{\alpha}\,(q_g - q_b)\,. \quad (10)$$

[14] Full neglect of prices in the consumption category captures in a stark way the fact that price is often not prominent in consumption situations (e.g. when purchase occurred at an earlier time). This is a specific instance of the role of bottom up attention discussed in Section 5. Qualitatively, the results go through if we allow for some attention to price paid during consumption as well.

[15] We could set zero distance along a context feature if in κ_t such feature is a subset of the category feature (e.g., seeing prices encountered in the past perfectly fits with *buy* price context).

Because both representations neglect some features, the average consumer can be insensitive to q and p compared to a fully attentive consumer. Different consumers, however, adopt different representations, each of which exhibits excess *relative* sensitivity to the category-relevant feature compared to the irrelevant one: consumers who represent their choice as buying are too sensitive to price relative to quality, causing under-valuation of the high quality good g relative to b. Consumers who represent choice as consuming are too sensitive to quality relative to price, causing over-valuation of g relative to b.

Some differences in representations reflect experiences F_{buy} and F_{con}. Others reflect the random shock ϵ_c or changes in irrelevant context, such as a situation (being in a store) that increases similarity with consuming, reducing $d_Z\left(\{z_t\}, Z_{con}\right)$. Changes in representation affect sensitivity to features. A consumer switching from *buy* to *con* becomes very sensitive to the higher quality of g relative to its higher price: her quality weight increases by $\overline{\alpha} - \underline{\alpha}$ while her price weight drops by $\overline{\alpha}$, boosting her preference for the higher quality product.

4.1.1 Experiences with Poverty and Price Sensitivity

Shah, Zhao, Mullainathan, and Shafir [119] show that poor people are more likely to deem price as relevant, and to exhibit high price elasticity, across many situations. This phenomenon can lead to large mistakes, such as avoiding medical out-of-pocket costs (Baicker, Mullainathan, Schwartzstein [5], Chandra, Flack, Obermeyer [25]). We explain such mistakes by the rigidity of the buy category due to the frequency with which the poor experience high opportunity costs. Suppose for simplicity that $q_b = p_b = 0$. Then Proposition 7 in the Appendix shows that:

$$\frac{\partial v(g)}{\partial F_{buy}} \propto -\lambda\left[(\overline{\alpha} - \underline{\alpha}) \cdot q_g + (\overline{\alpha} - 0) \cdot \eta \cdot p_g\right] \cdot \Pr\left(buy\,|t\right) \cdot \left[1 - d\left(t, buy\right)\right] < 0. \quad (11)$$

Due to a "mental set" of price-benefit evaluations, a person with many poverty experiences values the good less than a person with fewer such experiences, and is more price sensitive, $\partial v(g)/\partial p_g \partial F_{buy} < 0$. Unlike with rational inattention, price focus is suboptimal: it reflects overgeneralization of experiences. This has two implications.

1. Mistakes. A poor consumer may forsake valuable expenditures such as health copays due to her excessive focus on keeping down spending, which arises because $\underline{\alpha} < \overline{\alpha} \leq 1$ in Equation (11). A formerly poor consumer may exhibit high price elasticity even if she is no longer poor, i.e., even if η is small, inconsistent with neoclassical and rational inattention models. Past experiences cause price elasticity to depend on characteristics beyond current income and wealth (Hoch, Kim, Montgomery, and Rossi [65]).

2. Instability. Price focus depends on spurious context. By Proposition 2, $\Pr(buy\,|t)$ in (11) decreases in situation distance $d_Z(\{z_t\}, Z_{buy})$, which boosts quality focus, reducing price elasticity. The poor can "splurge" on festivals or "treat" goods such as cigarettes (Banerjee and Duflo [4]). These situations are associated with consumption pleasure, which increases $d_Z(\{z_t\}, Z_{buy})$. Relatedly, the poor are more price elastic if costs are monetary rather than in kind: out of pocket costs increase similarity to buying, reducing $d_Z(\{z_t\}, Z_{buy})$. This is in line with the compatibility principle (Tversky, Sattath, and Slovic [137] and Slovic, Griffin, and Tversky [122]). A previously-poor consumer could be more price elastic on items she used to buy, say clothes, compared to new goods, say i-Phones. Having only *buy* experiences with clothes, the distance $d_Z(\{z_t\}, Z_{buy})$ is lower for these goods compared to i-Phones.

Overgeneralization can be tested by comparing the price elasticity of two consumers who have similar current endowments such as income and wealth but different experiences. Rick, Cryder, and Loewenstein ([107]) develop a survey measure of thriftiness and show systematic (e.g., age-based) heterogeneity in consumers' focus on paying. We offer an "economic" theory of these differences: ceteris paribus, people with a poorer past have more frequently experienced high opportunity costs. They extrapolate the past high relevance of price to the present, exhibiting a higher price elasticity than an otherwise identical consumer.

Context specificity can be tested by comparing the price elasticity of the same consumer across situations that vary in “spurious” attributes. Wakefield and Inman ([142]) show that a consumer’s price elasticity is highly 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 generating these choices. Neoclassical economics may explain situation-specificity by allowing the consumer’s utility function to depend on the situation. Our approach does not assume ad hoc hedonics: it accounts for preferences based on measurable variation in past experiences and context. These factors can also cause violations of consistency axioms such as WARP, ruling out *any* utility explanations.

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, and so on. 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 [126]). 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 already been paid.

Opportunity cost neglect has been attributed to the temporal remoteness of the wine’s purchase price (Gourville and Soman [56]), while the sunk cost fallacy to diminishing sensitivity (Thaler [126]) or distaste for “waste” (Shafir and Thaler

[118]). In our account, both phenomena arise from categorization of decisions into consuming versus buying, which unifies some of the earlier intuitions but also yields new testable predictions.[16]

The consumer's choice is now ex post: the hedonics of g are realized. In the wine problem, "drinking" gives utility q_g and a wealth loss equal to the price p_g plus the capital gain $\Delta p_g > 0$. "Not drinking" gives zero. In the football problem, "driving" gives utility q_g plus the cost of driving $\Delta q_g < 0$ and a wealth loss equal to the ticket's price p_g. "Not driving" entails depreciation of p_g. The current problem is described by these hedonics and by the vignette context (the situation z_t).

Attention to hedonics $\alpha_O = (\alpha_Q, \alpha_{\Delta Q}, \alpha_P, \alpha_{\Delta P})$ shapes the dollar cost felt after drinking and the relative value of going to the game:

$$\begin{aligned} v(drinking(\alpha)) &= \alpha_P \cdot p_g + \alpha_{\Delta P} \cdot \Delta p_g, & (12) \\ v(driving(\alpha)) - v(not\ driving(\alpha)) &= \alpha_Q \cdot q_g + \alpha_{\Delta Q} \cdot \Delta q_g. & (13) \end{aligned}$$

Full attention $\alpha_O = (1, 1, 1, 1)$ yields rationality. In wine, $\alpha_P = \alpha_{\Delta P} = 1$ recovers in Equation (12) the market price, $p_g + \Delta p_g = \$75$. For estimating monetary costs, only attention to prices matter, q_g is not relevant. In football, $\alpha_Q = \alpha_{\Delta Q} = 1$ also recovers the rational rule: go to the game if and only if $q_g + \Delta q_g > 0$. The sunk price p_g is not relevant now. Each vignette creates a context κ_t matching diagnostic features of "consuming" (it reports qualities) and of "buying" (it reports price). It also describes a situation z_t. The competing categories prompt evaluations:

$$\begin{aligned} v(drinking(\alpha)) &= \begin{cases} 0 & if\ \alpha = \alpha_{con} \\ \overline{\alpha} p_g & if\ \alpha = \alpha_{buy} \end{cases}, & (14) \\ v(driving(\alpha)) - v(not\ driving(\alpha)) &= \begin{cases} \overline{\alpha} \cdot (q_g + \Delta q_g) & if\ \alpha = \alpha_{con} \\ \underline{\alpha} \cdot q_g & if\ \alpha = \alpha_{buy} \end{cases}. & (15) \end{aligned}$$

[16] Kőszegi and Matějka [78] 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 problem, "consuming" focuses the DM on the pleasure of drinking, triggering opportunity cost neglect. "Buying" focuses the DM on the price p_g, neglecting the capital gain, which is not a standard feature of most buy decisions.

In the football problem, "consuming" focuses the DM on the game q_g and the blizzard Δq_g, making the rational evaluation. "Buying" instead focuses the DM on the pain of paying p_g and the benefit q_g the payment had secured. This consumer represents driving as "enjoying a game I paid for" and not driving as "waste of p_g"as in Shafir and Thaler ([118]). The blizzard shock is neglected: it is not a standard feature of buy decisions.

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

1. Frequency. People who have recently bought a lot of wine but have not yet drunk it have high F_{buy}, which favors category *buy* and hence the $20 mode. In football, people who bought season tickets face only one buying decision but many consuming experiences; they thus 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 [86] (and will have a "sell wine" category, prompting attention to the capital gain). Having a higher 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 the vignette context. Describing the wine situation as z_t = "having drunk the wine" prompts similarity to *con* and dissimilarity to *buy*, fostering full price neglect.[17]

[17] During many consumption experiences, with prices not explicitly mentioned, we often feel no opportunity costs. Frederick, Novemsky, Wang, Dhar, and Nowlis [43] show experimentally that describing the option "not buy" in terms of keeping the money for other purchases substantially

Describing the football situation as z_t = "going versus not going to the game" evokes the past buying choice: paying or not to see a game. This fosters the 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 *con* and reduce that on *buy*, reducing the fallacy. We study the role of description in Section 5.

The same mechanism explains non-fungibility: transferring money into a category cues buying in that category, promoting in-category spending.[18] Consider account-based commitment (Thaler [126]): setting up a "rainy day" account focuses the DM on future financial risks, and hence on opportunity costs, prompting similarity to *buy*. The account thus triggers price focus, hindering spending. Categorization thus explains why some accounts are more tempting than others: different account names or purposes focus the consumer on different features of options, causing departures from the fungibility of money.

4.2 Statistical Problems

Probability judgments in i.i.d. draws and inference exhibit systematic biases (Benjamin [11]). When estimating the relative probability of obtaining sequences $H_1 = hhhhhh$ versus $H_2 = htthht$ from a fair coin, many people commit the Gambler's Fallacy (GF): they overestimate H_2 relative to H_1. Bordalo, Conlon, Gennaioli, Kwon, and Shleifer [21] propose a theory where these biases reflect attention-driven multi-modality and instability. The GF arises because in the problem above many people attend to the share of heads of the two sequences.

But why is the share of heads so prominent? Our answer is that, due to superficial similarity, many people confuse the problem with inferring whether the coin is biased or not. Thus, they focus on the share of heads, which is relevant in inference but irrelevant in the current problem, and commit the GF. As with

decreases the probability of purchase.

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

the previously-poor consumer's focus on price, focus on the share of heads reflects overgeneralization of familiar inference experiences.

Consider a general class of problems: a coin with probability θ of heads is selected according to a prior $\mathbb{P}_0$ in a set of coins $\Theta \subseteq [0,1]$, and generates i.i.d. sequences of up to $D \in \mathbb{N}$ flips in $\{h,t\}$. Hypotheses $o \in O$ are events a sigma algebra $\mathcal{F}$, whose probability can be computed with $\mathbb{P}$. To ease notation we restrict to coins but we can generalize, including to non-binary devices.[19] An atom is an elementary event, $\omega \in \Omega$, with feature vector:

$$y(\omega) = (\theta, n, e_1, ..., e_f, s, e_V) \tag{16}$$

where $\theta \in \Theta$ is the probability of heads of the selected coin; n is the number of draws; e_j, $j \in \{1, ..., n\}$ is the realization of the j-th draw; s is the event corresponding to the share of heads; and e_V is the name of the hypotheses, a feature indicating to which hypothesis this atom belongs.

The context of the GF problem, specifies a coin type prior $\mathbb{P}_0$ degenerate on 0.5, so $\Theta_t = \{0.5\}$, the number of flips $N_t = 6$, the j-th flip feature $E_{jt} = \{h,t\}$, the share of heads feature $S_t = \{x/6\}_{x\in\{0,...,6\}}$, and the hypotheses' names $V_t = \{H_1, H_2\}$. As with consumption, the problem is multidimensional: there is a known fairness of the coin, 6 individual flips, etc. The DM must decide what to attend to. Two crude categories may come to mind, emerging from frequent experiences with specific statistical problems.

Frequency estimation, category $freq$, collects problems where the coin type θ is known and the DM estimates the probability of $\{h,t\}$. The diagnostic features of $freq$ are thus "coin type is a singleton", "one draw", "draw can be h or t", and

[19] Formally, $\Omega = \{(\theta, d, (r_1, ..., r_d)) : \theta \in \Theta, d \in \{1, ..., D\}, \forall i \in \{1, ..., d\}, r_i \in \{h,t\}\}$. The option corresponding to $H \in \mathcal{F}$ 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 [113], Section 9, we could allow for an incentive compatible elicitation for multiple events.

"hypotheses are specific draws". Formally:

$$\kappa_{freq} = \left(\Theta_{freq}, N_{freq}, E_{freq}, V_{freq}\right) = \left(\{\theta\}, 1, \{h, t\}, V_{freq}\right) \qquad (17)$$

where $V_{freq} \subseteq \{h, t\}$. When evaluating an atom the DM focuses on the names of hypotheses, setting $\alpha_{freq,V} = 1$ and neglects the rest $\alpha_{freq,i} = 0$ for $i \in M_O \backslash \{\Theta, V\}$. For problems of frequency estimation, this process is both intuitive and correct.

Agnostic inference, category inf, collects problem where θ is not known and must be inferred from a signal, such as assessing whether $\theta = 0.7$ versus $\theta = 0.5$ based on one flip h. The diagnostic features of inf are: "there are multiple coin types", "there is more than one draw", "one draw is selection of a coin, other draws are coin flips", and "hypotheses are coin types". Formally:

$$\kappa_{inf} = \left(\Theta_{inf}, N_{inf}, \left(E_{j,inf}\right)_{j=1}^{N_{inf}}, V_{inf}\right), \qquad (18)$$

where $\Theta_{inf} \in \{\{\theta\} : \theta \in [0, 1]\}$, $N_{inf} \in \mathbb{N} \backslash \{1\}$, $E_{j,inf} = \{h, t\}$ for all $j \in \{1, ..., N_{inf}\}$ and $V_{inf} \subseteq \Theta_{inf}$. When evaluating an atom the DM focuses on its share of heads s and neglects the rest, including coin selection, setting $\alpha_{inf,S} = 1$ and $\alpha_{inf,i} = 0$ for $i \in M_O \backslash \{S\}$. With an uninformative prior this process is intuitive and correct: the share of positive signals is a sufficient statistic for θ. Since we often have little prior information, category inf generally works well for inference.[20]

Neither category fits perfectly. Compared to the problem, $freq$ in Equation (17) differs along the number of flips $d_N(6, 1) > 0$ and along the hypotheses, which do not correspond to a single flip, $d_V(V_t, \{h, t\}) > 0$. Similarly, inf in (18) differs along the set of coin types, which is not a singleton, $d_\Theta(\{0.5\}, \Theta_{inf}) > 0$, and hypotheses, which are not coin types, $d_V(V_t, V_{inf}) > 0$. Since both frequency estimation and

[20] We could allow for richer frequency categories with $r > 1$ i.i.d flips. Bordalo, Conlon, Gennaioli, Kwon, and Shleifer [21] allow for them in reduced form. These categories would not change our basic results, but allow the model to produce insensitivity to sample size. It is unnecessary to allow for a category of lopsided inference because it endogenously emerges from $freq$ when when statistical contrast is introduced in Section 5. See Appendix A.2.

inference are often encountered, some people focus on the coin's known fairness and represent the problem as $freq$, others on the length of sequences and select inf.

A DM relying on $freq$ deems the hypothesis events as relevant $\alpha_{freq,V} = 1$ and she neglects the other features, $\alpha_{freq,i} = 0$ for $i \in M_O \backslash \{V\}$. By Equation (3):

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

Hypotheses are deemed equally likely. The DM does not commit the GF but she also does not 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 inf deems only the share of heads to be relevant, $\alpha_{inf,S} = 1$ and neglects the other features $\alpha_{inf,i} = 0$ for $i \in M_O \backslash \{S\}$. By Equation (3),

$$v\left(H_1\left(\alpha\right)\right) = \mathbb{P}\left(s = 1\right) = \left(0.5\right)^6 \quad \text{and} \quad v\left(H_2\left(\alpha\right)\right) = \mathbb{P}\left(s = 0.5\right) = 5 \cdot \left(0.5\right)^4.$$

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

The GF is thus caused by the DM's confusion of the problem with inference, due to neglect that the coin *is* fair. This yields heterogeneity. People neglecting this feature commit GF. People who focus on fairness focus on the 50:50 nature of individual flips, avoiding the GF. Endogenous heterogeneity in similarity and representations explain disagreement despite common information and incentives.

Our theory also predicts instability in the GF: it should become less prevalent with fewer coin flips, which renders the problem more similar to frequency estimation of an individual flip. It should also become less prevalent when the names of hypotheses coincide with individual flips, which again render the problem more similar to $freq$. Experimental evidence in Bordalo, Conlon, Gennaioli, Kwon, and Shleifer [21] is consistent with both predictions. Specifically, when asked to judge *hhhhht* vs. *hhhhhh*, the incidence of the GF is reduced if subjects are asked "the first five flips are $hhhhh$, is the last flip h or t?"

As we show in Appendix A.2, competition between frequency estimation and agnostic inference categories also accounts for multi-modality and instability in inference. Anchoring to base rates in balls and urns experiments is due to the fact that the name of hypotheses "is the drawn ball from urn A or B?" is very similar to the event of selecting the urn. This makes the problem similar to $freq$, triggering focus on the base rate and the neglect of the signal. In the same problem, however, people focusing on the lack of knowledge about the DGP form an agnostic inference representation and anchor on the likelihood.[21]

Consumer choice and probability estimation entail very different tasks, but several long-standing puzzles can be unified by the use of partial representations focusing attention on features of options that are relevant in a superficially similar category of problems. At the same time, when payoff risks are salient, such as when choosing among financial assets, the problem requires both hedonic evaluation and statistical estimation. In this case, consumption and statistical categories compete with one another, as we show when studying risky lotteries next.

5 Bottom-up Attention and Representations

Choice is shaped by bottom-up salience. Consumers are more sensitive to taxes they already know about if those are shown on the price tag (Chetty, Looney, and Kroft [27]) and prefer goods that are physically present (Bushong, Camerer, and Rangel [23]). Contrasting prices, payoffs, or statistics draw attention (Bordalo,

[21] The model also explains the instability caused by the taxicabs format. Now the name of the hypotheses "is the cab green as claimed by the witness?", is very similar to the simple event "is the witness accurate or not?" for which the likelihood statistic is provided. Similarity to frequency estimation prompts anchoring to the likelihood. The effect is very strong also because the DGP is described in terms of the "accuracy" feature, making it bottom-up salient (see Section 5). Consistent with this account, instability in the taxicab frame is almost entirely due to switchers from the base rate to the likelihood, see Bordalo, Conlon, Gennaioli, Kwon, and Shleifer [21].

Gennaioli, and Shleifer [16],[17], Koszegi and Szeidl [79]).[22] We next show how sensory prominence and contrast affect categorization, producing framing effects.

Sensory prominence depends not on the choice task but on the problem description. The description is a prominence 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 $\alpha_{\delta,i}$), partly designed (e.g., a feature can be shrouded as in Gabaix and Laibson [47]). A shrouded feature $i \in M_O \cup M_K$, has $\alpha_{\delta,i} = 0$. A fully described feature has $\alpha_{\delta,i} = 1$. Described context is all the DM perceives, so $\kappa_t = \kappa_\delta$. Compared to our prior analysis, we next study the role of prominence α_δ of choice features. Visual prominence of a feature can be measured as in Li and Camerer [84]; related methods may be used to measure prominence of text.

Contrast is instead related to the choice task. It increases in the variability of the hedonics and events of choice options. High variability of relevant features draws attention. Unlike with goal optimal attention (Sims [120], Woodford [143], Gabaix [46]), however, contrast may *excessively* focus the DM on striking payoffs or probabilities, causing neglect of other relevant dimensions. Let Y be the set of atoms. Contrast of $i \in M_O$ is given by a real valued function σ_i of the $(y_i)_{y \in Y}$, $\sigma_i = \sigma \left[(y_i)_{y \in Y} \right] \geq 1$, a special case of which is:

$$\sigma_i = 1 + \frac{\sum_{(y,y') \in Y^2 : y \neq y'} d_i\left(y_i, y_i'\right) / |Y||Y-1|}{\sum_{y \in Y} d_i\left(y_i, \widetilde{y}\right) / |Y| + \epsilon}, \tag{19}$$

where $\epsilon > 0$ and $\widetilde{y}$ is a reference feature vector. For real-valued features,with $d\left(y_i, y_i'\right) = |y_i - y_i'|$ and $\widetilde{y} = 0$, it nests the exemplar function in Bordalo, Gennaioli, and Shleifer [17]. In Equation (19), price contrast increases if price variability is large (via the numerator), creating context dependence. Conversely, price contrast is reduced if average price is high (via the denominator), which yields diminishing sensitivity as in the Weber-Fechner law. Event contrast depends on event probabilities, as in Bordalo, Conlon, Gennaioli, Kwon, and Shleifer [21].

[22] See Lanzani [83] for the axioms underpinning this model.

Contrast arises both for features in the problem description and for features experienced in categories. Described contrast $\sigma_{\delta,i}$ is computed using the described values of feature i. If the feature is shrouded, $\alpha_{\delta,i} = 0$, its bottom-up contrast is minimal, $\sigma_{\delta,i} = 1$. Category contrast of the same feature, $\sigma_{c,i}$, is computed using the feature's experienced values in c. If i was neglected ($\alpha_{c,i} = 0$), then $\sigma_{c,i} = 1$. Thus, experiences create "top-down" contrast. Before flying, some people tend to think about crashes: even if not described, crashes are contrasting in category "flying".[23]

Matching. The DM matches description $(\alpha_\delta, \kappa_t)$ with each category $c \in C$, using a similarity function where more contrasting features receive higher weight (see Nosofsky [101] and Reed [106]):

$$S\left[\left(\alpha_t, \kappa_t\right), \left(\alpha_x, \kappa_x\right) | \sigma_x\right] = \\ \frac{\sum_{i \in M} \sigma_{x,i} \cdot \left[1 - d\left(\left|\alpha_{t,i} - \alpha_{x,i}\right|\right)\right] + \sum_{i \in M_K} \left[1 - \alpha_{t,i}\alpha_{x,i} d_i\left(\kappa_{t,i}, \kappa_{x,i}\right)\right]}{|M| + |M_K|},$$

where $x \in \{\delta, c\}$ covers features in description and across categories, and $\sigma_{x,i} = 1$ for $i \in M_K$ as context features do not vary across options. Equation (6) becomes:

$$S\left(t, c, \delta | \sigma\right) = \max_{\alpha_t \in [0,1]^{M_O \cup M_K}} F_c \cdot S\left[\left(\alpha_t, \kappa_t\right), \left(\alpha_c, \kappa_c\right) | \sigma_c\right] + S\left[\left(\alpha_t, \kappa_t\right), \left(\alpha_\delta, \kappa_t\right) | \sigma_\delta\right]. \tag{20}$$

The DM trades off similarity to c, the first term, against similarity to the description, the second term, capturing the intuition that the description anchors the representation. The interior optimal attention satisfies:

$$\frac{\partial}{\partial \alpha_{t,i}} d\left(\left|\alpha_{t,i} - \alpha_{c,i}\right|\right) + d_i\left(\kappa_{t,i}, \kappa_{c,i}\right) \cdot \mathbb{I}_{\{1\}}(\alpha_{c,i}) + \frac{\sigma_{\delta,i}}{\sigma_{c,i}} \cdot \frac{1}{F_c} \cdot \frac{\partial}{\partial \alpha_{t,i}} d\left(\left|\alpha_{t,i} - \alpha_{\delta,i}\right|\right) = 0. \tag{21}$$

Compared to (7), bottom-up salience adds the third term. The analysis of Section 2, in which description only affects context κ_t, is a special case for $F_c \to \infty$.

[23] The feature values used to compute contrast are encoded in context, κ_t and κ_c, which report the possible values of choice features.

Proposition 4 describes the impact of description on attention during matching to a category.

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 prominence $\alpha_{\delta,i}$ and category c attention $\alpha_{c,i}$. The description is more influential, namely $|\alpha_{t,i}(c, \delta, \sigma) - \alpha_{\delta,i}|$ is lower, when $\frac{\sigma_{\delta,i}}{\sigma_{c,i}} \cdot \frac{1}{F_c}$ is higher.*

If a feature such as price is visually prominent, high $\alpha_{\delta,P}$, it is more attended to in any category c, as in Li and Camerer [84], and especially so if current prices are contrasting, high $\sigma_{\delta,P}$, as in Bordalo, Gennaioli, and Shleifer [17],[16]. As we argue in the conclusion, this role of bottom up attention can shape the process of category formation itself. The sensitivity of attention to bottom up forces depends on the category being matched: it is smaller if the category is more frequent, higher F_c, or if price has higher contrast in the category, $\sigma_{c,P}$.[24]

Through its impact on attention, description shapes similarity and categorization:

Proposition 5 *Let $i \in M_O$ and $\alpha_{\delta,i} > \alpha_{c,i}$. increasing $\alpha_{\delta,i}$ or $\sigma_{\delta,i}$ increases similarity more for categories in which that feature is more relevant, $\alpha_{c,i}$ is higher. That is*

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

Categories focused on prominent and contrasting features are more likely to be selected. This is a theory of framing effects: a bottom-up salient feature can produce a shift in representation, causing preference reversals. Advertising a good as a “bargain”, or writing prices in larger font, increases price prominence $\alpha_{\delta,P}$, promoting retrieval of the “buying”category. This increases the weighting of price relative to quality, and favors cheaper goods.[25]

[24] The perception of sensory stimuli themselves are mediated by categories, Kay and Ross [77].

[25] Bottom-up framing explains why people tend to focus on prominent descriptions, neglecting the rest (What You See is All There Is, Kahneman [70], Enke [36], Graeber [57]).

Changes in intrinsic attributes can also cause re-framing. A more contrasting price difference, higher $\sigma_{\delta,P}$, also draws attention and favors a switch to the "buying" category. In fact, the strongest *average* sensitivity to change in a choice feature may be observed when it prompts a category switch.

A key implication of Proposition 5 is that the impact of framing on categorization and decisions is limited by strong reliance on certain categories. If the DM is accustomed to "consuming", F_{con} is high, she may neglect price even if visually prominent. Conversely, a poor consumer has large opportunity costs, high $\sigma_{c,P}$, so she attends to prices even if shrouded. In fact, the interaction between bottom-up attention and categorization sheds new light on choice under risk, statistical problems, and similarity judgments. We next examine this interaction for lottery choice, where systematic biases have been documented (see Kahneman and Tversky [75]) and linked to bottom-up attention (Bordalo, Gennaioli, Shleifer [16]).

5.1 Choice Implications

Lottery choice covers both payoffs *and* probabilities. Reasoning about payoffs cues experiences of "consuming" similar amounts, and hence category *con* (Section 4.1). Reasoning about events cues experiences of "frequency estimation", category $freq$ (Section 4.2).[26] Expected utility integrates hedonic and statistical features. Competition between categories instead creates heterogeneous and unstable representations *across* domains. We show that instability based on bottom up salience generates well-documented framing effects that cannot be explained by existing theories, including payoff contrast in Bordalo, Gennaioli, and Shleifer ([16]).

Consider the choice between monetary lotteries

$$X = (x_g, x_b; \pi) \text{ and } W = (w_g, w_b; \beta),$$

[26] 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 multiple draws as in inference.

which pay their upside prizes x_g and w_g with probabilities π and β, respectively, and pay x_b and w_b otherwise. X is a mean preserving spread of W, so $x_g > w_g$ and $\pi < \beta$. Each lottery has four features: its maximum and minimum payoffs (*hedonic*) and the maximum and minimum payoff events (*event*). Each lottery has two atoms of the form (u_g, u_b, e_g, e_b). The atom of a lottery state $s \in \{g, b\}$ specifies: i) in u_s the payoff in this state and in e_s its delivery event, and ii) the neutral values $u_{-s} = 0$ and $e_{-s} = \Omega$ for the features not corresponding to s.[27]

With quadratic distance d, Proposition 4 implies that, when matching to $c \in \{con, freq\}$, attention satisfies:[28]

$$\alpha_{t,u_s}(con) = \frac{\overline{\alpha} + \alpha_{\delta,u_s} \cdot \sigma_{\delta,u_s}/F_{con}}{1 + \alpha_{\delta,u_s} \cdot \sigma_{\delta,u_s}/F_{con}}, \qquad \alpha_{t,e_s}(con) = \frac{\alpha_{\delta,e} \cdot \sigma_{\delta,e}/F_{con}}{1 + \alpha_{\delta,e} \cdot \sigma_{\delta,e}/F_{con}}, \tag{22}$$

$$\alpha_{t,u_s}(freq) = \frac{\alpha_{\delta,u_s} \cdot \sigma_{\delta,u_s}/F_{freq}}{1 + \alpha_{\delta,u_s} \cdot \sigma_{\delta,u_s}/F_{freq}}, \qquad \alpha_{t,e_s}(freq) = 1, \quad s \in \{g, b\}\,. \tag{23}$$

Ceteris parisbus, compared to "frequency estimation", categorization in "consuming" boosts attention to payoffs but dampens it to probabilities, $\alpha_{t,u_s}(con) > \alpha_{t,u_s}(freq)$ and $\alpha_{t,e_s}(con) < \alpha_{t,e_s}(freq)$. Higher sensory prominence or contrast of payoffs in state s (higher α_{δ,u_s} or σ_{δ,u_s}) boosts attention to the payoff in this state. Higher event prominence or contrast (higher $\alpha_{\delta,e}$ or $\sigma_{\delta,e}$) boosts attention to probabilities. The bottom up forces are weaker in frequent categories (due to the terms $\alpha_{\delta,u_s} \cdot \sigma_{\delta,u_s}/F_c$ and $\alpha_{\delta,e} \cdot \sigma_{\delta,e}/F_c$).

At the equilibrium categorization, the DM's valuation gap for X over W is given by:

[27] Formally, lottery $Z \in \{X, W\}$ consists of two atoms $(z_g, 0, \{\omega \in \Omega : Z(\omega) = z_g\}, \Omega)$ and $(0, z_b, \Omega, \{\omega \in \Omega : Z(\omega) = z_b\})$. In particular, when applying category $freq$, for each atom $s \in \{g, b\}$ of lottery $Z \in \{X, W\}$ the feature V describing the hypothesis under evaluation is $e_s = \{\omega \in \Omega : Z(\Omega) = z_s\}$. In a more complete formalization, the atom of a lottery also includes the two event features of the alternative lottery, 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 redundant in our case because the lotteries are independent.

[28] To ease notation we set category contrast to $\sigma_{c,i} = 1$ for all $i \in M_C$. Our qualitative results do not depend on this assumption.

$$v(X) - v(W) = \Pr(con|t) \cdot \widetilde{v}(con) + \Pr(freq|t) \cdot \widetilde{v}(freq), \tag{24}$$

where the valuation gap $\widetilde{v}(c)$ between X and W is computed using the above attention weights, which we later illustrate for specific lotteries.

Our new insight is that prominence $\alpha_{\delta,i}$ shapes sensitivity to *given* payoffs and probabilities within a category c and it also shapes, together with their contrast $\sigma_{\delta,i}$, categorization. By Proposition 5, higher prominence or contrast of payoffs promotes "consuming" because in *con* payoffs are relevant. Higher event prominence or contrast promotes "frequency estimation" because in $freq$ probabilities are relevant. These forces account for a range of puzzles in risky choice.

Common Ratio Effect. Suppose $x_b = w_b = 0$. It is well known that if $(100, 0; 0.2) \sim (25, 0; 0.8)$, then for many people $(100, 0; 0.02) \succ (25, 0; 0.08)$. This pattern violates expected utility, in which preferences are invariant to uniform scaling of probabilities. We explain this puzzle: probabilities shrink, event contrast drops. This triggers a shift towards a representation that focuses on payoffs, which benefits the riskier lottery.

Formally, the valuation gap $\widetilde{v}(c)$ between X and W in category c is given by:

$$\widetilde{v}(c) = \pi^{\alpha_{t,e_g(c)}} \cdot \alpha_{t,u_g(c)} \cdot (x_g - w_g) + \left[\pi^{\alpha_{t,e_g(c)}} - \beta^{\alpha_{t,e_g(c)}}\right] \cdot w_g(\alpha_{t,u_g(c)}). \tag{25}$$

Since $\pi \cdot x_g = \beta \cdot w_g$, the right hand side increases in attention to payoffs $\alpha_{t,u_g}(c)$ and decreases in attention to probabilities $\alpha_{t,e_g}(c)$. Thus, categorization in *con* boosts risk taking while categorization in $freq$ boosts risk aversion.

In the problem above, the reduction of probabilities reduces event contrast, e.g. $|\pi - \omega|$ drops from $|0.8 - 0.2| = 0.6$ to $|0.08 - 0.02| = 0.06$. Payoff contrast is instead constant.[29] The drop in event contrast reduces $\alpha_{t,e}(c)$ in any category and boosts categorization into consuming. Overall, when differences in probabilities

[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.

are “peanuts” the DM’s attention is drawn away from them and towards the \$100 versus \$25 payoff difference, promoting risk taking.

This shift in representations offers a foundation for Rubinstein’s [109] intuition that the common ratio effect arises from a change in the perceived similarity between probabilities, and thus unstable weights on them (an effect not considered in Bordalo, Gennaioli, and Shleifer [16]). The same mechanism can also contribute to the neglect of small background risks in everyday life (Gennaioli, Shleifer, and Vishny [49],[50]), because their probabilities under alternative courses of action are similar, which promotes a focus on the evaluation of payoffs.

To illustrate additional implications, suppose that W pays a sure amount $\widetilde{w} = \pi \cdot x_g$. Hedonic features are whether the lottery pays more or less than $\widetilde{w}$, with corresponding event features.[30] Payoff contrast σ_{δ,u_s} in state $s \in \{g, b\}$ increases in $|x_s - \widetilde{w}|$. The valuation gap for X over W in category c is now given by:[31]

$$\widetilde{v}(c) = x_g \cdot \left[\alpha_{t,u_g}(c) \cdot \pi^{\alpha_{t,e}(c)} \cdot (1-\pi) - \alpha_{t,u_b}(c) \cdot (1-\pi)^{\alpha_{t,e}(c)} \cdot \pi\right]. \qquad (26)$$

The gap is higher, fostering risk seeking, when the DM pays more attention to the lottery upside, higher $\alpha_{t,u_g}(c)$, or pays less attention to its downside, lower $\alpha_{t,u_b}(c)$.

Prominence. In our model, risk attitudes change with the description of lotteries even if payoffs and probabilities are constant. Consider two descriptions of X.

$$\text{Full Prominence} \; : \; \text{win } x_g \text{ with probability } \pi \text{ and 0 otherwise,} \qquad (27)$$

$$\text{Shrouded Downside} \; : \; \text{win } x_g \text{ with probability } \pi. \qquad (28)$$

[30] This is equivalent to the previous formalization of atoms provided that, for the sure thing, we set $w_g = w_b = \widetilde{w}$ and $\beta = \pi$.

[31] The next formula is obtained by formalizing a safe lottery paying some amount c, when compared to a risky alternative (x_g, π), as composed by two atoms $(c, 0, \{\omega : X(\omega) = x_g\}, \Omega)$ and $(0, c, \Omega, \{\omega : X(\omega) = x_b\})$. In words, it is modeled as a lottery with a good and a bad outcome that have respectively the probability π and $(1-\pi)$ of the alternative lotteries, but were both the good and bad outcome are equal to c. Qualitatively analogous results would be obtained by modeling the safe lottery as having equal good and bad outcomes with probability 0.5 that does not depend on the alternative under consideration. However, modeling it as a single atom would change the behavior described below.

In (27) the upside and downside payoffs are prominently described, $\alpha_{\delta,u_s} = 1$, as is often done with lotteries in the lab. In (28), in contrast, the downside is shrouded, $\alpha_{\delta,u_b} < 1$. By Equations (22) and (23), a less prominent downside reduces attention to this feature, promoting risk taking in any category.

This effect explains important phenomena. Advertising high returns fosters investment and neglect of risk (especially when in good times the crash event is shrouded, see Mullainathan and Shleifer [98], Célérier and Vallée [24]).[32] Risk attitudes can also shift depending on the prominence of the DM's gain/loss state, or of an insider/outsider view (Kahneman and Lovallo [72]). When prompted to think as a trader, people are less loss averse (Sokol-Hessner, Hsu, Curley, Delgado, Camerer, and Phelps [123]). These phenomena arise because: i) the framing shapes the prominence of payoffs, explicitly or implicitly, and ii) a more prominent payoff promotes a switch to a payoff evaluation representation, which triggers a neglect of probabilities.

Discontinuities. Categories can also yield aversion to minuscule risks, as in Kahneman and Tversky's [75] "certainty effect". Consider first two lotteries X and W that pay x and w with certainty, with $x > w$. Categorization in *con* and choice of X are straightforward: probabilities are not involved.[33] Adding a small risk ε of a zero payoff to lottery X, so that the probability of the upside x is now $\pi = 1 - \varepsilon$, creates: i) a downside payoff feature, whose salience increases in contrast $|w - 0|$, and ii) an event feature, increasing similarity with "frequency estimation".

Critically, as the probability of the downside gradually increases from 0 to ε, similarity to a frequency estimation problem does not change much. On the other hand, the downside payoff contrast has a sharp, discontinuous jump to $|w - 0|$, which reinforces categorization in *con*. Thus, the DM sticks to the original payoff-

[32] 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.

[33] 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.

evaluation representation, with the entailed attention to the contrasting downside payoff and insensitivity to its small probability (fully so, $\alpha_{t,e_s} = 0$, $s \in \{g, b\}$, if "consuming" is very frequent $F_{con} \to \infty$).[34] Strong aversion to the small risk follows.

Barseghyan, Molinari, O'Donoghue, and Teitelbaum [7] show that discontinuous probability weighting at 0 (as suggested in Kahneman and Tversky [75] but subsequently abandoned) is important to understand insurance demand. Haigh and List [60] document discontinuities also with professional traders.[35] Discontinuity cannot be explained by continuous payoff weights as in Bordalo, Gennaioli, and Shleifer [17]. Here it arises because the small probability risk sharply boosts payoff contrast but modestly increases similarity to frequency estimation problems, causing a discontinuous focus on the downside.

The Fourfold Pattern and "Simplicity Equivalents". Equation (26) also yields the so-called "fourfold pattern" in risky choice (Kahneman and Tversky [75]) but crucially reconciles it with Oprea's [103] evidence of similar behavior when choosing among riskless mirrors. This arises due the use of a common "consuming" representation in both domains.

To see this, consider risky lotteries first. Given that $x_g = \widetilde{w}/\pi$, 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 (22) and (23), then, in every category the DM focuses more on the upside if and only if $\pi < 0.5$. The DM is risk seeking for $\pi < 0.5$ and

[34] 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.

[35] Another important example of discontinuity arises in situations involving social norms. Gneezy and Rustichini [54] show that a small payment reduces 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 [80]). However, adding the possibility of taking away money from the opponent leads to a discontinuous change in behavior towards not sharing (List [87]).

risk averse for $\pi > 0.5$, as in the fourfold pattern.[36]

Consider next a riskless choice where options have a similar description, what Oprea [103] calls "riskless mirrors". Option A consists of 90 boxes with $x_g = \$2.5$ and 10 boxes with $x_b = \$0$, and option B consists of 100 boxes with $\widetilde{w} = \$2.25$. The subject is paid the total value of the chosen option divided by 100. The two options pay the same, but (when going down a multiple price list) many subjects exhibit a preference for B. Options have two features: the payoff in boxes $s \in \{g, b\}$ and their frequencies n_g and n_b. Also in this case, reasoning about payoffs prompts the "consuming" category, while reasoning about frequencies prompts the "frequency estimation" category.

When evaluating A versus B, the contrast between A's payoff in its g and b boxes with the payoff in B boxes prompts categorization as *con*. The DM focuses on evaluating payoffs of different boxes, reducing attention to their precise frequencies. From Equation (3), the value of the generic atom y of A in state $s \in \{g, b\}$ is

$$v\left(y\left(\alpha_O\right)\right) = \frac{n_s^{\alpha_{t,e}} \cdot \left[\alpha_{t,u_s} \cdot x_s + (1 - \alpha_{t,u_s}) \cdot (x_s + \widetilde{w})/2\right]}{100},$$

so that in category c the valuation gap for A over B is given by:

$$\widetilde{v}\left(c\right) = x_g \cdot \left[\alpha_{t,u_g}(c) \cdot \left(\frac{n_g}{100}\right)^{\alpha_{t,e}(c)} \cdot \left(\frac{n_b}{100}\right) - \alpha_{t,u_b}(c) \cdot \left(\frac{n_b}{100}\right)^{\alpha_{t,e}(c)} \cdot \left(\frac{n_g}{100}\right)\right]. \quad (29)$$

Equation (29) is equivalent to Equation (26) with risky lotteries, and so yields equivalent choice behavior. The fourfold pattern cannot come from preferences for risk, because the domains are different. It comes from the payoff evaluation

[36] Relative to the Bordalo, Gennaioli, and Shleifer [16], here payoff-contrast triggers a consuming representation, causing probability neglect ($\alpha_e = 0$ for $F_{con} \to \infty$). Thus, the fourfold pattern can also arise if the upside and downside payoffs are equally attended to, $\alpha_{t,u_g} = \alpha_{t,u_b}$, which is relevant for testing the model using attention data. We also predict that 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 dollar units like payoffs, certainty equivalents are more similar to payoff evaluation, $c = con$.

representation, induced by high payoff contrast in both domains.

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

Riskless choice can be contaminated by numerical tasks in other domains. In the left digit bias (e.g. List, Muir, Pope, and Sun [88]), consumers perceive $9.99 to be dissimilar from $10 despite the metric proximity. A number's digits are akin to event features in a sequence of i.i.d. draws. The fact that the left digit is usually more relevant prompts people to focus on this feature and neglect others, boosting the perceived difference between $9.99 and $10.

Our model also accounts for framing effects in statistical and riskless choice, as we show with two examples in the Appendix. First, Appendix A.2 shows that contrasting statistics offer a foundation for instability in judgment of i.i.d. draws and inference documented by Bordalo, Conlon, Gennaioli, Kwon, and Shleifer [21].

Second, Appendix A.3 shows that contrast yields the famous context-dependent similarity judgments in Tversky [133]. 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. We explain this as follows: when Poland is on the list, political differences are contrasting, so the problem is represented as "evaluating political proximity". Sweden and Austria are then deemed similar. When Norway is on the list, geographic differences are contrasting, so the problem is represented as "evaluating geographical proximity." Sweden and Austria are then deemed dissimilar.

6 Conclusion

Mental representations that can vary based on experience and contextual cues unify puzzling decisions within and across domains. This mechanism addresses a key shortcoming of neoclassical and behavioral theories, namely the predicted stability of preferences 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 open several avenues. We discuss four key challenges ahead.

First, a major task for future work is to measure or experimentally control the key ingredients and cognitive concepts in the model, namely: i) mental representations/categories, ii) contextual similarity (see Bordalo, Conlon, Gennaioli, Kwon and Shleifer [21], Bordalo, Burro, Coffman, Gennaioli and Shleifer [14]) and iii) experiences. These new parameters should then be used in conjunction with choice data (Malmendier and Nagel [92]). Our approach is highly complementary to AI methods, which can help recover representations from self-reported reasons for choice (Haaland, Roth, Stantcheva, and Wohlfart [59], Link, Peschl, Roth, and Wohlfart, [85]), from which similarity within and across domains can be extracted, as well as by unveiling subtle context features as in Ludwig and Mullainathan [89]. These unveiled features complement prominently described parameters such as the choice set or the data generating process and their role can be interpreted through the cognitive mechanisms in our framework.

Our model can also shed light on how non-choice data, whose availability is growing fast, can be incorporated into structured economic analysis and linked to precise out-of-sample predictions, thus increasing testability and external validity. This is possible because our theory, rather than specifying stable choice parameters, links representations and choice to measurable context features.

A second set of implications concerns the design of experiments. Current practice favors the use of abstract protocols to better identify "universal" choice biases and minimize experimenter demand. In our model, abstraction is desirable for studying general cognitive mechanisms, as it enables precise control of 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 removing the 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 of naturalistic contexts.

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 [28]) 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 [71]), to cooperation with strangers versus in groups (Enke [35] and Malmendier [91]), to beliefs about social preferences across groups (Exley, Hauser, Moore, and Pezzuto [39]), and so on. Our approach can also help explain strategic behavior. Goke, Weintraub, Mastromonaco, and Seljan ([55]) 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 big 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. For a poor DM, the opportunity cost of spending is large, so prices are bottom up contrasting for her. This causes her to frequently focus on price, causing the formation of a large price-focused buying category. Later, even if the DM becomes better off, she will be more likely to retrieve buying compared to someone with the same income but without the same experiences with poverty. The visual prominence of prices in stores also facilitate the use of a buying category in these contexts, and hence its subsequent retrieval in the same contexts. On the other hand, sensory prominence of pleasure and remoteness from prices makes the consumption experience immediately different from choosing whether to buy, leading to a consuming category focused on quality.

These bottom up forces imply that categories will not be free parameters. They have a precise structure that reflects measurable visual features and economic incentives. In fact, categories increase the predictive power of standard economic factors, making past economic incentives and conditions predictive of otherwise hard to explain heterogeneity in choices under similar current conditions. Past knowledge and incentives can also explain why people make choice errors and why they exhibit precise forms of instability, improving explanatory power over assumed heterogeneity in tastes or biases. These effects may shed 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] Ba, C., Bohren, J.A. and Imas, A., 2024. Over-and underreaction to information. Available at SSRN 4274617.

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

[5] Baicker, K., Mullainathan, S., and Schwartzstein, J., 2015. Behavioral hazard in health insurance. The Quarterly Journal of Economics, 130, pp.1623-1667.

[6] 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.

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

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

[9] 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.

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

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

[12] 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, pp.331 - 346.

[13] Bohren, A., Imas, A., Ungeheuer, M. and Weber, M., 2024. A Cognitive Foundation for Perceiving Uncertainty. Available at SSRN.

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

[15] Bordalo, P., Coffman, K., Gennaioli, N., Schwerter, F. and Shleifer, A., 2021. Memory and representativeness. Psychological Review, 128(1), p.71-85.

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

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

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

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

[20] 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.

[21] Bordalo. P., Conlon, J., Gennaioli, N., Kwon S., and Shleifer A., 2024. How people use statistics. The Review of Economic Studies, forthcoming.

[22] 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.

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

[24] 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.

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

[26] 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.

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

[28] 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.

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

[30] Conlon, J., 2024, Attention, information, and persuasion, mimeo.

[31] Conlon, J., and Kwon, S.,2024, Persuasion through cues, mimeo.

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

[33] Einhorn, H. and Hogarth, R., 1986. Decision making under ambiguity. Journal of Business, pp.S225-S250.

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

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

[36] Enke, B., 2020. What you see is all there is. The Quarterly Journal of Economics, 135(3), pp.1363-1398.

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

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

[39] Exley, C., Hauser, O., Moore, M. and Pezzuto, J., 2025. Believed gender differences in social preferences. The Quarterly Journal of Economics, 140, pp. 403-458.

[40] 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.

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

[42] 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.

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

[44] Fudenberg, D., Lanzani, G. and Strack, P., 2024. Selective memory equilibrium. Journal of Political Economy, 132, pp. 3978-4020.

[45] Fudenberg, D., Lanzani, G. and Strack, P., 2024. Learning, Memory, and Stochastic Choice. Available at SSRN 4923469.

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

[47] 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.

[48] Gagnon-Bartsch, T., Rabin, M. and Schwartzstein, J., 2023. Channeled attention and stable errors, Mimeo.

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

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

[51] Gennaioli, N. and Tabellini, G., 2023. Identity politics, Mimeo.

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

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

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

[55] 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.

[56] Gourville, J.T. and Soman, D., 1998. Payment depreciation: The behavioral effects of temporally separating payments from consumption. Journal of Consumer Research, 25(2), pp.160-174.

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

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

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

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

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

[62] Handel, B. and Schwartzstein, J., 2018. Frictions or mental gaps: what's behind the information we (don't) use and when do we care?, Journal of Economic Perspectives, 32(1), pp.155-178.

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

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

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

[66] Hsee, C., 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.

[67] 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.

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

[69] Jehiel, P., 2005. Analogy-based expectation equilibrium. Journal of Economic Theory, 123(2), pp.81-104.

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

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

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

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

[74] Kahneman, D. and Tversky, A., 1982. The simulation heuristic, Judgment under Uncertainty, pp. 201-208, Cambridge University Press.

[75] Kahneman, D. and Tversky, A., 1979. Prospect Theory: an analysis of decision under risk. Econometrica, 47(2), pp.263-292.

[76] Kant, I., 1781, Critique of Pure Reason.

[77] 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.

[78] 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.

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

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

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

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

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

[84] 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.

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

[86] List, J., 2004. Neoclassical theory versus prospect theory: Evidence from the marketplace. Journal of Political Economy, 72(2), pp.615-625.

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

[88] List, J., Muir, I., Pope, D. and Sun, G., 2023. Left-digit bias at lyft. The Review of Economic Studies, 90(6), pp.3186-3237.

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

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

[91] Malmendier, U., 2021. Experience effects in finance: Foundations, applications, and future directions. Review of Finance, 25, pp. 1339-1363.

[92] Malmendier, U. and Nagel, S., 2016. Learning from inflation experiences. The Quarterly Journal of Economics, 131(1), pp.53-87.

[93] Mas-Colell, A., Whinston, M., and Green, J., 1995. Microeconomic Theory, Oxford University Press.

[94] McFadden, D., 1974. Conditional logit analysis of qualitative choice behavior. Frontiers in Econometrics.

[95] Mohlin, E., 2014. Optimal categorization. Journal of Economic Theory, 152, pp.356-381.

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

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

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

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

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

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

[102] Ok, E.A., 2011. Real analysis with economic applications. Princeton University Press.

[103] Oprea, R., 2024. Decisions under risk are decisions under complexity. American Economic Review, 114(12), pp.3789-3811.

[104] 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.

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

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

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

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

[109] 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.

[110] Rudin, W., 1964. Principles of mathematical analysis (Vol. 3). New York: McGraw-hill.

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

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

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

[114] Schacter, D., Addis, D., Hassabis, D., Martin, V., Spreng, R. and Szpunar, K., 2012. The future of memory: remembering, imagining, and the brain. Neuron, 76(4), pp.677-694.

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

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

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

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

[119] 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.

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

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

[122] 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.

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

[124] 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.

[125] 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.

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

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

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

[129] Thaler, R. and Johnson, E., 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.

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

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

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

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

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

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

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

[137] Tversky, A., Sattath, S. and Slovic, P., 1988. Contingent weighting in judgment and choice. Psychological Review, 95(3), p.371 - 384.

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

[139] 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.

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

[141] Wachter, J.A. and Kahana, M.J., 2024. A retrieved-context theory of financial decisions. The Quarterly Journal of Economics, 139(2), pp.1095-1147.

[142] 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.

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

[144] 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 of all, observer that if $d_i\left(\kappa_{t,i}, \kappa_{c,i}\right) = 0$, the maximal similarity similarity of 1 can be achieved by letting $\alpha_{t,i}(c) = \alpha_{c,i}$, so clearly the weights fully follow the category ones.

Then, we will consider the case when $\alpha_{c,i} = 0$. By Equation (5), we have that $\alpha_{t,i}$ only affects $S\left[\left(\alpha_t, \kappa_t\right), \left(\alpha_c, \kappa_c\right)\right]$ 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.

Finally, if $\alpha_{c,i} = 1$ and $d_i\left(\kappa_{t,i}, \kappa_{c,i}\right) > 0$, then any $\alpha_{t,i}(c)$ weakly shrunk attention towards 0, concluding the proof of the first part of the statement.

We now prove the asserted relation between $d_i\left(\kappa_{t,i}, \kappa_{c,i}\right)$ and $\alpha_{t,i}(c)$. It is trivial if $\alpha_{c,i} = 0$, so consider the case where $\alpha_{c,i} = 1$. Because d is strictly convex, $S\left[\left(\alpha_t, \kappa_t\right), \left(\alpha_c, \kappa_c\right)\right]$ 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:

$$\frac{\partial S\left[\left(\alpha_t, \kappa_t\right), \left(\alpha_c, \kappa_c\right)\right]}{\partial \alpha_{t,i}} \begin{cases} = 0 \text{ and } \alpha_{t,i} \in [0,1] \\ > 0 \text{ and } \alpha_{t,i} = 1 \\ < 0 \text{ and } \alpha_{t,i} = 0 \end{cases} . \tag{30}$$

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

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

the first order condition simplifies to

$$G(\alpha_{t,i}, d_i\left(\kappa_{t,i}, \kappa_{c,i}\right)) \begin{cases} = 0 \text{ and } \alpha_{t,i} \in [0,1] \\ < 0 \text{ and } \alpha_{t,i} = 1 \\ > 0 \text{ and } \alpha_{t,i} = 0 \end{cases}.$$

Note that $G(1, d_i\left(\kappa_{t,i}, \kappa_{c,i}\right)) = \frac{\partial}{\partial \alpha_{t,i}} d\left(0\right) + d_i\left(\kappa_{t,i}, \kappa_{c,i}\right) \geq d_i\left(\kappa_{t,i}, \kappa_{c,i}\right) > 0$, so the second case for the first order condition can never be satisfied. But then, either $\alpha_{t,i} = 0$, and so it is at least weakly shrunk towards 0, or $\frac{\partial}{\partial \alpha_{t,i}} d\left(1 - \alpha_{t,i}\right) + d_i\left(\kappa_{t,i}, \kappa_{c,i}\right) = 0$

So if $\alpha(d_i\left(\kappa_{t,i}, \kappa_{c,i}\right))$ is implicitly defined by

$$\alpha(d_i\left(\kappa_{t,i}, \kappa_{c,i}\right)) = \begin{cases} 0 & \text{if } G(0, d_i\left(\kappa_{t,i}, \kappa_{c,i}\right)) > 0 \\ \text{solution to } G(\alpha(d_i\left(\kappa_{t,i}, \kappa_{c,i}\right)), d_i\left(\kappa_{t,i}, \kappa_{c,i}\right)) = 0 & \text{if } G(0, d_i\left(\kappa_{t,i}, \kappa_{c,i}\right)) \leq 0 \end{cases},$$

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

Next, we compute $\frac{\partial \alpha(d_i(\kappa_{t,i}, \kappa_{c,i}))}{\partial d_i(\kappa_{t,i}, \kappa_{c,i})}$. If $G(0, d_i\left(\kappa_{t,i}, \kappa_{c,i}\right)) > 0$, then for all $\chi \in [0,1]$ in a neighborhood of $d_i\left(\kappa_{t,i}, \kappa_{c,i}\right)$, we still have $G(0, \chi) > 0$ and $\alpha(\chi) = 0$. So in this case, $\frac{\partial \alpha(d_i(\kappa_{t,i}, \kappa_{c,i}))}{\partial d_i(\kappa_{t,i}, \kappa_{c,i})} = 0$. If $G(0, d_i\left(\kappa_{t,i}, \kappa_{c,i}\right)) \leq 0$, then we can implicitly differentiate $G(\alpha(d_i\left(\kappa_{t,i}, \kappa_{c,i}\right)), d_i\left(\kappa_{t,i}, \kappa_{c,i}\right)) = 0$ with respect to $d_i\left(\kappa_{t,i}, \kappa_{c,i}\right)$ to obtain

$$\frac{\partial^2 d(1 - \alpha(d_i\left(\kappa_{t,i}, \kappa_{c,i}\right)))}{\partial \alpha_{t,i}^2} \frac{\partial \alpha(d_i\left(\kappa_{t,i}, \kappa_{c,i}\right))}{\partial d_i\left(\kappa_{t,i}, \kappa_{c,i}\right)} + 1 = 0$$

and so

$$\frac{\partial \alpha(d_i\left(\kappa_{t,i}, \kappa_{c,i}\right))}{\partial d_i\left(\kappa_{t,i}, \kappa_{c,i}\right)} = -\frac{1}{\frac{\partial^2 d(1 - \alpha(d_i(\kappa_{t,i}, \kappa_{c,i})))}{\partial \alpha_{t,i}^2}} < 0,$$

due to the strict convexity of d. Therefore, $\frac{\partial \alpha(d_i(\kappa_{t,i}, \kappa_{c,i}))}{\partial d_i(\kappa_{t,i}, \kappa_{c,i})} \leq 0$ in general. ■

Lemma 6 *The maximum total similarity satisfies:*

$$\begin{aligned} \partial S\left(t,c\right)/\partial F_c &= 1-d\left(t,c\right)>0, & (31) \\ \partial S\left(t,c\right)/\partial d_i\left(\kappa_{t,i},\kappa_{c,i}\right) &\propto -F_c\cdot\alpha_{t,i}\left(c\right)\leq 0. & (32) \end{aligned}$$

Proof. Observe that

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

Then it follows that

$$S(t,c)=F_c\cdot\left(1-d\left(t,c\right)\right) \tag{33}$$

and (31) follows immediately.

Finally,

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

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

Proof of Proposition 2. Equation (8) follows by, e.g., Lemma 1 in McFadden

[94]. Then, using Equation (8) we calculate

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

Combining this with Lemma 6 (in particular, Equations (31) and (32)) and using the chain rule immediately yields the desired result. ■

Proof of Proposition 3. For (i), note that following Proposition 2,

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

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

Now, consider two different individuals. We first recall that by Proposition 2,

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

Observe that $d\left(t,c\right)$ 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 $F_{c^*}\left(j\right) \neq F_{c^*}\left(j'\right)$. Without loss of generality, let $F_{c^*}\left(j\right) > F_{c^*}\left(j'\right)$. Now select some arbitrary category $c' \in C \setminus \{c^*\}$, and we have

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

Therefore

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

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

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

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

For (ii), recall from the proof of Proposition 2 that

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

and also the fact that, by Equation (33), $\frac{\partial S(t,c)}{\partial d(t,c)} = -F_c$, we have

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

Then applying Proposition 2 again, we get

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

So

$$\frac{\partial^2 \Pr(c|t)}{\partial d(t,c)\,\partial F_c} \le 0 \iff \lambda F_c\,(1-d(t,c))\,[2\Pr(c|t)-1] \le 1$$

Since $\Pr(c|t)$ is increasing in F_c, the left hand side of the condition is monotonically increasing in F_c. The condition holds for F_c small enough (and whenever $\Pr(c|t) \le 1/2$), but does not hold for F_c large enough, which in particular guarantees $\Pr(c|t) > 1/2$ (from Equation (8) the latter condition holds provided $exp\,(\lambda F_c\,(1-d(t,c))) > \sum_{c'\neq c} exp\,(\lambda F_{c'}\,(1-d(t,c'))))$. Therefore, there exists a threshold F_c^*, which increases in $d(t,c)$, such that the condition holds if and only if $F_c < F_c^*$, establishing (iii). ■

Proposition 7 *Suppose that there are not nontrivial event features and let*

$$\overline{\alpha_i} = \sum \alpha_{t,i}(c)\Pr(c\,|t)\,.$$

At database $(F_{c'})_{c'\in C}$, the average valuation $\overline{v(o)}$ of o satisfies:

$$\frac{\partial\overline{v(o)}}{\partial F_c} = \lambda\,(1-d(t,c))\Pr(c\,|t)\,\langle \alpha_{H,t}(c)-\overline{\alpha_H}, u_o\rangle\,, \tag{34}$$

$$\frac{\partial\overline{v(o)}}{\partial d_i(\kappa_{t,i},\kappa_{c,i})} = -\lambda F_c \alpha_{i,t}(c)\Pr(c\,|t)\,\langle \alpha_{H,t}(c)-\overline{\alpha_H}, u_o\rangle\,. \tag{35}$$

Proof of Proposition 7. Let $\overline{\alpha_{O-c}} = \frac{\overline{\alpha_O}-\alpha_{O,t}(c)\Pr(c|t)}{[1-\Pr(c|t)]}$ be the average attention weights conditional on not being categorized in c. We then have that

$$\begin{aligned}
\frac{\partial \overline{v(o)}}{\partial F_c} &= \left\langle \alpha_{O,t}(c)\partial \Pr\left(c\,|t\right)/\partial F_c - \overline{\alpha_{O-c}}\partial \Pr\left(c\,|t\right)/\partial F_c, u_0\right\rangle \\
&= \left\langle \alpha_{O,t}(c)\partial \Pr\left(c\,|t\right)/\partial F_c - \frac{\overline{\alpha_O} - \alpha_{O,t}(c)\Pr\left(c\,|t\right)}{\left[1-\Pr\left(c\,|t\right)\right]}\partial \Pr\left(c\,|t\right)/\partial F_c, u_0\right\rangle \\
&= \partial \Pr\left(c\,|t\right)/\partial F_c \left\langle \frac{\alpha_{O,t}(c) - \overline{\alpha_O}}{\left[1-\Pr\left(c\,|t\right)\right]}, u_0\right\rangle \\
&= \lambda\left(1 - d\left(t,c\right)\right)\cdot \Pr\left(c\,|t\right)\left\langle \alpha_{O,t}(c) - \overline{\alpha_O}, u_0\right\rangle
\end{aligned}$$

where the fourth equality follows by Proposition 2, proving Equation (34). The proof of Equation (35) is completely analogous. ■

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

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

Let $i \in M_O$ and define $\sigma^* = \frac{\sigma_{\delta,i}}{\sigma_{c,i}}\cdot\frac{1}{F_c}$. If we define

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

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

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

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

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

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

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

then this also characterizes the optimal attention $\alpha_{t,i}(c,\delta,\sigma)$ (see, for example, Proposition K.10 and K.11 in Ok [102]). Also, observe that since d is twice continuously differentiable and $d'(0)=0$, G is continuously differentiable.

Now, we can compute $\partial\alpha_{t,i}(c,\delta,\sigma)/\partial\alpha_{c,i}\geq 0$. If $\alpha_{t,i}(c,\delta,\sigma)=0$, an increase of $\alpha_{c,i}$ can only weakly increase $\alpha_{t,i}(c,\delta,\sigma)$, therefore $\frac{\partial\alpha\left(\alpha_{c,i},\alpha_{\delta,i},\sigma^{*}\right)}{\partial\alpha_{c,i}}\geq 0$. In the case $\alpha_{t,i}(c,\delta,\sigma)>0$, we can implicitly differentiate $G\left(\alpha\left(\alpha_{c,i},\alpha_{\delta,i},\sigma^{*}\right),\alpha_{c,i},\alpha_{\delta,i},\sigma^{*}\right)=0$ with respect to $\alpha_{c,i}$ to get (see, for example, Theorem 9.28 in Rudin [110])

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

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

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 $\alpha_{t,i}(c,\delta,\sigma) = 0$ we immediately get $\frac{\partial\alpha\left(\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right)}{\partial\alpha_{\delta,i}} \geq 0$. If $\alpha_{t,i}(c,\delta,\sigma) > 0$, 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 (see, for example, Theorem 9.28 in Rudin [110])

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

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 compute $\frac{\partial\alpha\left(\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right)}{\partial\sigma^*}$. If $\alpha\left(\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right) = 0$, there are two cases. If $\alpha_{\delta,i} > 0$ we trivially have that a marginal change in σ can only make $|\alpha_{t,i}(c,\delta,\sigma)-\alpha_{\delta,i}|$ weakly smaller. If $\alpha_{\delta,i} = 0$, $d'(0) = 0$ implies that $\alpha_{c,i} = 0$, so that $\alpha_{t,i}(c,\delta,\sigma)$ remains constant at 0 after a change in σ^*, making $|\alpha_{t,i}(c,\delta,\sigma)-\alpha_{\delta,i}|$ weakly smaller.

If $\alpha\left(\alpha_{c,i},\alpha_{\delta,i},\sigma^*\right) > 0$, 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 (see, for example, Theorem 9.28 in Rudin [110])

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

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. By $d'(0) > 0$, $\alpha_{d,i} > \alpha_{c,i}$ implies that $\alpha_{t,i}(c,\delta,\sigma) \in (0,1)$. With this, we can compute $\frac{\partial S(t,c,\delta|\sigma)}{\partial \alpha_{c,i}}$ for $i \in M_O$ using the envelope theorem (see, e.g., Theorem M.L.1 in Mas-Colell, Whinston, and Green [93]) to obtain

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

It therefore follows that

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

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

Next, then

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

where the third step follows from the second part of Proposition 4, and the last step from the immediate observation that $\alpha_{t,i}(c,\delta,\sigma) \in [\alpha_{c,i}, \alpha_{\delta,i}]$. ■

A.2 Bottom-up Contrast in Statistical Problems

In Bordalo, Conlon, Gennaioli, Kwon, Shleifer [21], 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_2 versus a full heads sequence H_1. 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 inf compared to frequency estimation $freq$, 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 U "urn selection" 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 inf is more likely, triggering focus on the green share and base rate neglect. This is in line with the evidence in Bordalo, Conlon, Gennaioli, Kwon, and Shleifer [21].

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.

A.3 Top-Down Contrast and Unstable Similarity

Tversky [133] 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 (n) or Poland (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 $(\alpha_{t,G}, \alpha_{t,P})$ to hedonics entails estimated distance

$$v((a,s)\,(\alpha_{t,G}, \alpha_{t,P})) = \alpha_{t,G} \cdot u_G\,(a,s) + \alpha_{t,P} \cdot u_P\,(a,s)\,,$$

which is used as an inverse measure of similarity. 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 of category G in which geographic features are learned or judged, and problems of category 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 category 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 category 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 5, when the most contrasting category is 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.