

Choice by Contextual Associations*

Jose Apesteguia[†]

Andrea Salvanti[‡]

October 17, 2026

Abstract

Choice behavior is significantly influenced by context. Evidence from cognitive science suggests that this dependency may be driven by associations. In this paper, we propose a choice-theoretic model of contextual associations. We formalize contexts as sets of topics. This allows us to define associations as links between alternatives and topics that directly impact utility. We characterize the model’s empirical content, establish conditions for identification, analyze its comparative statics, and propose several extensions. To demonstrate the model’s portability, we apply it to a setting of probabilistic voting, providing a new, association-based mechanism for political polarization.

Keywords: Associations, Contexts, Topics, Framing, Moral values, Random utility model.

JEL: D01, D03, D60

*We are sincerely grateful to Francesco Cerigioni who helped shape the project’s initial stages and modelling ideas. We are also grateful to Larbi Alaoui, Enriqueta Aragonès, Miguel Ballester, Lukas Bantle, Yusufcan Demirkan, Kfir Eliaz, Ruben Enikolopov, Itzhak Gilboa, Gaël Le Mens, Humberto Llavador, Stefania Minardi, Jawwad Noor, Antonio Penta, Ariel Rubinstein, Marta Serra, Patrick Sewell and Ran Spiegler, as well as to seminar audiences at Princeton University, Tel Aviv University, Université Libre de Bruxelles, Universitat Pompeu Fabra, Lund University, the BSE Summer Forum “Preferences and Bounded Rationality,” the SITE Conference “Psychology and Economics” and the Zurich Workshop on Economics and Psychology for helpful comments and discussion. We thank Raquel Lorenzo, Paul Richter and Maxim Ventura Bolet for outstanding research assistance. Jose Apesteguia gratefully acknowledges financial support from FEDER/Ministerio de Ciencia e Innovación (Agencia Estatal de Investigación) through Grant PID2021-125538NB-I00 and through the Severo Ochoa Programme for Centers of Excellence in R&D (Barcelona School of Economics CEX2024-001476-S).

[†]ICREA, Universitat Pompeu Fabra and BSE. E-mail: jose.apesteguia@upf.edu

[‡]Universitat Pompeu Fabra and BSE. E-mail: andrea.salvanti@upf.edu

1 Introduction

There is overwhelming evidence that associations affect the evaluation of options and influence individual behavior. For example, in marketing, sales are affected by associations between products and values such as prestige or sustainability; in politics, voting behavior is impacted by associations between policies and principles such as freedom, patriotism or democracy; in some media outlets, attitudes toward minorities like immigrants are changed by associations between the specific minority and negative values such as illegality; in finance, stock prices respond to associations between a firm’s name and technological trends like the dotcom era.¹ Crucially, different contexts can activate different associations, thus making the evaluation of options context-dependent.

The variety and importance of the preceding examples underscore the relevance of associations for economic behavior. Yet, to the best of our knowledge, no economic theory has examined the role of contextual associations from first principles. In this paper, we develop a choice-based model of how context shapes choice through the associations it activates. This raises two novel challenges: first, developing a tractable theory of context, and second, understanding how different contexts generate different associations and thereby shape choices. The paper’s main contribution is to provide a first unifying framework for analyzing contexts and associations, with broad applicability across diverse economic settings.

A formal definition of what a context is in general decision problems is difficult and elusive. There may be many features in a context that can be behaviorally relevant, and these features may interact with each other, making the formalization of the behavioral influence of contexts challenging. Our approach entails reducing the dimensionality of the problem by identifying the meaning of sets of features, which we call topics. For instance, politicians’ words, campaign speeches, media texts, and public statements can be analyzed to identify topics such as fairness, patriotism, or democracy; these topics together form the context in which voters evaluate political candidates.² As we discuss in Section 2.1, recent developments in machine learning make this approach empirically feasible and allow us to take topics as primitives of our framework.

Once contexts are formalized as sets of topics, we introduce associations between

¹See Keller (1993) and Hainmueller et al. (2015) for the role played by associations in marketing; Enke et al. (2023) and Lauderdale and Blumenau (2025) for the associations of policies to values; Djourelouva (2023) for the association of minorities to illegality and Cooper et al. (2001) for the association of corporate names to technological trends.

²Throughout the paper, we use the term context to denote the set of active topics in the decision environment. This differs from other uses in which context refers to the composition of the choice set.

alternatives and topics as the main driver of contextual preferences. Our contribution here is to formalize them by way of a link between the topics in a context and the available alternatives. In our model, such associations enter into the utility valuation of an alternative, potentially influencing individual behavior. We then understand contextual utility as being the result of two separate components: a standard context-independent intrinsic utility and a context-dependent associative utility, which in turn is separable in the associative values between a given alternative and the different topics. For example, voters may have an intrinsic preference for a party, while the party’s value is also shaped by its associations with the political topics active in the campaign, such as patriotism, fairness, or democracy.

We embed this utility representation in a Luce choice framework. The resulting model is characterized by restrictions within and across contexts. Within each context, choice satisfies Luce’s independence of irrelevant alternatives. A change in context affects relative choice through the topics it adds and removes. Additive Consistency requires changes in choice probabilities to add up consistently across contexts: whenever the additions and removals of topics cancel out overall, the corresponding changes in relative choice probabilities must cancel out as well. In Section 3, we show that Contextual Luce’s IIA and Additive Consistency are necessary and sufficient for our representation. We then establish the uniqueness of contextual utility, give a sufficient condition for the identification of intrinsic utilities and association values, and derive comparative statics with respect to changes in association values and contexts.

Section 4 is devoted to the study of variations of the base model and their revealed-preference foundations. In Section 4.1 we start by considering the case of associations being triggered by the attributes of the alternative, instead of the whole alternative. This represents a natural variation of the model, since it is often the case that associations are activated by certain attributes. For example, when choosing between candidates, demographic attributes such as gender or ethnicity may play a primary role. In Section 4.2, we consider the case in which topics may be present in a context with different intensities. Specifically, we model a context by means of a distribution over the set of possible topics, and the associations between alternatives and topics are modulated by the corresponding topic weights.

Throughout the paper, we discuss the applicability of our framework to a wide range of settings, including framing effects, risk attitudes, moral values, discrimination, priming, identity, and marketing. Furthermore, Section 5 illustrates the framework through a model of probabilistic voting and contextual polarization. Here, the context is given by the political topics at stake. Voters associate parties with the topics on which

they are most salient, and this association is stronger when their moral values align. We show that polarization emerges endogenously as a “demand-side” phenomenon: purely office-seeking parties diverge in their moral values to align with the voters who perceive them as most salient. Crucially, this mechanism generates polarization among ex-ante symmetric candidates, offering a novel rationalization for empirical findings such as those documented by Enke (2020), without relying on drivers such as policy motivation or asymmetric party characteristics.

In Section 6 we conclude and briefly discuss further possible extensions. Together, they underscore the flexibility of our framework and its potential for diverse future applications.

Related Literature. The significance of the decision environment in influencing choice behavior has been established in the foundational works of Tversky and Kahneman (1981), with Salant and Rubinstein (2008) and Bernheim and Rangel (2009) providing key theoretical frameworks. Our contribution to this literature is twofold. First, we formalize the notion of a context in terms of topics and identify associations as a tractable mechanism through which contexts influence choice. Second, we embed this mechanism in a model that restricts the algebraic structure across contexts, thereby linking them behaviorally and delivering testable cross-context predictions.

Associations play a central role in models of cue-dependent consumption and addiction (Laibson, 2001; Bernheim and Rangel, 2004), as well as in models of memory and belief formation (Mullainathan, 2002; Enke et al., 2020; Noor, 2021; Conlon and Kwon, 2025; Gayer et al., 2026). Our contribution is to represent contexts as sets of topics and provide behavioral restrictions linking choice across contexts through associations between topics and alternatives.

Our model also speaks to the growing literature on narratives, broadly understood as stories that individuals tell themselves and others to make sense of the world. Following Shiller (2017), the study of narratives in economics has been increasingly formalized, with applications ranging from persuasion (e.g., Spiegler, 2016; Eliaz and Spiegler, 2020; Schwartzstein and Sunderam, 2021; Aina, 2025) to cultural transmission (e.g., Bénabou et al., 2018). In our framework, narratives can be understood as stories that shape the association value between alternatives and topics, thereby influencing decision-making by reshaping the relative value of alternatives in different contexts. With this definition, we contribute to this literature by providing a choice-theoretic model for analyzing association-based narratives, rather than emphasizing the role of narratives in altering agents’ beliefs.

Building on Moral Foundations Theory (Haidt, 2012), research in social psychology shows that the framing of moral values significantly impacts the valuation of alternatives. For instance, it has been shown that conservatives express greater support for environmental protection when pro-environmental messages are framed in terms of ‘purity’ rather than ‘harm,’ or greater support for same-sex marriage when framed in terms of ‘loyalty’ rather than ‘fairness’ (Feinberg and Willer, 2015; Wolsko et al., 2016). By treating moral values as topics, our paper offers a formal structure to rationalize these shifts. Relatedly, our framework provides a theoretical foundation for empirical findings in economics linking morality to consumer decisions (e.g., De Pelsmacker et al., 2005; Loureiro and Lotade, 2005; Bartling et al., 2015; Hainmueller et al., 2015) and voting behavior (e.g., Bornschier et al., 2021; Buser, 2025). In this domain, the closest contribution is Minardi et al. (2025), who axiomatize an additive utility function composed of an alternative’s intrinsic utility and a value-consistency indicator. While they treat the relevance of a value as fixed, our contribution is to model the link between values and alternatives as context-dependent, allowing the same alternative to trigger different moral associations depending on the active context.

A complementary strand of research on salience examines how decision makers overweigh aspects of the environment that attract attention, thereby distorting evaluations and choices. This mechanism has been formalized in the context of extreme payoffs in risky choice (Bordalo et al., 2012), prominent prices or quality in consumption choices (Bordalo et al., 2013), and salient features of hypotheses in inference problems (Bordalo et al., 2026). In these models, the context is typically endogenous to the menu, defined by the comparison of available alternatives (e.g., against a reference point derived from the choice set). We contribute to this literature by proposing a model where the context is exogenous to the menu, defined by the set of active topics, and where it directly impacts the valuation of alternatives through associative utility rather than by reweighting existing attributes.

From a conceptual standpoint, this paper also connects to the growing trend in behavioral economics aimed at unifying various boundedly rational behaviors (see Enke, 2024 for a comprehensive literature review). While much of this research emphasizes the impact of cognitive limitations on belief formation, our approach complements these efforts by offering a unified model that highlights the role of associations in shaping the perceived value of alternatives.

2 The Model

2.1 From Contexts to Topics

In the last decades, evidence has accumulated on the importance of the context of choice for individual behavior. However, identifying what defines a context and how it affects behavior is challenging. We start by proposing a formalization of contexts that provides a structure through which they may influence behavior in a predictable manner. We then build on this structure by presenting a cognitive channel that specifies how the choice context affects individual behavior.

A context can be seen as a set of potentially many relevant *features*, and to make things more complex, the relevance of a certain feature might depend on the presence or absence of other features. In this paper, we reduce the dimensionality of choice contexts by using higher order elements, that we call *topics*.³ Here, a topic captures the underlying meaning conveyed by a set of features. That is, we exploit the complementarities between features to obtain a set of relevant topics whose meanings are independent of each other. Accordingly, we identify contexts by the set of topics that are present.

The second main ingredient in our framework deals with the influence of contexts on individual behavior. Our central assumption is that associations transfer the value of topics to the different alternatives. This is grounded in research in cognitive science showing that associations, which are largely unconscious and beyond the decision-maker’s volitional control, can directly influence the evaluation of alternatives (Wimmer and Shohamy, 2012; Luettgau et al., 2020). Moreover, the relationship between topics and alternatives may be idiosyncratic, both at the individual and the alternative levels. That is, the influence of a topic may vary across alternatives and across individuals. Given these premises, we model associations as real numbers, reflecting the positive or negative intensity of the association between alternatives and topics.

Having presented the two main building blocks of our framework, we now discuss how to empirically identify the different contexts, while we leave the identification of associations for the following sections. This is needed given that contexts will be treated as primitives of the model while associations and their behavioral meaning are the central object of interest in our identification strategy.

³Topics, or concepts, or ideas have been central to the understanding of human reasoning since the ancient philosophers, and currently play a key role in philosophy, psychology, sociology, linguistics and computer science (see Kunda, 1999, Murphy, 2002 and Hannan et al., 2019).

The rapidly growing literature on computational cognitive science provides practical tools to identify contexts.⁴ Machine-learning algorithms from natural language processing offer a useful illustration. These algorithms learn relationships between linguistic units, such as words or sentences, based on their meaning in texts, and can be used to identify the most relevant topics in a document. Here, the context is the text, and the features are its linguistic units. Although linguistic units may take on different meanings depending on the text in which they are embedded, these algorithms reduce the complexity of a full text to a few relevant topics. Approaches range from dimensionality reduction techniques and static embeddings (e.g., Mikolov et al., 2013; Iliev et al., 2015; Bhatia, 2017; Bhatia et al., 2018; Gentzkow et al., 2019; Enke, 2020; Le Mens et al., 2023) to modern Large Language Models (LLMs), which can now be used to directly extract semantic topics by utilizing their capacity to interpret context-dependent meaning. To illustrate the ease of implementation, we revisit the research on moral framing discussed in our literature review. We focus on the experimental design by Feinberg and Willer (2015), who showed that political arguments resonate differently depending on their moral content. We used ChatGPT 6 Astra to replicate the coding of the moral values embedded in their experimental messages regarding universal health care. Using a fixed prompt to classify the texts according to the five moral foundations of Moral Foundations Theory mentioned above, ChatGPT 6 Astra successfully reproduced the authors’ original classifications along the purity and fairness dimensions.⁵

Similarly, in image analysis, a picture represents a context, the features are all the pixels in the picture, and the algorithm translates sets of pixels into human interpretable sets of topics to ultimately classify the picture (see, e.g., Zhou et al., 2018, Lake et al., 2015, or Ghorbani et al., 2019). For example, Caprini (2024) develops a dictionary-based “visual vocabulary” to analyze media images, showing that partisan visual language is polarized and shapes political attitudes. Recently, LLMs have demonstrated the ability to perform this task directly, extracting complex social topics from raw images with accuracy comparable to human coders.⁶

The relationship between machine learning and our choice-theoretic framework runs in both directions: machine-learning methods identify candidate contextual structures

⁴Topics may alternatively be identified by domain experts or applied researchers. We emphasize machine-learning methods because they provide a systematic and scalable way to extract topics from high-dimensional, unstructured data, such as text and images.

⁵See Appendix C for the exact prompt, analyzed texts and the outcome.

⁶See Santavirta et al. (2025) for a validation of GPT-4V in extracting high-dimensional social features and reproducing human perceptual structures.

from high-dimensional data, while the revealed-preference restrictions of the model provide a behavioral criterion for assessing them. In Appendix B, we establish that if two contextual structures are both consistent with the same observed choice behavior, then, under additional regularity conditions, intrinsic utility is invariant and the contextual structures coincide up to a relabeling of topics. Moreover, we show that if CbCA fails under a proposed contextual structure, this failure cannot generically be repaired by replacing it with another admissible structure.⁷

Our framework, therefore, falls into what Liang (2025) classifies as “machine-learning-augmented” economic theory: we employ machine learning to identify the primitives from high-dimensional data, which we then embed in an economic model to produce testable behavioral implications. This methodological strategy aligns with recent work such as Li and Camerer (2022), who use machine learning to assess the salience of actions in hide-and-seek games, and Compiani et al. (2025), who incorporate unstructured text and image data into demand estimation to recover substitution patterns.

To sum up, we emphasize three points in our approach: (i) the mechanism through which context influences choice operates via the associations it activates; (ii) the activation of associations depends on the meaning of features in the context, which we call topics; and (iii) topics can be identified using recent advances in computational cognitive sciences. Given the previous discussion, in the rest of the paper we abstract from features and directly talk about contexts and topics.

2.2 Choice by Contextual Associations

Let X be the finite set of alternatives, with typical elements denoted by $x, y, z \in X$, and $|X| \geq 2$. We assume a universal domain on menus, given by $\mathcal{X} := \{M \subseteq X : |M| \geq 2\}$. Let K be the finite set of topics, with typical elements denoted by $t, t' \in K$. A context $k \subseteq K$ is a non-empty set of topics, and $\mathcal{D}_K \subseteq 2^K \setminus \{\emptyset\}$ denotes the non-empty domain of observable contexts. Thus, a decision problem is a pair $(M, k) \in \mathcal{X} \times \mathcal{D}_K$. Finally, for every $M \in \mathcal{X}$, every $x \in M$, and every $k \in \mathcal{D}_K$, let $\rho(x, M|k) \in (0, 1)$ denote the probability of choosing alternative x from menu M under context k , with $\sum_{x \in M} \rho(x, M|k) = 1$.

⁷Our “behavioral” approach complements econometric methods that provide robust ways to identify latent topics despite environment-specific linguistic variation (Biderman et al., 2025), thereby limiting dependence on the particular extraction algorithm employed.

Definition 1 (Choice by Contextual Associations) For all $x \in X$ and $k \in \mathcal{D}_K$, let $V(x|k)$ be:

$$V(x|k) = u(x) + \sum_{t \in k} A(x, t),$$

where $u : X \rightarrow \mathbb{R}$ represents intrinsic utility and $A : X \times K \rightarrow \mathbb{R}$ represents the value of the association between alternative x and topic t . A stochastic choice function ρ is a choice by contextual associations (CbCA) whenever for all $M \in \mathcal{X}$, $x \in M$ and $k \in \mathcal{D}_K$:

$$\rho(x, M|k) = \frac{e^{V(x|k)}}{\sum_{y \in M} e^{V(y|k)}}.$$

CbCA uses the convenient and well-understood Luce functional form in every given context k . The novelty is the structure of the model across contexts, given by the contextual utility value $V(x|k)$. The function $V(x|k)$ has two components. The first one, $u(x)$, is context independent and represents the intrinsic value of alternative x , that is, the utility value that is not affected by the context. This component reflects the standard utility representation in economics; if there are no associations, CbCA reduces to the standard Luce framework. The second component of $V(x|k)$, $\sum_{t \in k} A(x, t)$, is the proper association value of x with the topics that constitute context k . The associations follow an additive structure that distorts the intrinsic evaluation of alternatives. It is important to highlight that the association between a topic t and an alternative x , i.e., $A(x, t)$, can be zero. Furthermore, notice that the way we model associations imposes structure across contexts. Varying k directly affects the second component of $V(x|k)$ in a predictable way. For example, eliminating a topic t from context k eliminates the association value of x with t , i.e. $A(x, t)$, keeping active all the other associations of x with topics in $k \setminus \{t\}$.

Here we illustrate the applicability of the model by revisiting the famous *Asian disease* problem which is often used as the main example of framing effects (Tversky and Kahneman, 1981). To recall, in the Asian disease problem the choice of policy is highly influenced by the semantics of the available options. When the options are presented in terms of the lives saved, participants seem to prefer the safe option x over the risky one y , while the reverse is true when the options are presented in terms of deaths. Using our framework, there are two contexts, each one being defined by a single topic, $k_1 = \{Death\}$ and $k_2 = \{Life\}$. In order to focus on the role of associations, let $u(x) = u(y)$. Then, the findings imply that $A(x, Death) < A(y, Death)$ and

$A(x, Life) > A(y, Life)$.⁸ Now, we can study the novel context $k = \{Death, Life\}$, which would imply presenting the alternatives in terms of both lives saved and deaths. Since the context-dependent utility in k incorporates both $A(\cdot, Death)$ and $A(\cdot, Life)$, these opposing associations act as countervailing forces. CbCA thus predicts that $\rho(x, \{x, y\} \mid k_1) < \rho(x, \{x, y\} \mid k) < \rho(x, \{x, y\} \mid k_2)$. Remarkably, several experimental studies report patterns consistent with this behavioral moderation in context k (e.g., Druckman, 2001; Gigerenzer, 2018).⁹ In Section 3.2, we complement this analysis by showing how the parameters can be identified from the data.

In passing, we note that the framework also applies beyond choice to probabilistic judgments and can account for the *conjunction fallacy* (Kahneman and Tversky, 1982). In the Linda problem, subjects often select “feminist bank teller” as more likely than “bank teller”, although the former implies the latter. Topics evoked by Linda’s description, such as feminism and egalitarianism, can generate associations strong enough to favor the former and overturn an intrinsic ranking favoring the latter. Text-based evidence in Bhatia (2017) supports this association pattern.

3 Characterization, Identification and Comparative Statics

3.1 Axiomatic Characterization

In this section we delineate the choice behaviors that can be accounted for by CbCA. We do so by proposing two properties on choice: a variation of Luce’s IIA axiom and a new cross-context axiom, Additive Consistency. We show that the two properties together are necessary and sufficient for choice behavior to sit within CbCA.

We maintain the universal domain on menus, given by $\mathcal{X} = \{M \subseteq X : |M| \geq 2\}$.¹⁰ For contexts, however, we allow $\mathcal{D}_K$ to be any non-empty subset of $2^K \setminus \{\emptyset\}$. The following notation is convenient in the analysis throughout: for every $x, y \in M$ and

⁸In this illustration and in Section 3.2, we interpret u and A as representative-agent utilities that rationalize the aggregate choice frequencies observed in these studies.

⁹Prospect theory explains the standard gain-loss reversal via reference-dependence, but in the mixed context k there is no natural reference point and therefore the theory does not deliver an unambiguous prediction. Other models attribute framing effects to mechanisms such as anticipated regret or limited attention (e.g., Loomes and Sugden, 1982; Bordalo et al., 2022). However, they do not yield a determinate prediction in the mixed-frame case k without auxiliary assumptions regarding the integration of conflicting forces.

¹⁰This assumption may be relaxed following Alós-Ferrer and Mihm (2025), who characterize the Luce rule on arbitrary domains.

$k \in \mathcal{D}_K$, $R(x, y, M|k) = \ln \left(\frac{\rho(x, M|k)}{\rho(y, M|k)} \right)$.¹¹ Our first property, Contextual Luce’s IIA (CLIIA), imposes the structure of Luce’s independence of irrelevant alternatives in every possible context k . To emphasize, CLIIA allows for violations of Luce’s IIA across contexts but not within a context.

Contextual Luce’s IIA (CLIIA): For all $k \in \mathcal{D}_K$ and all $x, y \in M \cap M'$, $R(x, y, M|k) = R(x, y, M'|k)$.

The second property, Additive Consistency (AC), imposes structure across contexts. A change in context affects relative choice through the topics it adds and removes. AC requires consistency across such changes: whenever the additions and removals of every topic exactly offset across a collection of contextual comparisons, their effects on log choice odds must also offset.¹²

Additive Consistency (AC): For every pair of distinct alternatives $x, y \in X$, every integer $m \geq 1$, and every two sequences $k_1, \dots, k_m$ and $k'_1, \dots, k'_m$ in $\mathcal{D}_K$,

$$\sum_{i=1}^m (1_{\{t \in k'_i\}} - 1_{\{t \in k_i\}}) = 0 \quad \forall t \in K \Rightarrow \sum_{i=1}^m [R(x, y, \{x, y\}|k'_i) - R(x, y, \{x, y\}|k_i)] = 0.$$

Note that AC has empirical content only when the observed contexts are sufficiently related in their topic composition to permit nontrivial cancellations. Together the two properties give empirical content to the model.¹³

Theorem 1 *Let ρ be defined on $\mathcal{X} \times \mathcal{D}_K$. ρ satisfies CLIIA and AC if and only if ρ is a choice by contextual associations.*

We conclude this section by making four remarks on the CbCA framework. First, we note that $V(x | k)$ is unique up to a context-dependent constant, and that under an additional richness condition u and A are uniquely pinned down. This delivers a clear identification of the model’s parameters, which we discuss in Section 3.2.

¹¹Note that this ratio is well-defined since by assumption $\rho(x, M|k) > 0$ for every $x \in M \subseteq X$ and every $k \in \mathcal{D}_K$.

¹²AC is related to the cancellation conditions used to characterize additive representations in measurement theory (Scott, 1964); here the restriction applies to changes in log choice odds across contexts.

¹³All proofs are in the Appendix.

Remark 1 [Uniqueness of V] By CLIA, $V(x|k)$ is unique up to a constant that is alternative independent, but possibly context-dependent. Then, for every $M \in \mathcal{X}$, $x \in M$ and $k \in \mathcal{D}_K$,

$$\rho(x, M|k) = \frac{e^{V(x|k)}}{\sum_{y \in M} e^{V(y|k)}} = \frac{e^{V'(x|k)}}{\sum_{y \in M} e^{V'(y|k)}}$$

where $V'(x|k) = V(x|k) + \beta(k)$, with $\beta(k) \in \mathbb{R}$. As shown in Step 1 of the proof in Appendix A.1, taking any $y_0 \in X$ and imposing the normalization $V(y_0|k) = 0$ for all $k \in \mathcal{D}_K$ uniquely determines $V(x|k) = R(x, y_0, X|k)$. The decomposition of this normalized value function into intrinsic utility and association values need not be unique. A sufficient condition for its uniqueness is the following.

Richness: For any $t \in K$, there exist $k, k' \in \mathcal{D}_K$ such that $t \in k$ and $k' = k \setminus \{t\}$.

Under Richness, differencing any two contexts that vary only in the presence of topic t pins down $A(x, t)$ for every x and t , and $u(x)$ then follows as the residual.

Second, as discussed in Section 2.1, we treat the set of topics K and domain of contexts $\mathcal{D}_K$ as primitives of the model. From a theoretical standpoint, it is then natural to ask whether the same choice data can be rationalized under different contextual structures, and whether a change in contextual structure can restore CbCA when it fails.

Remark 2 [Uniqueness and testability of the contextual structure] Suppose two distinct machine-learning algorithms identify different domains of contexts, $(K, \mathcal{D}_K)$ and $(K', \mathcal{D}'_{K'})$, that both rationalize the same observable choice behavior ρ under CbCA. In Appendix B, we show that, when both domains satisfy Richness and the two additional assumptions stated there hold, the agent’s normalized intrinsic utility is invariant across algorithms, and there is a direct one-to-one translation between the two sets of identified topics. In other words, both representations ultimately capture the same underlying primitives with identical behavioral meanings. One might also worry that, if CbCA fails under one contextual structure, the model could be restored simply by adopting another. We show that, under the conditions considered in Appendix B, such a repair is generically impossible.

Third, we highlight that we do not assume the existence of the empty context.

Remark 3 [The empty context] In general settings, the empty context is better understood as a theoretical construct rather than a readily identifiable context in practice.

Therefore, we have chosen not to assume its existence. In Section 3.3, we discuss how a neutral context can play an instrumental role in analyzing the influence of an entire context on choice behavior.

Finally, we comment on the adoption of the logit functional form.

Remark 4 [The Logit functional form] CbCA inherits Luce’s IIA, that may be restrictive in settings with highly similar alternatives, as in the classic “red bus/blue bus” problem. As noted by Apesteguia and Ballester (2018), the logit model may also fail to satisfy certain monotonicity properties in some economic applications. Nevertheless, the logistic function has several advantages, most notably its analytical tractability and widespread applicability, which have made it a benchmark across many fields. These include empirical work on the role of social responsibility and morality in choice (e.g., Bartling et al., 2015; Hainmueller et al., 2015) and in voting (e.g., Bornschier et al., 2021; Buser, 2025), both of which are relevant for our political economy application (see Section 5). Moreover, the Luce framework provides a baseline for richer models that relax its core assumptions, as illustrated by advances in rational inattention (Matějka and McKay, 2015), nested logit models (Kovach and Tserenjigmid, 2022), moderate utility models (He and Natenzon, 2024), and weighted linear choice models (Chambers et al., 2025).

3.2 Identification

We now analyze the extent to which the model’s parameters, namely the context-independent utilities $u(x)$ and the association values $A(x, t)$, can be identified from observed choices. This is important both for interpretation and to enable out-of-sample predictions.

For the identification analysis in this subsection, we assume Richness and normalize $V(y_0 \mid k) = 0$ for all $k \in \mathcal{D}_K$, for a baseline alternative $y_0 \in X$, as introduced in Remark 1. For every $t \in K$, Richness provides contexts $k, k' \in \mathcal{D}_K$ such that $t \in k$ and $k' = k \setminus \{t\}$. Hence, $A(y_0, t) = V(y_0 \mid k) - V(y_0 \mid k') = 0$ for every $t \in K$, and $u(y_0) = 0$ follows from the representation. Adding topic t to k' alters the choice probability of an alternative $x \in X$, relative to alternative y_0 , by

$$\zeta(\{x, y_0\}, X \mid k', t) = R(x, y_0, X \mid k' \cup \{t\}) - R(x, y_0, X \mid k') = A(x, t). \quad (1)$$

Equation (1) yields one identifying restriction for each $x \in X \setminus \{y_0\}$ and $t \in K$, for a total of $(|X| - 1) \times |K|$ restrictions. Now, we use the identified association values to

identify the intrinsic utilities. Consider the additional $|X| - 1$ equations that capture the difference in value between any alternative $x \in X \setminus \{y_0\}$ and the alternative y_0 , $V(x|k) - V(y_0|k) = \ln \left(\frac{\rho(x, X|k)}{\rho(y_0, X|k)} \right)$, which leads to

$$u(x) = \ln \left(\frac{\rho(x, X|k)}{\rho(y_0, X|k)} \right) - \sum_{t \in k} A(x, t). \quad (2)$$

Then, by the $(|X| - 1) \times |K|$ conditions from equation (1) and the $|X| - 1$ conditions from equation (2), it is possible to retrieve $A(x, t)$ and $u(x)$ for all $x \in X \setminus \{y_0\}$ and all $t \in K$ from choice data.

Let us return to the Asian disease example to illustrate the identification of the parameters. We use the three contexts $k_1 = \{Death\}$, $k_2 = \{Life\}$, and $k = \{Death, Life\}$, and now allow $u(x)$ and $u(y)$ to differ. Using the data in Druckman (2001) and taking y as the baseline alternative, with $A(y, Death) = A(y, Life) = 0$ and $u(y) = 0$, we compute $\zeta(\{x, y\}, \{x, y\}|k_2, Death)$, which gives:

$$\begin{aligned} A(x, Death) &= \ln \left(\frac{\rho(x, \{x, y\}|k)}{\rho(y, \{x, y\}|k)} \right) - \ln \left(\frac{\rho(x, \{x, y\}|k_2)}{\rho(y, \{x, y\}|k_2)} \right) \\ &= \ln \left(\frac{.436}{.564} \right) - \ln \left(\frac{.681}{.319} \right) = -1.02 < 0. \end{aligned}$$

Similarly, we compute $\zeta(\{x, y\}, \{x, y\}|k_1, Life)$, which gives

$$\begin{aligned} A(x, Life) &= \ln \left(\frac{\rho(x, \{x, y\}|k)}{\rho(y, \{x, y\}|k)} \right) - \ln \left(\frac{\rho(x, \{x, y\}|k_1)}{\rho(y, \{x, y\}|k_1)} \right) \\ &= \ln \left(\frac{.436}{.564} \right) - \ln \left(\frac{.228}{.772} \right) = 0.96 > 0. \end{aligned}$$

Then we can derive $u(x)$ through the following equation:

$$u(x) = \ln \left(\frac{\rho(x, \{x, y\}|k_1)}{\rho(y, \{x, y\}|k_1)} \right) - A(x, Death) = \ln \left(\frac{0.228}{0.772} \right) + 1.02$$

which implies that $u(x) = -0.20 < u(y)$. The same value would be found using k_2 or k to retrieve $u(x)$. These values provide a simple interpretation of the aggregate choice pattern: the safe option has lower intrinsic utility but is highly sensitive to context. Its negative association with *Death* and positive association with *Life* help explain the observed reversal in choice shares. The observed moderation in the mixed context is consistent with these estimates. In line with Peters and Levin (2008), the

framing effect in the Asian disease problem can thus be interpreted as the interaction of contextual associations and intrinsic valuations. Our framework offers a structured way to decompose these forces at the aggregate level.

3.3 Comparative Statics

We now study the behavioral consequences of changes in associations, from three perspectives: (i) variations of the association value, (ii) the addition of a new topic, and (iii) the impact of the full context.

First, notice that, *ceteris paribus*, an increase in the value of the associations of alternative $x \in M$ unambiguously increases the probability of choosing x and reduces the probability of choosing any other alternative $y \in M \setminus \{x\}$. That is, for any M and k with $t \in k$,

$$\frac{\partial \rho(x, M|k)}{\partial A(x, t)} = \rho(x, M|k)(1 - \rho(x, M|k)) > 0, \text{ and} \quad (3)$$

$$\frac{\partial \rho(y, M|k)}{\partial A(x, t)} = -\rho(x, M|k)\rho(y, M|k) < 0. \quad (4)$$

Interestingly, (3) describes a symmetric, inverse U-shaped curve, with its maximum attained at $\rho(x, M|k) = 1/2$. Hence, extreme values of $\rho(x, M|k)$ will be mildly affected by changes in the association value, while intermediate values will experience larger consequences. As for (4), note that it is always negative, and it is increasingly so with both $\rho(x, M|k)$ and $\rho(y, M|k)$.

Second, we study the consequences of adding a new topic t to a given context k' on the choice probability of an alternative x . The overall impact of adding a new topic intuitively depends on the values of its newly incorporated associations, namely, on whether alternative x gains a greater advantage relative to the other alternatives. To account for that we introduce the following measure,

$$\psi(x, M|k', t) = A(x, t) - \ln \left(\sum_{y \in M} \rho(y, M|k') \cdot e^{A(y, t)} \right),$$

and argue that we can expect a positive impact of adding topic t to context k' on the choice probability of x whenever $\psi(x, M|k', t) > 0$.¹⁴ In particular, the probability of choosing x increases when its association with the new topic is sufficiently strong

¹⁴In Appendix A.2 we provide the mathematical argument.

relative to the associations of the other alternatives in the menu, taking into account their initial choice probabilities.

Finally, we analyze the influence of a whole context on choice. For this comparison, we assume Richness and require a “reference context” in which evaluations are *not distorted* by associations. Accordingly, for every $M \in \mathcal{X}$, we assume the existence of a neutral context $n_M \in \mathcal{D}_K$ such that, for every $k \in \mathcal{D}_K$, $k \cup n_M \in \mathcal{D}_K$ and, for every $x \in M$, $\rho(x, M \mid k \cup n_M) = \rho(x, M \mid k)$. It follows that $\rho(x, M \mid n_M) = \frac{e^{u(x)}}{\sum_{y \in M} e^{u(y)}}$ for every $x \in M$.¹⁵ We first quantify the influence of the associations triggered in context k from a revealed-preference perspective using the log-ratio of choice probabilities between k and n_M :

$$ch(x, M \mid k) = \ln \frac{\rho(x, M \mid k)}{\rho(x, M \mid n_M)}.$$

We now seek to identify the model-based counterpart to this observable measure. Accordingly, we define the average utility of a menu as the log of the average exponential value: $\hat{u}(M) = \ln \left(\frac{1}{|M|} \sum_{y \in M} e^{u(y)} \right)$ and $\hat{V}(M, k) = \ln \left(\frac{1}{|M|} \sum_{y \in M} e^{V(y \mid k)} \right)$. Using these aggregates, we define the model-based measure $md(x, M \mid k)$ as the extent to which the change in alternative x ’s value triggered by the associations in context k exceeds the average change in the menu:

$$md(x, M \mid k) = \ln \frac{e^{V(x \mid k)} / e^{u(x)}}{e^{\hat{V}(M, k)} / e^{\hat{u}(M)}}.$$

The following proposition confirms that this model-based measure is exactly equivalent to the observable variation in choice. This result highlights that a decision maker is “more influenced” by associations with respect to alternative x when the context produces a larger alternative-specific associative effect for x than the average associative effect across the menu.

Proposition 1 *For every $M \in \mathcal{X}$, $x \in M$, and every $k \in \mathcal{D}_K$, $ch(x, M \mid k) = md(x, M \mid k)$.*

Let us use the familiar setting of risk preferences to illustrate. In the following example, individual 1’s stronger associations with risk lead to more pronounced risk-averse behavior in one context and risk-loving behavior in the other. To appreciate

¹⁵To see this, take $x, y \in M$. The neutral context and CbCA imply $\sum_{t \in n_M \setminus k} [A(x, t) - A(y, t)] = 0$ for every $k \in \mathcal{D}_K$. For any $t \in n_M$, Richness provides $k, k' \in \mathcal{D}_K$ with $t \in k$ and $k' = k \setminus \{t\}$. Subtracting the two equalities gives $A(x, t) = A(y, t)$. Thus, $\sum_{t \in n_M} A(x, t)$ is independent of $x \in M$ and cancels from the choice probabilities.

this, consider a toy example involving a risky lottery y and a safe, degenerate lottery x paying the expected value of y with certainty, and let $u_i(x) = u_i(y)$ for $i = 1, 2$. We consider two contexts: an insurance context $k_1 = \{Danger\}$ and a casino context $k_2 = \{Joy\}$. Let $A_1(y, Danger) < A_2(y, Danger) < 0$, reflecting individual 1's stronger dislike of risk in k_1 , and $A_1(y, Joy) > A_2(y, Joy) > 0$, capturing individual 1's stronger affinity for risk in k_2 . For the safe option, let $A_i(x, \cdot) = 0$. In the insurance context k_1 , we have:

$$ch_1(x, \{x, y\}|k_1) = md_1(x, \{x, y\}|k_1) > md_2(x, \{x, y\}|k_1) = ch_2(x, \{x, y\}|k_1) > 0.$$

Both inequalities rely on the fact that the measures simplify to $\ln\left(\frac{2}{1+e^{A_i(y, Danger)}}\right)$, which is strictly decreasing in $A_i(y, Danger)$. The first inequality follows directly from our assumption that $A_1(y, Danger) < A_2(y, Danger)$. The second inequality follows because $A_2(y, Danger) < 0$. This implies that purely based on associations, both individuals exhibit risk-aversion, but individual 1 does so more intensely.

Clearly, the same reasoning applies to the casino context k_2 , with reverse behavioral effects. We note, in passing, that CbCA rationalizes the seemingly puzzling behavior of an individual who both insures against certain risks and gambles in a casino, simply by the activation of distinct contexts that trigger opposing associations with risk.

4 More Structure into Choice by Contextual Associations

In this section we propose two variations of the CbCA model, where we impose different structures on the nature of contextual associations in order to address special settings of interest. The first variation takes into account the possibility that associations are driven by the attributes that compose the different alternatives. In the second variation we allow topics to be present in a context with varying degrees of intensity.

4.1 Attribute-Based Associations

On occasions, alternatives are defined by a collection of *attributes*, which in turn may trigger associations. The color, the size, the geographical origin, the brand or other attributes of the alternatives may represent the mediator in the associations between alternatives and topics. This may involve novel patterns of associations and, accordingly,

in this section we build on our framework to build a model of choice by attribute-based associations.

Let $\mathcal{A}$ be a finite set of attributes, with a typical element denoted by $a \in \mathcal{A}$. An alternative x is defined as a non-empty subset of attributes, $x \subseteq \mathcal{A}$. Let $X \subseteq 2^{\mathcal{A}} \setminus \{\emptyset\}$ denote a finite set of alternatives containing at least one non-singleton alternative. We say that ρ is a choice by contextual attribute-based associations (CbCABA) whenever $V(x|k)$ in Definition 1 is replaced by

$$V_a(x|k) = u(x) + \sum_{a \in x} \sum_{t \in k} A_a(a, t),$$

with $A_a : \mathcal{A} \times K \rightarrow \mathbb{R}$ representing the value of the association between attribute a and topic t . That is, $V_a(x|k)$ shares with $V(x|k)$ the separation of intrinsic utility from the value of the associations, but establishes that associations depend on the attributes defining the alternative.

In order to be able to identify the link between attributes and topics we require richer data. We say that X is partitioned if for every non-singleton alternative $x \in X$ there are $x_1, x_2 \in X$ such that $x_1 \cap x_2 = \emptyset$ and $x_1 \cup x_2 = x$. In a similar fashion, we say that the domain of menus $\mathcal{X}$ is partitioned when X is partitioned. Denote by $P(x)$ the set of bipartitions of x , that is $P(x) = \{(x_1, x_2) \in X^2 | x_1 \cap x_2 = \emptyset \text{ and } x_1 \cup x_2 = x\}$. We now use the bipartitions to present the following property.

Attribute Separability (AS) For all $k, k' = k \setminus \{t\} \in \mathcal{D}_K$ with $t \in k$ and for all $x, y \in X$, with $(x_1, x_2) \in P(x)$ and $(y_1, y_2) \in P(y)$, we have that $R(x, y, X|k) - R(x, y, X|k') = [R(x_1, y_1, X|k) - R(x_1, y_1, X|k')] + [R(x_2, y_2, X|k) - R(x_2, y_2, X|k')]$.

AS imposes that the behavioral influence of associations is separable on the attributes of the alternatives. Now, Theorem 2 draws from Theorem 1 to characterize the model of CbCABA.

Theorem 2 *Let ρ be defined on $\mathcal{X} \times \mathcal{D}_K$ with $\mathcal{D}_K$ satisfying Richness and $\mathcal{X}$ partitioned. ρ satisfies CLIA, AC, and AS if and only if ρ is a choice by contextual attribute-based associations.*

Choice by attribute-based associations may be instrumental in providing a framework for the study of discrimination, as it allows for the incorporation of a new potential source: discrimination motivated by contextual associations. To see this, suppose there

are two candidates for a job, x and y , who only differ in their ethnic origin, which is irrelevant for the performance on the job. Imagine then that the topics in a context are more negatively associated with the ethnic origin of x than with that of y . These associations reduce the probability of x being chosen for the job relative to the probability based on intrinsic utilities alone. Unfortunately, such patterns of behavior are not difficult to find.¹⁶

4.2 Weighted Contexts

Topics may be present with varying degrees of *intensity*. For example, in a political campaign, topics such as patriotism or freedom may be most prevalent, and other topics such as equality may also be present in the political discourse but to a lesser extent. In this section we propose a model to study contextual associations where contexts may differ in terms of the weights given to the different topics.

Denote by $\Delta(K)$ the simplex over the set of topics K . A context is defined by a probability distribution over topics, denoted by $\sigma \in \Delta(K)$, where $\sigma(t)$ is interpreted as the intensity, or weight, of topic t in context σ . As before, contexts are observable. The computational approach mentioned in Section 2.1 oftentimes identifies the distribution of topics in a context, hence giving practical support to the assumption of the observability of σ .

We say that ρ , defined on $\mathcal{X} \times \Delta(K)$, is a choice by contextual associations in weighted contexts (CbCAWC) whenever $V(x|k)$ in Definition 1 is substituted by

$$V(x|\sigma) = u(x) + \sum_{t \in K} \sigma(t)A(x, t).$$

That is, the utility of alternative x in context σ is the intrinsic utility of x , $u(x)$, plus the sum of the association values of x with all the possible topics K , weighted by the prevalence of topics in context σ .

The axiomatic foundations of the model follow the same logic as the one of the main model. We characterize this specification using a weighted version of CLIIA and Weighted Independence. Consider, then, the following version of CLIIA, called Weighted Luce's IIA, that also imposes Luce's IIA over every possible context σ .

Weighted Luce's IIA (WLIIA): For all $\sigma \in \Delta(K)$ and all $x, y \in M \cap M'$, it holds

¹⁶As an example, see Lu et al. (2021) for evidence of discrimination against Asian-Americans when individuals are primed using COVID-19.

that $R(x, y, M|\sigma) = R(x, y, M'|\sigma)$.¹⁷

Weighted Independence requires that the behavioral influence of the context is linear in the weights of topics.

Weighted Independence (WI): For all $\sigma, \sigma' \in \Delta(K)$, all $x, y \in X$ with $x \neq y$ and all $\alpha \in [0, 1]$, $\alpha R(x, y, \{x, y\}|\sigma) + (1 - \alpha)R(x, y, \{x, y\}|\sigma') = R(x, y, \{x, y\}|\alpha\sigma + (1 - \alpha)\sigma')$.¹⁸

The two properties are necessary and sufficient to characterize the new specification.

Theorem 3 *Let ρ be defined on $\mathcal{X} \times \Delta(K)$. ρ satisfies WLIIA and WI if and only if ρ is a choice by contextual associations in weighted contexts.*

The CbCAWC model can serve as a framework for analyzing how consumer choices are shaped by an individual’s contextual identity. For example, topics can be understood as identity-related dimensions, while the context is the salience of these dimensions in a given situation. Consider, for instance, a consumer faced with the decision between purchasing lab-grown meat or traditional farm-raised meat. This decision highlights a conflict between two competing dimensions of her identity. On one hand, lab-grown meat resonates with the self-topic of being a “sustainable individual,” linked to topics such as environmental responsibility. On the other hand, farm-raised meat appeals to an identity grounded in tradition, invoking associations with topics such as authenticity. The consumer’s ultimate decision will thus hinge not only on the intrinsic value of the options, but also on the context that shapes which identity dimension becomes most salient and therefore most influential in guiding behavior. In this way, the model allows us to capture how consumer choices are shaped by a person’s contextual identity.¹⁹

¹⁷Where $R(x, y, M|\sigma) = \ln \left(\frac{\rho(x, M|\sigma)}{\rho(y, M|\sigma)} \right)$.

¹⁸We write $\alpha\sigma + (1 - \alpha)\sigma'$ to refer to the resulting context from combining contexts σ and σ' with weights α and $(1 - \alpha)$, respectively, such that the weight of every $t \in K$ is $\alpha\sigma(t) + (1 - \alpha)\sigma'(t)$.

¹⁹Our framework captures the intuition of Akerlof and Kranton (2000), who incorporate identity into the utility function, and is consistent with the empirical evidence in Benjamin et al. (2010) as well as the formal modeling in Pennesi (2024), which show that making different aspects of identity salient can affect individual choices.

5 An Application to Probabilistic Voting

The aim of this section is to show the applicability and tractability of the CbCA framework in an economic problem. We use our baseline model to provide a contextual association-based account of political polarization, understood as parties adopting increasingly dissimilar moral values (Enke, 2020; Desmet and Wacziarg, 2021; Bertrand and Kamenica, 2023). In this framework, two parties represent the alternatives available to voters, and the context is formed by a set of (political) topics, which we assume are identified from the public debate via a computational algorithm. Voters are described by a CbCA stochastic choice function that depends on both their intrinsic utility for a party (partisanship) and an association function that shifts the party’s value depending on the context.

The association function encompasses two main components that have played an increasingly central role in political economy. First, the function captures voters’ tendency to link parties to specific topics, a form of topic ownership or saliency (Petrocik, 1996; Walgrave et al., 2012; Mahieux et al., 2024). Second, this link is modulated by the alignment between parties’ and voters’ moral values (Enke et al., 2023).²⁰

We show that when voters are polarized, in equilibrium parties align with those who exhibit stronger associations between the party and their salient topics, rationalizing the empirical evidence in Enke (2020). This confirms polarization as a “demand-side” phenomenon, yet one driven by contextual associations rather than the standard channels of convex preferences (Kamada & Kojima, 2014), social identity (Bonomi et al., 2021), or the rising weight of cultural topics (Gennaioli & Tabellini, 2023). Crucially, we generate this divergence in a symmetric environment with purely office-seeking candidates, departing from the bulk of the literature where divergence typically requires ex-ante asymmetries (e.g., Schofield, 2007; Buisseret and Van Weelden, 2022) or policy motivations (Besley and Coate, 1997; Callander, 2008).²¹ Sections 5.1 and 5.2 formalize the model and derive the main theoretical result.

5.1 Setting

Keeping the notation of the general model discussed in Section 2.2, we denote by X the set of parties (alternatives in the main model) and by K the set of political topics.

²⁰Although we do not explicitly model policy choices, one may view them as manifestations of underlying moral values.

²¹A notable exception is Bonomi (2024), who shows that risk aversion over office rents (diminishing marginal utility) can drive polarization even among symmetric, office-seeking candidates.

We consider a two-party system, with $M = \{A, B\} \subseteq X$, and an infinite set of voters $\mathcal{I}$. We denote the set of topics raised in the political debate by $k = \{a, b\} \subseteq K$. The utility function of voter $i \in \mathcal{I}$ for party $x \in M$, given the set of topics k , follows Definition 1. In this framework, we interpret $u_i(x)$ as the unconditioned prejudice in favor of a particular party and $A_i : X \times K \rightarrow \mathbb{R}$ as the value that individual i associates with party x when it comes to topic $t \in k$. To focus on the role of associations, we assume that individuals have no intrinsic preferences for any party, which translates into $u_i(x) = 0$ for all individuals and all parties. In what follows, we formalize the association function $A_i(x, t)$.

Positioning in moral values. Following Enke (2020), we replace the policy positions typically found in standard models of multi-dimensional competition with positions on a *morality index*.²² This index captures the relative emphasis on *universalist* values – such as individual rights, justice, and impartial fairness – versus *particularist* values – such as community and tradition. Let $s^i \in \mathbb{R}$ denote the moral position of voter $i \in \mathcal{I}$, where a larger s^i reflects more particularist values. We model the electorate $\mathcal{I}$ as the union of two disjoint populations, $\mathcal{I}_a$ and $\mathcal{I}_b$, each of measure $\frac{1}{2}$, where voters in $\mathcal{I}_a$ on average hold more universalist values.²³ We assume that moral positions are heterogeneous not only across populations, but also within them, and impose the following symmetry condition:

Assumption 1 *For $i \in \mathcal{I}_a$, $s^i = -\alpha + z^i$, while for $i \in \mathcal{I}_b$, $s^i = \alpha + z^i$, with $\alpha > 0$. The variable z^i is i.i.d. according to a unimodal and symmetric pdf f , with bounded support $[-\bar{z}, \bar{z}] \subset \mathbb{R}$, with $\bar{z} > 0$.*

Assumption 1 provides a measure of ex-ante values’ polarization, represented by α , which quantifies the distance between the average positions of voters in the two subgroups. A higher α indicates greater polarization.²⁴ Denoting by F the cdf of z we notice that the cdf of s can be written as $H^{\mathcal{I}_a}(s) = F(s + \alpha)$ for $i \in \mathcal{I}_a$ and $H^{\mathcal{I}_b}(s) = F(s - \alpha)$ for $i \in \mathcal{I}_b$. The positioning of party $x \in M$ on the morality index is $s^x \in [-\alpha - \bar{z}, \alpha + \bar{z}]$.²⁵

²²A one-dimensional representation is also consistent with models in which public debate compresses multidimensional disagreement into a single dimension (Spector, 2000) or rational voter learning aligns issues along the principal axis of partisan disagreement (Vaeth, 2023).

²³For example, $\mathcal{I}_a$ may represent voters living in urban areas, while $\mathcal{I}_b$ may represent voters living in rural areas, consistent with documented urban–rural differences in moral values (Enke, 2020).

²⁴The assumption allows the supports of the moral-position distributions of $\mathcal{I}_a$ and $\mathcal{I}_b$ to overlap whenever $\bar{z} > \alpha$.

²⁵ Assuming that $s^x \in [-\alpha - \bar{z}, \alpha + \bar{z}]$ instead of $s^x \in \mathbb{R}$ is without loss of generality, since parties will

Party-topic salience. The salience of a party depends on the topic $t \in k$ and on the population group to which the voter belongs, reflecting the fact that parties have advantages on different topics and that voters prioritize topics differently. Formally, we assume that salience is measurable with respect to $I \in \{\mathcal{I}_a, \mathcal{I}_b\}$ and $t \in k$, defining the salience of party $x \in M$ on topic $t \in k$ for individual $i \in I$ as $q_t^{I,x} \in \mathbb{R}^+$. We denote by $q^I = (q_t^{I,A}, q_t^{I,B})_{t \in k}$ the salience vector for population $I \in \{\mathcal{I}_a, \mathcal{I}_b\}$. To maintain symmetry and formalize the advantage of each party in different topics, we impose the following assumption.

Assumption 2 *Vector q is strictly positive, satisfies $\sum_{I,t} q_t^{I,x} = 1/2$ for all x , and:*

- (i) *The advantage of party A on topic a satisfies: $q_a^{\mathcal{I}_a,A} - q_a^{\mathcal{I}_a,B} > q_a^{\mathcal{I}_b,A} - q_a^{\mathcal{I}_b,B} > 0$.*
- (ii) *For all distinct candidates $x \neq y$: $q_a^{\mathcal{I}_a,x} = q_b^{\mathcal{I}_b,y}$ and $q_a^{\mathcal{I}_b,x} = q_b^{\mathcal{I}_a,y}$.*

Assumption 2(i) captures two important aspects of salience for party A. The first is party salience, reflecting the historical associative advantage of party A on topic a. That is, $q_a^{I,A} > q_a^{I,B}$ for $I \in \{\mathcal{I}_a, \mathcal{I}_b\}$. The second aspect is topic-salience, arising from the fact that voters with more universalist values pay greater attention to the universalist topic a. This implies that they perceive the difference in party salience on topic a, given by $q_a^{\mathcal{I}_a,A} - q_a^{\mathcal{I}_a,B}$, as greater than the perceived difference for voters in $\mathcal{I}_b$. One interpretation is that voters who attach greater value to a topic are more likely to self-select into media outlets that cover it, leading them to perceive parties as particularly salient on that topic (e.g., Gentzkow and Shapiro, 2010). Assumption 2(ii) renders salience effects symmetric for party B and, together with the normalization of salience, ensures that neither party enjoys a relative or absolute salience advantage. This simplification allows us to isolate the core mechanism of the model, namely the interaction between party-topic salience and positioning.

We now formalize the association between a party $x \in M$ and topic $t \in k$ for voter $i \in I$ as

$$A_i(x, t) \equiv \bar{A}_i(s^x, s^i, q_t^{I,x}) = \left(1 - \left(\frac{s^x - s^i}{2(\bar{z} + \alpha)}\right)^2\right) \cdot q_t^{I,x}, \quad (5)$$

where the term $\left(\frac{s^x - s^i}{2(\bar{z} + \alpha)}\right)^2 \in [0, 1]$ is the standardized squared distance in moral values between party $x \in M$ and individual $i \in I$. The standardization $2(\bar{z} + \alpha)$ guarantees that salience and positioning are measured on the same scale, therefore bearing the

never choose s^x outside the specified interval in equilibrium.

same importance in the maximization program. Importantly, the specification in equation (5) implies a complementarity between how salient the party is and the distance between the individual's and the party's moral values. The idea is that when there is misalignment in values, even highly salient parties may fail to generate a strong association (and vice-versa). Given the CbCA function proposed in Section 2.2, the probability of voter $i \in I$ choosing party $x \in M$, given the set of topics k , the positioning vector (s^x, s^{-x}, s^i) and the salience vector q^I , is defined as

$$\rho_i(x, M|k, s^i, s^x, s^{-x}, q^I) = \frac{e^{\left(1 - \left(\frac{s^x - s^i}{2(\bar{z} + \alpha)}\right)^2\right) \cdot \sum_{t \in k} q_t^{I,x}}}{\sum_{y \in M} e^{\left(1 - \left(\frac{s^y - s^i}{2(\bar{z} + \alpha)}\right)^2\right) \cdot \sum_{t \in k} q_t^{I,y}}}. \quad (6)$$

To simplify notation, we write $\rho_i(x, M|k)$ instead of $\rho_i(x, M|k, s^i, s^x, s^{-x}, q^I)$. Party $x \in M$ chooses $s^x \in [-\alpha - \bar{z}, \alpha + \bar{z}]$ to maximize its expected vote share given the choice of the other party s^{-x} , namely

$$P(x, M|k, s^x, s^{-x}) = \frac{1}{2} \left(\int_{\mathbb{R}} \rho_i(x, M|k) dH^{\mathcal{I}_a}(s) + \int_{\mathbb{R}} \rho_i(x, M|k) dH^{\mathcal{I}_b}(s) \right). \quad (7)$$

5.2 Analysis

We define parties as being polarized in equilibrium if $s^A \neq s^B$, and measure the degree of polarization by the distance $|s^A - s^B|$. The next proposition establishes that a population of voters with heterogeneous moral values, combined with heterogeneity in party-topic salience, leads to polarization.

Proposition 2 *Under Assumptions 1-2, parties are polarized in the unique Nash equilibrium, with $s^A < 0 < s^B$.*

The underlying idea behind the result is that parties adjust their positions to better align with voters who prioritize topics where the party is more salient. This reflects the interplay between positioning and salience: voters favor parties whose positioning aligns with their own, and even more so when party-topic salience is higher.

Two final remarks are in order. First, in the proof of Proposition 2 it is possible to see that if $\alpha = 0$, a standard median voter theorem applies, with parties adopting the same values. Polarization of voters drives polarization of parties and voting behavior in equilibrium. Second, to emphasize how the complementarity between party-topic salience and positioning in moral values drives the association and polarization, we

assumed that neither party had an absolute advantage over the other. As a result, both parties have equal expected vote shares in every equilibrium. However, this assumption is not needed and it can be shown that there exist equilibria in which $s^A < s^B$, but parties have different expected vote shares. For example, asymmetries can arise from differences in the distributions of moral values across the two populations (e.g., $H^{\mathcal{I}_a}(s) \neq 1 - H^{\mathcal{I}_b}(-s)$), from asymmetric saliency terms (e.g., $q_a^{\mathcal{I}_a, A} \neq q_b^{\mathcal{I}_b, B}$) or from different intrinsic values (e.g., $u_i(A) \neq u_j(B)$ with $i \in \mathcal{I}_a$ and $j \in \mathcal{I}_b$).

6 Discussion

This paper develops a model of choice by contextual associations, which formalizes the influence of contextual factors on decision-making. Drawing on the machine learning literature, we maintain that the context can be identified via the set of topics it activates, introducing a structured representation of contexts via topics. Leveraging this structure, our model offers a testable and tractable framework for unifying a range of behavioral phenomena, including, as we have mentioned throughout the paper, framing effects, risk attitudes, moral values, discrimination, priming and identity. Moreover, we demonstrate the model’s applicability through a detailed analysis of political polarization. Looking forward, our framework provides a foundation for exploring other context-dependent behaviors, such as advertising effects or the influence of social norms and culture. We conclude with a brief discussion of two possible extensions of the model and a potential application.

The association function could capture the idea that, perhaps due to cognitive limitations, individuals might only consider a subset of topics. Chong and Druckman (2007) discuss how individuals base their attitudes on a subset of evaluative dimensions. One possibility within our framework is that only the most relevant topics affect contextual utility:

$$V_r(x|k) = u(x) + \sum_{t \in \arg \max(r_x, k)} A(x, t),$$

with $r_x \subseteq K \times K$ being a transitive ranking of topics for x , which represents their alternative-specific relevance, and $\arg \max(r_x, k)$ denoting the set of topics in k that are maximal in this ranking.

To provide a more general account of framing effects mediated by associations, the model can be extended to allow associations to be modulated through two distinct channels. The first, shared with the CbCA model, captures the association value

between topics and alternatives. The second incorporates the influence of how alternatives are presented on the strength of their association with a topic. This may reflect, for example, whether an alternative constitutes the default option, the emphasis placed on it, or its position in the menu. A way to formalize this in our framework would be to replace $V(x|k)$ in Definition 1 with

$$V_\omega(x, M|\pi_M^x, k) = u(x) + \sum_{t \in k} \omega(\pi_M^x, t) A(x, t),$$

with π_M^x describing the presentation of alternative x in menu M , and $\omega(\pi_M^x, t)$ the corresponding weight when combined with topic t .

We close by suggesting that the model can also be used to understand how individuals may exercise some control over the weights they assign to topics.²⁶ Identifying topics as ethical values and a context as the (endogenous) weights individuals assign to each value, as in the weighted context version of our model of Section 4.2, we want to capture situations in which consumers trade off conflicting ethical values that products embody. Continuing with the example presented in Section 4.2, an individual may prioritize authenticity over environmental responsibility to “rationalize” her choice of choosing farm-raised meat. A possible way to analyze this using our model is to consider a finite set of menus $\mathcal{M} = \{M_1, \dots, M_n\}$ and a measure of frequencies $\gamma \in \Delta(\mathcal{M})$. The decision maker selects the weights over different values in a “self-serving” way, according to the following objective function,

$$\max_{\sigma \in \Delta(K)} W(\sigma) \equiv \sum_{M_i \in \mathcal{M}} \gamma(M_i) \sum_{x \in M_i} \rho(x, M_i | \sigma) V(x | \sigma),$$

where $V(x|\sigma) = u(x) + \sum_{t \in K} \sigma(t) A(x, t)$, as in Section 4.2, with u and A normalized as in the proof of Theorem 3.

²⁶See Bernheim et al. (2021) for another account of a theory of chosen preferences.

Appendix

A Proofs

A.1 Proof of Theorem 1

The necessity of the axioms is evident and thus omitted. We organize the proof of sufficiency in three steps.

Step 1. Fix a baseline alternative $y_0 \in X$. Using standard arguments, CLIIA implies that, for every context $k \in \mathcal{D}_K$, we can set $V(y_0 | k) = 0$ and define, for every $x \in X \setminus \{y_0\}$, $V(x | k) = \ln \left(\frac{\rho(x, \{x, y_0\} | k)}{\rho(y_0, \{x, y_0\} | k)} \right)$. It follows that, for every $M \in \mathcal{X}$ and every $x \in M$, $\rho(x, M | k) = \frac{e^{V(x|k)}}{\sum_{y \in M} e^{V(y|k)}}$.

We now show that $V(x|k)$ can be decomposed into a term measuring the intrinsic value of x and another term additively measuring the values of the associations of x with the topics in k , by first establishing the existence of association values (Step 2) and then defining the intrinsic utility (Step 3).

Step 2. Fix $x \in X \setminus \{y_0\}$. We show that there exist association values $A(x, t)$, $t \in K$, such that

$$\sum_{t \in k'} A(x, t) - \sum_{t \in k} A(x, t) = V(x|k') - V(x|k) \quad \text{for every } k, k' \in \mathcal{D}_K. \quad (8)$$

If $|\mathcal{D}_K| = 1$, the claim is immediate, since (8) imposes no nontrivial restriction; hence suppose $|\mathcal{D}_K| \geq 2$. By way of contradiction, suppose there are no association values satisfying these equalities.²⁷ Consider an equality in which some $A(x, t)$ appears with a nonzero coefficient and use it to express $A(x, t)$ as a function of the remaining association values and the corresponding difference in contextual values. Set this equality aside and substitute the resulting expression for $A(x, t)$ into every other equality in which it appears. The remaining system therefore no longer contains $A(x, t)$. Repeat the same procedure for another association value appearing in the remaining system,

²⁷For $|\mathcal{D}_K| = P$, equation (8) gives P^2 equalities over ordered pairs (k, k') , of which P are trivial since they involve the same k and each equation for (k', k) is the negative of that for (k, k') ; thus there are at most $\binom{P}{2}$ distinct nontrivial equalities.

and continue until no association value can be further eliminated. If the final nontrivial equality is $A(x, t) = c$ for some $c \in \mathbb{R}$, set $A(x, t) = c$; if all remaining equalities are $0 = 0$, assign arbitrary values to any still-undetermined association values. Reversing the substitutions then yields a solution. However, since by assumption no solution exists, the procedure must leave an equality of the form $0 = c$ with $c \neq 0$.

Note that for every pair of contexts (k, k') in the equalities (8), topic t contributes $A(x, t)$ to the left-hand side if $t \in k' \setminus k$, and $-A(x, t)$ if $t \in k \setminus k'$. Topics present in both contexts cancel, while topics absent from both do not enter the equality. Thus, the number multiplying each association value $A(x, t)$, its coefficient, is $1_{\{t \in k'\}} - 1_{\{t \in k\}} \in \{-1, 0, 1\}$. Therefore, since each elimination is obtained by multiplying previously derived equalities by rational numbers and adding them together, the resulting contradiction $0 = c$ is itself obtained by multiplying the original equalities (8) by rational numbers and summing them.

Multiplying by a common denominator of the rational numbers used in this combination gives an integer combination of the original equalities. If the equality corresponding to a pair (k, k') is multiplied by a positive integer q , we repeat this pair q times. If it is multiplied by a negative integer q , we reverse the pair and repeat (k', k) $|q|$ times. After carrying out these repetitions and reversals for every pair entering the integer combination, relabel all the resulting pairs as $(k_1, k'_1), \dots, (k_m, k'_m)$, where $m \geq 1$ is the total number of repetitions. Since the left-hand side of the integer combination is zero, the coefficient of every association value $A(x, t)$ in this sum is zero. By construction, these pairs satisfy

$$\sum_{i=1}^m (1_{\{t \in k'_i\}} - 1_{\{t \in k_i\}}) = 0 \quad \text{for every } t \in K,$$

whereas

$$\sum_{i=1}^m [V(x|k'_i) - V(x|k_i)] \neq 0.$$

By Step 1, $V(x|k) = R(x, y_0, \{x, y_0\}|k)$ for every $k \in \mathcal{D}_K$. The last two equations contradict AC applied to x and y_0 . Therefore, the required association values exist.

Step 3. By Step 2, for each $x \neq y_0$, the difference $V(x|k) - \sum_{t \in k} A(x, t)$ is independent of $k \in \mathcal{D}_K$. Define the intrinsic utility $u(x)$ as this common value, and set $u(y_0) = 0$

and $A(y_0, t) = 0$ for every $t \in K$. Then

$$V(x|k) = u(x) + \sum_{t \in k} A(x, t) \quad \text{for every } x \in X, k \in \mathcal{D}_K.$$

Together with Step 1, this establishes the CbCA representation.

A.2 Mathematical Argument for $\psi(x, M|k', t)$

Let $k', k' \cup \{t\} \in \mathcal{D}_K$, with $t \notin k'$ and let $x \in M \subseteq X$. The probability of choosing x increases after adding t if and only if the ratio $\frac{\rho(x, M|k' \cup \{t\})}{\rho(x, M|k')} > 1$, which in turn can be written as

$$\frac{\rho(x, M|k' \cup \{t\})}{\rho(x, M|k')} = \frac{\frac{e^{V(x|k')} e^{A(x, t)}}{\sum_{y \in M} e^{V(y|k')} e^{A(y, t)}}}{\frac{e^{V(x|k')}}{\sum_{y \in M} e^{V(y|k')}}} = e^{A(x, t)} \cdot \frac{\sum_{y \in M} e^{V(y|k')}}{\sum_{y \in M} e^{V(y|k')} e^{A(y, t)}}.$$

We can rewrite the denominator of the second term by factoring out the total sum $\sum_{z \in M} e^{V(z|k')}$ and using the definition of $\rho(y, M|k')$ as follows:

$$\sum_{y \in M} e^{V(y|k')} e^{A(y, t)} = \left(\sum_{z \in M} e^{V(z|k')} \right) \cdot \sum_{y \in M} \rho(y, M|k') e^{A(y, t)}.$$

Substituting this back into the ratio we arrive at:

$$\frac{\rho(x, M|k' \cup \{t\})}{\rho(x, M|k')} = \frac{e^{A(x, t)}}{\sum_{y \in M} \rho(y, M|k') e^{A(y, t)}}$$

Rearranging, for the probability of x to increase it must be the case that $\ln \left(\frac{\rho(x, M|k' \cup \{t\})}{\rho(x, M|k')} \right) > 0$, which in turn implies $A(x, t) - \ln \left(\sum_{y \in M} \rho(y, M|k') \cdot e^{A(y, t)} \right) > 0$, as desired.

A.3 Proof of Proposition 1

The claim follows immediately by noticing the following.

$$\begin{aligned}
ch(x, M|k) &= \ln \frac{\rho(x, M|k)}{\rho(x, M|n_M)} = \ln \frac{e^{V(x|k)} / \sum_{y \in M} e^{V(y|k)}}{e^{u(x)} / \sum_{y \in M} e^{u(y)}} \\
&= V(x|k) - u(x) - \left(\ln \sum_{y \in M} e^{V(y|k)} - \ln \sum_{y \in M} e^{u(y)} \right) \\
&= V(x|k) - u(x) - (\widehat{V}(M, k) - \widehat{u}(M)) \\
&= \ln \frac{e^{V(x|k)} / e^{u(x)}}{e^{\widehat{V}(M, k)} / e^{\widehat{u}(M)}} = md(x, M|k).
\end{aligned}$$

A.4 Proof of Theorem 2

The necessity of the axioms is evident and thus omitted. By Theorem 1, CLIIA and AC imply that there exist functions $u : X \rightarrow \mathbb{R}$ and $A : X \times K \rightarrow \mathbb{R}$ such that, for every $x \in X$ and $k \in \mathcal{D}_K$,

$$V(x|k) = u(x) + \sum_{t \in k} A(x, t).$$

Fix an arbitrary reference alternative $y_0 \in X$ and impose the normalization $V(y_0|k) = 0$ for every $k \in \mathcal{D}_K$. Fix $t \in K$. By Richness, there exist $k, k' = k \setminus \{t\} \in \mathcal{D}_K$ with $t \in k$. AS implies that for any $x, y \in X$ and bipartitions $(x_1, x_2) \in P(x)$ and $(y_1, y_2) \in P(y)$,

$$A(x, t) - A(x_1, t) - A(x_2, t) = A(y, t) - A(y_1, t) - A(y_2, t).$$

Since X contains a non-singleton alternative and is partitioned, this allows us to define

$$\delta(t) := A(y, t) - A(y_1, t) - A(y_2, t),$$

which is independent of the alternative and the bipartition used. By induction on $|x|$, with the base case $|x| = 1$ being trivial, we obtain:

$$A(x, t) = \sum_{a \in x} A(\{a\}, t) + (|x| - 1)\delta(t). \quad (9)$$

For every attribute a occurring in an alternative in X , the partitioned domain ensures that $\{a\} \in X$. Define

$$A_a(a, t) := A(\{a\}, t) + \delta(t)$$

for these attributes, and assign arbitrary values to $A_a(a, t)$ for attributes that do not occur in any alternative. Summing over the attributes of any x ,

$$\sum_{a \in x} A_a(a, t) = \sum_{a \in x} A(\{a\}, t) + |x|\delta(t).$$

Comparing this with equation (9) gives

$$A(x, t) = \sum_{a \in x} A_a(a, t) - \delta(t).$$

Substituting into the value function,

$$\begin{aligned} V(x|k) &= u(x) + \sum_{t \in k} \left(\sum_{a \in x} A_a(a, t) - \delta(t) \right) \\ &= \left[u(x) + \sum_{a \in x} \sum_{t \in k} A_a(a, t) \right] - \sum_{t \in k} \delta(t). \end{aligned}$$

The term $\beta(k) = \sum_{t \in k} \delta(t)$ is independent of the alternative x . Since choice probabilities are invariant to context-dependent translations of utility, the behavior generated by V is identical to that generated by

$$V^*(x|k) = u(x) + \sum_{a \in x} \sum_{t \in k} A_a(a, t).$$

Thus, ρ is a choice by contextual attribute-based associations.

A.5 Proof of Theorem 3

The necessity of the axioms is evident and thus omitted. We prove the sufficiency part. By WLIIA, for any $\sigma \in \Delta(K)$, there exists a value function $V(x|\sigma)$ that rationalizes choice behavior à la Luce. The WI axiom implies that the log-odds ratio $R(x, y|\sigma) = V(x|\sigma) - V(y|\sigma)$ is linear in σ . To ensure that $V(x|\sigma)$ is also linear, we normalize the model by choosing a baseline alternative $y_0 \in X$ and setting $V(y_0|\sigma) = 0$ for all σ .

Since $V(x|\sigma)$ is affine in σ over the simplex $\Delta(K)$, it can be written as $V(x|\sigma) =$

$\sum_{t \in K} \sigma(t) V(x|\mathbf{1}_t)$, that is, as a convex combination of its values at the vertices of the simplex. We can now construct the model's components. Define the intrinsic utility as the average value across pure contexts, $u(x) := \frac{1}{|K|} \sum_{t \in K} V(x|\mathbf{1}_t)$, and the association value as the deviation from this average, $A(x, t) := V(x|\mathbf{1}_t) - u(x)$, yielding the desired CbCAWC representation.

A.6 Proof of Proposition 2

We organize the proof in five steps and we adopt the convention that $\rho_{\mathcal{I}_a}(s^i) = \rho_i(A, M|k, s^i, s^A, s^B, q^{\mathcal{I}_a})$ and $\rho_{\mathcal{I}_b}(s^i) = \rho_i(A, M|k, s^i, s^A, s^B, q^{\mathcal{I}_b})$. Furthermore, by Assumption 1, with a slight abuse of notation, we also write $\rho_{\mathcal{I}_a}(z^i) = \rho_{\mathcal{I}_a}(s^i)$ for $i \in \mathcal{I}_a$ and $s^i = z^i - \alpha$, and we write $\rho_{\mathcal{I}_b}(z^i) = \rho_{\mathcal{I}_b}(s^i)$ for $i \in \mathcal{I}_b$ and $s^i = z^i + \alpha$.

Step 1. Given Assumptions 1 and 2, a Nash Equilibrium always exists. The first order derivative of equation (7) with respect to s^A is

$$- \sum_{I \in \{\mathcal{I}_a, \mathcal{I}_b\}} \sum_{t \in k} q_t^{I,A} \cdot \int_{\mathbb{R}} \frac{(s^A - s^i)}{4(\bar{z} + \alpha)^2} \cdot \rho_I(s^i) \cdot (1 - \rho_I(s^i)) dH^I(s). \quad (10)$$

Denote party A 's best response to s^B by $\bar{s}^A(s^B)$. The derivative in equation (10) is strictly positive at $s^A = -\alpha - \bar{z}$ and strictly negative at $s^A = \alpha + \bar{z}$, so the best response is interior. Equating equation (10) to zero and rearranging gives the following implicit characterization:

$$\bar{s}^A(s^B) = \frac{\sum_{I \in \{\mathcal{I}_a, \mathcal{I}_b\}} \sum_{t \in k} q_t^{I,A} \cdot \int_{\mathbb{R}} s^i \cdot \rho_I(s^i) \cdot (1 - \rho_I(s^i)) dH^I(s)}{\sum_{I \in \{\mathcal{I}_a, \mathcal{I}_b\}} \sum_{t \in k} q_t^{I,A} \cdot \int_{\mathbb{R}} \rho_I(s^i) \cdot (1 - \rho_I(s^i)) dH^I(s)}. \quad (11)$$

where $\rho_I(s^i)$ is evaluated at $s^A = \bar{s}^A(s^B)$. Thus, equation (11) characterizes the best response implicitly, since $\rho_I(s^i)$ depends on both parties' positions.

Now we show that equation (7) is strictly concave and $\bar{s}^A(s^B)$ is therefore the unique maximizer given s^B . Consider the second order derivative of equation (7):

$$\sum_{I \in \{\mathcal{I}_a, \mathcal{I}_b\}} \frac{\sum_{t \in k} q_t^{I,A}}{4(\bar{z} + \alpha)^2} \int_{\mathbb{R}} \rho_I(s^i) (1 - \rho_I(s^i)) \left[2 \left(\frac{s^A - s^i}{2(\bar{z} + \alpha)} \right)^2 \left(\sum_{t \in k} q_t^{I,A} \right) (1 - 2\rho_I(s^i)) - 1 \right] dH^I(s).$$

By Assumption 1, $\left(\frac{s^A - s^i}{2(\bar{z} + \alpha)}\right)^2 \leq 1$. Since $0 < \rho_I(s^i) < 1$ and $0 < \sum_{t \in k} q_t^{I,A} < \frac{1}{2}$ by Assumption 2, the expression in brackets satisfies

$$2 \left(\frac{s^A - s^i}{2(\bar{z} + \alpha)} \right)^2 \left(\sum_{t \in k} q_t^{I,A} \right) (1 - 2\rho_I(s^i)) - 1 \leq 2 \sum_{t \in k} q_t^{I,A} - 1 < 0.$$

All remaining factors are positive, so the second order derivative is strictly negative. The same reasoning applies to party B . Therefore, each party's objective is strictly concave in its own position, so each party has a unique best response.

Moreover, equation (7) is continuous in (s^A, s^B) . Indeed, the individual choice probabilities in equation (6) are continuous in the parties' positions and bounded between zero and one, so continuity of the integrals follows from the dominated convergence theorem. Since the strategy set $S = [-\alpha - \bar{z}, \alpha + \bar{z}]$ is compact, Berge's Maximum Theorem implies that each party's best-response correspondence is nonempty and upper hemicontinuous. By strict concavity of each party's objective in its own position, each best-response correspondence is singleton-valued. Hence the corresponding best-response functions $\bar{s}^A : S \rightarrow S$ and $\bar{s}^B : S \rightarrow S$ are continuous. Define

$$T : S^2 \rightarrow S^2, \quad T(s^A, s^B) = (\bar{s}^A(s^B), \bar{s}^B(s^A)).$$

The map T is continuous and, since S^2 is compact and convex, Brouwer's Fixed Point Theorem guarantees a fixed point $(s^{A*}, s^{B*}) \in S^2$. By the definition of T , each party is then choosing a best response to the other, so (s^{A*}, s^{B*}) is a Nash equilibrium.

Step 2. In every equilibrium, it must be the case that:

- (i) $P(A, M|k, s^A, s^B) = P(B, M|k, s^B, s^A)$, that is, both parties have the same expected vote share.
- (ii) Party A receives the same expected vote share from population $\mathcal{I}_a$ (respectively, $\mathcal{I}_b$) as party B receives from population $\mathcal{I}_b$ (respectively, $\mathcal{I}_a$), i.e.

$$\int_{\mathbb{R}} \rho_{\mathcal{I}_a}(z^i) dF(z) = \int_{\mathbb{R}} (1 - \rho_{\mathcal{I}_b}(z^i)) dF(z) \quad \text{and} \quad \int_{\mathbb{R}} (1 - \rho_{\mathcal{I}_a}(z^i)) dF(z) = \int_{\mathbb{R}} \rho_{\mathcal{I}_b}(z^i) dF(z),$$

To prove step 2(i), assume by contradiction that in equilibrium party A has a smaller expected vote share than party B (the argument for party A having a larger

expected vote share is analogous). Consider the deviation of party A to $s^A = -s^B$. By equation (6) and Assumptions 1-2 we have that, for every $z^i \in [-\bar{z}, \bar{z}]$,

$$\rho_{\mathcal{I}_a}(z^i) = 1 - \rho_{\mathcal{I}_b}(-z^i). \quad (12)$$

Integrating both sides of equation (12) we get

$$\int_{\mathbb{R}} \rho_{\mathcal{I}_a}(z^i) dF(z) = \int_{\mathbb{R}} (1 - \rho_{\mathcal{I}_b}(-z^i)) dF(z) = \int_{\mathbb{R}} (1 - \rho_{\mathcal{I}_b}(z^i)) dF(z),$$

where the last equality comes from the fact that the distribution f is symmetric around 0. Since the measure of both populations is the same, choosing $s^A = -s^B$ implies that both parties have the same expected vote share. Therefore party A would have a profitable deviation, contradicting that in equilibrium a party can have a larger expected vote share than the other.

To prove step 2(ii), since the populations have equal measure, step 2(i) implies

$$\int_{\mathbb{R}} \rho_{\mathcal{I}_a}(z^i) dF(z) + \int_{\mathbb{R}} \rho_{\mathcal{I}_b}(z^i) dF(z) = 1.$$

Rearranging gives both equalities in step 2(ii).

Step 3. Now we show that it must be the case that:

- (i) $s^A = s^B = 0$ is an equilibrium if and only if $\alpha = 0$.
- (ii) If $s^A = s^B > (<) 0$ then $P(B, M|k, s^B, s^A) > (<) P(A, M|k, s^A, s^B)$.

We first prove step 3(i).

[If] When $\alpha = 0$, by Assumption 1, the pdf of $H^I(s)$ coincides with $f(z)$. By equation (6), whenever $s^A = s^B = 0$, the choice probability depends only on the squared distance $(z^i)^2$, implying that $\rho_I(z^i) = \rho_I(-z^i)$ for all $z^i \in [-\bar{z}, \bar{z}]$. Since f is symmetric around 0 and the term $\rho_I(z^i) \cdot (1 - \rho_I(z^i))$ takes identical values at z^i and $-z^i$, it follows that

$$\int_{\mathbb{R}} z^i \cdot \rho_I(z^i) \cdot (1 - \rho_I(z^i)) dF(z) = 0 \quad \text{for all } I \in \{\mathcal{I}_a, \mathcal{I}_b\}.$$

Then, by equation (11) and strict concavity from Step 1, $\bar{s}^A(0) = 0$. The same reasoning holds for party B , with $\bar{s}^B(0) = 0$. Therefore $s^A = s^B = 0$ is an equilibrium with $\alpha = 0$.

[Only if] Assume by contradiction that $s^A = s^B = 0$ is an equilibrium while $\alpha > 0$. By Assumptions 1-2 and equation (6), we have $\rho_{\mathcal{I}_a}(z^i) = 1 - \rho_{\mathcal{I}_b}(-z^i)$ for all $z^i \in [-\bar{z}, \bar{z}]$. It follows that

$$\int_{\mathbb{R}} z^i \cdot \rho_{\mathcal{I}_a}(z^i) \cdot (1 - \rho_{\mathcal{I}_a}(z^i)) dF(z) = - \int_{\mathbb{R}} z^i \cdot \rho_{\mathcal{I}_b}(z^i) \cdot (1 - \rho_{\mathcal{I}_b}(z^i)) dF(z). \quad (13)$$

Substituting equation (13) in the first order condition of the maximization problem we get:

$$\left[\sum_{t \in k} (q_t^{\mathcal{I}_a, A} - q_t^{\mathcal{I}_b, A}) \right] \int_{\mathbb{R}} (z^i - \alpha) \rho_{\mathcal{I}_a}(z^i) (1 - \rho_{\mathcal{I}_a}(z^i)) dF(z) = 0.$$

By equation (6) and Assumption 2, whenever $s^A = s^B = 0$ and $\alpha > 0$, for any $z^i \in (0, \bar{z}]$, the voter at z^i is closer to the party than the voter at $-z^i$, i.e. $|\alpha + z^i| < |\alpha - z^i|$. Consequently, it holds that $\rho_{\mathcal{I}_a}(z^i) > \rho_{\mathcal{I}_a}(-z^i) > \frac{1}{2}$, where the second inequality follows from party A having a higher saliency for population $\mathcal{I}_a$. Observe that the function $\rho_{\mathcal{I}_a}(z^i) \cdot (1 - \rho_{\mathcal{I}_a}(z^i))$ is maximized at $\rho_{\mathcal{I}_a}(z^i) = \frac{1}{2}$ and is strictly decreasing in $\rho_{\mathcal{I}_a}(z^i)$ when $\rho_{\mathcal{I}_a}(z^i) > \frac{1}{2}$. Therefore, since negative realizations of z^i receive more weight relative to positive ones, by the symmetry of F , it follows that

$$\int_{\mathbb{R}} z^i \cdot \rho_{\mathcal{I}_a}(z^i) \cdot (1 - \rho_{\mathcal{I}_a}(z^i)) dF(z) < 0.$$

Since $\alpha > 0$ and $\rho_{\mathcal{I}_a}(z^i) \cdot (1 - \rho_{\mathcal{I}_a}(z^i)) > 0$, we obtain

$$\int_{\mathbb{R}} (z^i - \alpha) \cdot \rho_{\mathcal{I}_a}(z^i) \cdot (1 - \rho_{\mathcal{I}_a}(z^i)) dF(z) < 0.$$

Thus, the integral in the first order condition is strictly negative, while the first factor is nonzero by Assumption 2, a contradiction.

To prove step 3(ii), we show that $s^A = s^B < 0$ implies that party A has a larger expected vote share and hence, by step 2(i), this cannot be an equilibrium. The argument is analogous for $s^A = s^B > 0$, with party B having a greater expected vote share.

If $[-\bar{z} - \alpha, \bar{z} - \alpha] \cap [-\bar{z} + \alpha, \bar{z} + \alpha] = [-\bar{z} + \alpha, \bar{z} - \alpha] \neq \emptyset$, which happens if $\bar{z} > \alpha$, we have two cases:

- **Case 1:** Consider the interval $[-\bar{z} + \alpha, \bar{z} - \alpha]$. First, since $s^A = s^B$, by Assumptions 1-2 and equation (6) we have that $\rho_{\mathcal{I}_a}(s) = 1 - \rho_{\mathcal{I}_b}(s)$ for every $s \in [-\bar{z} + \alpha, \bar{z} - \alpha]$. Second, we can pair every $s \in [-\bar{z} + \alpha, 0]$ with $s' = -s \in [0, \bar{z} - \alpha]$. By the symmetry of f , we have that $dH^{\mathcal{I}_a}(s) = f(s + \alpha) = f(-s - \alpha) = f(s' - \alpha) = dH^{\mathcal{I}_b}(s')$ and $dH^{\mathcal{I}_a}(s') = f(s' + \alpha) = f(-s' - \alpha) = f(s - \alpha) = dH^{\mathcal{I}_b}(s)$. Third, since $s^A = s^B < 0$, the voter at s is closer to the location of the parties than the voter at s' , i.e. $|s^A - s| < |s^A - s'|$. From equation (6), and given that Party A has a salience advantage in $\mathcal{I}_a$, this implies that $\rho_{\mathcal{I}_a}(s) = (1 - \rho_{\mathcal{I}_b}(s)) > (1 - \rho_{\mathcal{I}_b}(s')) = \rho_{\mathcal{I}_a}(s')$, that is, $\rho_{\mathcal{I}_a}(s) - (1 - \rho_{\mathcal{I}_b}(s')) = (1 - \rho_{\mathcal{I}_b}(s)) - \rho_{\mathcal{I}_a}(s') > 0$. Therefore, since symmetry and weak unimodality imply $f(s + \alpha) \geq f(s - \alpha)$ for all $s \in [-\bar{z} + \alpha, 0]$, the preceding pairing argument implies:

$$\int_{-\bar{z} + \alpha}^{\bar{z} - \alpha} \rho_{\mathcal{I}_a}(s^i) dH^{\mathcal{I}_a}(s) \geq \int_{-\bar{z} + \alpha}^{\bar{z} - \alpha} (1 - \rho_{\mathcal{I}_b}(s^i)) dH^{\mathcal{I}_b}(s).$$

- **Case 2:** Consider the intervals $[-\bar{z} - \alpha, -\bar{z} + \alpha]$ and $[\bar{z} - \alpha, \bar{z} + \alpha]$. Again we can pair each $s \in [-\bar{z} - \alpha, -\bar{z} + \alpha]$ with $s' = -s \in [\bar{z} - \alpha, \bar{z} + \alpha]$. First, notice that $dH^{\mathcal{I}_a}(s') = dH^{\mathcal{I}_b}(s) = 0$. Second, by Assumptions 1-2 and equation (6), following a similar reasoning as in Case 1, $dH^{\mathcal{I}_a}(s) = dH^{\mathcal{I}_b}(s')$ and $\rho_{\mathcal{I}_a}(s) > 1 - \rho_{\mathcal{I}_b}(s')$, implying

$$\int_{-\bar{z} - \alpha}^{-\bar{z} + \alpha} \rho_{\mathcal{I}_a}(s^i) dH^{\mathcal{I}_a}(s) > \int_{\bar{z} - \alpha}^{\bar{z} + \alpha} (1 - \rho_{\mathcal{I}_b}(s^i)) dH^{\mathcal{I}_b}(s).$$

Combining the weak inequality in Case 1 with the strict inequality in Case 2, we obtain

$$\int_{\mathbb{R}} \rho_{\mathcal{I}_a}(s^i) dH^{\mathcal{I}_a}(s) > \int_{\mathbb{R}} (1 - \rho_{\mathcal{I}_b}(s^i)) dH^{\mathcal{I}_b}(s),$$

a contradiction to step 2(ii).

If $\bar{z} \leq \alpha$, the supports have disjoint interiors, and the argument in Case 2 applies directly to their full supports.

Step 4. We now show that when $\alpha > 0$, it must be the case that $s^A < s^B$. We proceed by ruling out all other orderings using the profitable deviations established in Step 3(ii).

First, assume $s^A > s^B$. By Step 2(i), equilibrium requires equal expected vote

shares, $P(A, M|k, s^A, s^B) = P(B, M|k, s^B, s^A) = \frac{1}{2}$. If $s^B < 0$, party A can deviate to $s_D^A = s^B$. This results in a tie at a negative location ($s_D^A = s^B < 0$), which by Step 3(ii) yields $P(A, M|k, s_D^A, s^B) > \frac{1}{2}$, a profitable deviation. Conversely, if $s^B \geq 0$, then $s^A > 0$, and party B can deviate to $s_D^B = s^A$. This results in a tie at a positive location ($s^A = s_D^B > 0$), which by Step 3(ii) yields $P(B, M|k, s^A, s_D^B) > \frac{1}{2}$. Thus, in equilibrium, it must be that $s^A \leq s^B$.

Finally, we rule out $s^A = s^B$. If $s^A = s^B = 0$, Step 3(i) implies that this profile cannot be an equilibrium when $\alpha > 0$. If $s^A = s^B \neq 0$, Step 3(ii) implies that one party has a strictly larger expected vote share than the other, contradicting Step 2(i). Therefore, $s^A \neq s^B$. Combining this with the arguments above, every equilibrium satisfies $s^A < s^B$.

Step 5. Since the parties' expected vote shares sum to one, Step 1 implies that party A 's payoff is strictly concave in s^A and strictly convex in s^B . Hence (s^A, s^B) is a Nash equilibrium if and only if, for every $s_D^A, s_D^B \in [-\alpha - \bar{z}, \alpha + \bar{z}]$,

$$P(A, M | k, s_D^A, s^B) \leq P(A, M | k, s^A, s^B) \leq P(A, M | k, s^A, s_D^B). \quad (14)$$

Suppose that (s^A, s^B) and $(\tilde{s}^A, \tilde{s}^B)$ are two Nash equilibria. Then

$$\begin{aligned} P(A, M | k, \tilde{s}^A, s^B) &\leq P(A, M | k, s^A, s^B) \leq P(A, M | k, s^A, \tilde{s}^B) \\ &\leq P(A, M | k, \tilde{s}^A, \tilde{s}^B) \leq P(A, M | k, \tilde{s}^A, s^B). \end{aligned}$$

All inequalities are therefore equalities. Strict concavity in s^A implies $s^A = \tilde{s}^A$, while strict convexity in s^B implies $s^B = \tilde{s}^B$. Thus the Nash equilibrium is unique.

As in Step 2(i), Assumptions 1 and 2(ii), together with the change of variables from z to $-z$, give

$$P(A, M | k, -s^B, -s^A) = 1 - P(A, M | k, s^A, s^B). \quad (15)$$

If (s^A, s^B) is a Nash equilibrium, applying (15) to (14), with deviations $-s_D^B$ and $-s_D^A$, gives

$$P(A, M | k, s_D^A, -s^A) \leq P(A, M | k, -s^B, -s^A) \leq P(A, M | k, -s^B, s_D^B).$$

Since the feasible interval is symmetric about zero, this holds for every feasible s_D^A, s_D^B . Hence $(-s^B, -s^A)$ is also a Nash equilibrium. Uniqueness therefore implies $s^A = -s^B$.

Together with Step 4, this gives $s^A < 0 < s^B$.

B Uniqueness and Testability of the Contextual Structure

Theorem 1 characterizes the empirical content of CbCA when the contextual structure $(K, \mathcal{D}_K)$ is fixed. This section studies whether the model remains well disciplined when the contextual structure itself is allowed to vary. First, we show that if the same observed choice data ρ admit CbCA representations on both $\mathcal{X} \times \mathcal{D}_K$ and $\mathcal{X} \times \mathcal{D}'_{K'}$, then, under the assumptions below, the two representations coincide up to relabeling of topics. Hence, different algorithms may assign different names to topics, but they recover identical behavioral primitives from the same CbCA data. Second, the model remains testable within the class of CLIIA data. If ρ satisfies CLIIA on $\mathcal{X} \times \mathcal{D}_K$ but violates AC, then this failure cannot generically be repaired by moving to another contextual structure satisfying the restrictions imposed below.

Consider two contextual structures $(K, \mathcal{D}_K)$ and $(K', \mathcal{D}'_{K'})$, and let $f : \mathcal{D}_K \rightarrow \mathcal{D}'_{K'}$ be a bijection. The data transported by f are denoted by ρ^f and defined by

$$\rho^f(x, M|k') = \rho(x, M|f^{-1}(k'))$$

for every $k' \in \mathcal{D}'_{K'}$, every $M \in \mathcal{X}$, and every $x \in M$. Thus, the transport only records the data observed at k as data for the context $f(k)$ in the alternative contextual structure. We say that the data admit CbCA representations on the two contextual structures whenever ρ is CbCA on $(K, \mathcal{D}_K)$ and ρ^f is CbCA on $(K', \mathcal{D}'_{K'})$. In this case, we write $u_{\mathcal{D}_K}, A_{\mathcal{D}_K}$ and $u_{\mathcal{D}'_{K'}}, A_{\mathcal{D}'_{K'}}$ for the corresponding representations. Throughout, representations are understood to be normalized at a common baseline alternative $y_0 \in X$, so that $V_{\mathcal{D}_K}(y_0|k) = V_{\mathcal{D}'_{K'}}(y_0|k') = 0$ for every context in the relevant domain.

With this common normalization, we can compare the topics identified by the two algorithms through their association values, independently of the labels assigned to them. To keep the labels used by the two algorithms distinct, we may relabel topics so that $K \cap K' = \emptyset$, without altering the corresponding contextual utilities or choice probabilities. Let $\tilde{K} = K \cup K'$. We then identify topics in $\tilde{K}$ that have identical association values.

For every $t \in \tilde{K}$, define $\bar{A}(\cdot, t) = A_{\mathcal{D}_K}(\cdot, t)$ if $t \in K$, and $\bar{A}(\cdot, t) = A_{\mathcal{D}'_{K'}}(\cdot, t)$ if $t \in K'$. Define the equivalence relation $\sim$ on $\tilde{K}$ by $t \sim t'$ if and only if $\bar{A}(\cdot, t) = \bar{A}(\cdot, t')$,

and let $\hat{K} = \tilde{K} / \sim$ denote the corresponding set of equivalence classes, which we call *primitive topics*. Let $\phi : K \rightarrow \hat{K}$ and $\phi' : K' \rightarrow \hat{K}$ be the mappings that assign each topic to its equivalence class, so that $\phi(t) = [t]$ for every $t \in K$ and $\phi'(t') = [t']$ for every $t' \in K'$. Finally, define $A : X \times \hat{K} \rightarrow \mathbb{R}$ by $A(\cdot, [t]) = \bar{A}(\cdot, t)$. This function is well-defined by the definition of $\sim$, and $\hat{K} = \phi(K) \cup \phi'(K')$.

Universal Primitive Topics (UPT): The pair of representations satisfies UPT if the collection $\{A(\cdot, \hat{t})\}_{\hat{t} \in \hat{K}}$ is linearly independent.

The idea is that $\hat{K}$ collects the topics identified by the two algorithms according to their behavioral meaning. Two topics belong to the same primitive topic if and only if they have identical association profiles. UPT then requires distinct primitive topics to have linearly independent association profiles. If the associations of one primitive topic were expressible as a linear combination of others, its behavioral impact could be reproduced by a combination of those existing topics in an alternative contextual representation. Since the theoretical role of primitive topics in CbCA is precisely to serve as the fundamental, irreducible building blocks of contexts, a linearly dependent “topic” cannot be regarded as genuinely primitive or new. Crucially, $\hat{K}$ and the mappings ϕ and ϕ' are not additional primitives of the model: they are constructed from the association values identified from the choice data.

Topic Non-Redundancy (TNR): A stochastic choice function ρ on a contextual structure $(K, \mathcal{D}_K)$ satisfies TNR if, for any two distinct topics $t, t' \in K$, there exist distinct alternatives $x, y \in M$, a menu $M \in \mathcal{X}$, and contexts k, k' with $t \in k$, $t' \in k'$, and $k, k \setminus \{t\}, k', k' \setminus \{t'\} \in \mathcal{D}_K$, such that $R(x, y, M \mid k) - R(x, y, M \mid k \setminus \{t\}) \neq R(x, y, M \mid k') - R(x, y, M \mid k' \setminus \{t'\})$.

TNR requires distinct topics to have distinct effects on relative choice probabilities, so that no two topics identified by the same algorithm are behaviorally indistinguishable.

The next proposition shows that, under Richness on both domains, UPT and TNR, the two representations coincide up to a relabeling of topics. For example, *global warming* in one representation and *climate crisis* in the other represent the same primitive topic if their associations with every alternative coincide.

Proposition 3 *Suppose ρ is CbCA on $(K, \mathcal{D}_K)$ and, for some bijection $f : \mathcal{D}_K \rightarrow \mathcal{D}'_{K'}$, the transported data ρ^f are CbCA on $(K', \mathcal{D}'_{K'})$. Let the corresponding representations*

be $(K, \mathcal{D}_K, u_{\mathcal{D}_K}, A_{\mathcal{D}_K})$ and $(K', \mathcal{D}'_{K'}, u_{\mathcal{D}'_{K'}}, A_{\mathcal{D}'_{K'}})$, with both domains of contexts satisfying Richness. Suppose that the pair of representations satisfies UPT, that ρ satisfies TNR on $(K, \mathcal{D}_K)$, and that ρ^f satisfies TNR on $(K', \mathcal{D}'_{K'})$. Then:

1. $u_{\mathcal{D}_K}(\cdot) = u_{\mathcal{D}'_{K'}}(\cdot) \equiv u(\cdot)$.
2. There is a bijection $\pi : K \rightarrow K'$ such that $A_{\mathcal{D}_K}(\cdot, t) = A_{\mathcal{D}'_{K'}}(\cdot, \pi(t))$, for all $t \in K$, and $f(k) = \{\pi(t) : t \in k\}$, for all $k \in \mathcal{D}_K$.

Proof of Proposition 3. Throughout the proof, we write $A(\cdot, \hat{t})$ for the association vector of primitive topic $\hat{t} \in \hat{K}$ as constructed above, and use the identities $A_{\mathcal{D}_K}(\cdot, t) = A(\cdot, \phi(t))$ and $A_{\mathcal{D}'_{K'}}(\cdot, t') = A(\cdot, \phi'(t'))$. By definition of ρ^f , and since both representations are Luce representations normalized by $V_{\mathcal{D}_K}(y_0|k) = V_{\mathcal{D}'_{K'}}(y_0|k') = 0$, we obtain

$$V_{\mathcal{D}_K}(x | k) = V_{\mathcal{D}'_{K'}}(x | f(k))$$

for all $x \in X$ and $k \in \mathcal{D}_K$. Equivalently, for every $x \in X$ and every $k \in \mathcal{D}_K$,

$$u_{\mathcal{D}_K}(x) - u_{\mathcal{D}'_{K'}}(x) = \sum_{t' \in f(k)} A(x, \phi'(t')) - \sum_{t \in k} A(x, \phi(t)). \quad (16)$$

Since the left-hand side of (16) does not depend on the context, the equation holds for every $k \in \mathcal{D}_K$. Fix any $t_0 \in K$. By Richness, there exists $k \in \mathcal{D}_K$ such that $t_0 \in k$ and $k \setminus \{t_0\} \in \mathcal{D}_K$. Writing (16) for k and for $k \setminus \{t_0\}$ and taking differences yields

$$A(\cdot, \phi(t_0)) = \sum_{t' \in f(k)} A(\cdot, \phi'(t')) - \sum_{t' \in f(k \setminus \{t_0\})} A(\cdot, \phi'(t')). \quad (17)$$

We now show that $\phi(K) = \phi'(K')$. Suppose, toward a contradiction, that $\phi(t_0) \notin \phi'(K')$. Then equation (17) expresses the vector $A(\cdot, \phi(t_0))$ as a linear combination of vectors in $\{A(\cdot, \hat{t})\}_{\hat{t} \in \phi'(K')}$, contradicting UPT. Hence $\phi(t_0) \in \phi'(K')$ for every $t_0 \in K$, so $\phi(K) \subseteq \phi'(K')$. By the symmetric argument, interchanging the two representations and using f^{-1} , $\phi'(K') \subseteq \phi(K)$. Therefore, $\phi(K) = \phi'(K')$.

Next, we show that both ϕ and ϕ' are injective. Under CbCA, the effect of topic t on the relative choice probabilities of x and y is $A_{\mathcal{D}_K}(x, t) - A_{\mathcal{D}_K}(y, t)$. Since ρ satisfies TNR on $(K, \mathcal{D}_K)$, distinct topics in K cannot have identical association vectors. By the definition of the equivalence relation $\sim$, $\phi(t) \neq \phi(t')$ whenever $t \neq t'$, and ϕ is injective. Since ρ^f satisfies TNR on $(K', \mathcal{D}'_{K'})$, the symmetric argument shows that ϕ' is also injective.

We now construct the bijection $\pi : K \rightarrow K'$. Since $\phi(K) = \phi'(K')$, for each $t \in K$ there exists at least one $t' \in K'$ such that $\phi(t) = \phi'(t')$. Because ϕ' is injective, this t' is unique. Define $\pi(t) = t'$. Then $A_{\mathcal{D}_K}(\cdot, t) = A(\cdot, \phi(t)) = A(\cdot, \phi'(\pi(t))) = A_{\mathcal{D}_{K'}}(\cdot, \pi(t))$. Thus π is well-defined. It is injective because ϕ and ϕ' are injective, and surjective because $\phi(K) = \phi'(K')$. Hence π is a bijection.

We next prove that the intrinsic utilities are identical. Rewrite equation (16) by grouping the right-hand side by primitive topic:

$$u_{\mathcal{D}_K}(\cdot) - u_{\mathcal{D}_{K'}}(\cdot) = \sum_{\hat{t} \in \phi(K)} c_k(\hat{t}) A(\cdot, \hat{t}), \quad (18)$$

where $c_k(\hat{t}) = |\{t' \in f(k) : \phi'(t') = \hat{t}\}| - |\{t \in k : \phi(t) = \hat{t}\}|$. Since ϕ and ϕ' are injective, both counts are in $\{0, 1\}$, so $c_k(\hat{t}) \in \{-1, 0, 1\}$. Because the left-hand side of (18) is independent of k , linear independence implies that $c_k(\hat{t})$ is constant across all $k \in \mathcal{D}_K$; write $c(\hat{t})$ for this constant. Fix any $\hat{t} \in \phi(K)$ and let $t = \phi^{-1}(\hat{t})$, which is well-defined by injectivity. By Richness, there exists $k \in \mathcal{D}_K$ with $t \in k$ and $k \setminus \{t\} \in \mathcal{D}_K$. Evaluating $c(\hat{t})$ at k gives $c(\hat{t}) = |\{t' \in f(k) : \phi'(t') = \hat{t}\}| - 1 \leq 0$, while evaluating it at $k \setminus \{t\}$ gives $c(\hat{t}) = |\{t' \in f(k \setminus \{t\}) : \phi'(t') = \hat{t}\}| \geq 0$. Therefore, $c(\hat{t}) = 0$ for every $\hat{t} \in \phi(K)$. Substituting into equation (18) gives $u_{\mathcal{D}_K}(\cdot) = u_{\mathcal{D}_{K'}}(\cdot) \equiv u(\cdot)$.

Finally, we show that the relabeling also preserves contexts. Fix $k \in \mathcal{D}_K$. Since the intrinsic utilities are identical, equation (16) gives $\sum_{t \in k} A(\cdot, \phi(t)) = \sum_{t' \in f(k)} A(\cdot, \phi'(t'))$. Using $\phi'(\pi(t)) = \phi(t)$ and the fact that π is a bijection, this becomes $\sum_{t \in k} A(\cdot, \phi(t)) = \sum_{t \in \pi^{-1}(f(k))} A(\cdot, \phi(t))$. By UPT, and since ϕ is injective, the vectors $\{A(\cdot, \phi(t))\}_{t \in K}$ are linearly independent. Hence the two sums contain the same terms, so $k = \pi^{-1}(f(k))$. Equivalently, $f(k) = \{\pi(t) : t \in k\}$ for every $k \in \mathcal{D}_K$. $\blacksquare$

The previous proposition shows that, under Richness on both domains, UPT and TNR, whenever two CbCA representations rationalize the same data, the contextual structure is pinned down up to relabeling. We now address the complementary case in which the data have the Luce structure within each context but fail the cross-context restriction imposed by CbCA on the proposed contextual structure. Formally, suppose that ρ satisfies CLIIA on $\mathcal{X} \times \mathcal{D}_K$ but violates AC.²⁸ By Theorem 1, ρ is not CbCA on this contextual structure. The question is whether this failure has empirical content, or whether a CbCA representation can be restored by replacing $(K, \mathcal{D}_K)$ with another

²⁸We focus on AC violations since, if ρ violates CLIIA at some $k \in \mathcal{D}_K$, then, for every bijection $f : \mathcal{D}_K \rightarrow \mathcal{D}_{K'}$, the transported data ρ^f violate CLIIA at $f(k)$.

contextual structure $(K', \mathcal{D}'_{K'})$. The next result shows that, when the alternative representation uses fewer than $|\mathcal{D}_K| - 1$ topics, such a restoration is possible only for a measure-zero set of CLIIA data. Moreover, if each topic in the alternative contextual structure is empirically falsifiable, as defined next, the same measure-zero conclusion holds even without imposing an explicit bound on the number of topics.

Topic Falsifiability (TF): A contextual structure $(K, \mathcal{D}_K)$ satisfies TF if, for every $t \in K$, there exist distinct contexts $k, k' \in \mathcal{D}_K$ such that $t \in k \cap k'$, $k \setminus \{t\} \in \mathcal{D}_K$, and $k' \setminus \{t\} \in \mathcal{D}_K$.

TF requires that the effect of every topic can be compared across at least two different contexts.

Proposition 4 *The set of CLIIA data on $\mathcal{X} \times \mathcal{D}_K$ for which there exist a finite topic set K' with $|K'| < |\mathcal{D}_K| - 1$, a domain of contexts $\mathcal{D}'_{K'}$ with $|\mathcal{D}'_{K'}| = |\mathcal{D}_K|$, and a bijection $f : \mathcal{D}_K \rightarrow \mathcal{D}'_{K'}$ such that ρ^f is CbCA on $(K', \mathcal{D}'_{K'})$, has Lebesgue measure zero in $\mathbb{R}^{|\mathcal{D}_K| \times (|X|-1)}$. Moreover, if the alternative contextual structure $(K', \mathcal{D}'_{K'})$ satisfies TF, then the same conclusion holds without imposing any upper bound on $|K'|$.*

Proof of Proposition 4. Take a baseline alternative $y_0 \in X$ and enumerate $\mathcal{D}_K = \{k_i\}_{i=1}^P$, where $P = |\mathcal{D}_K|$. Let $n = |X| - 1$. By CLIIA and the normalization in Remark 1, each data set is equivalently described by the normalized values

$$V(x|k_i) = \ln \left(\frac{\rho(x, \{x, y_0\}|k_i)}{\rho(y_0, \{x, y_0\}|k_i)} \right),$$

for $i = 1, \dots, P$ and $x \neq y_0$. Thus CLIIA data are identified, after normalization, with the matrix $(V(x|k_i))_{i, x \neq y_0}$ in $\mathbb{R}^{P \times n}$.

Consider an alternative contextual structure $(K', \mathcal{D}'_{K'})$ and a bijection $f : \mathcal{D}_K \rightarrow \mathcal{D}'_{K'}$. Let $b_{t'}^f \in \{0, 1\}^P$ be the incidence vector of topic $t' \in K'$ across the transported contexts, with entries

$$b_{t'}^f(k_i) = |\{t'\} \cap f(k_i)|,$$

and define

$$\mathcal{L}_f = \text{span}\{\mathbf{1}, \{b_{t'}^f\}_{t' \in K'}\} \subseteq \mathbb{R}^P.$$

Let V^f denote the normalized values of the transported data ρ^f . By definition of ρ^f ,

for every $i = 1, \dots, P$ and every $x \neq y_0$,

$$V^f(x|f(k_i)) = \ln \left(\frac{\rho^f(x, \{x, y_0\}|f(k_i))}{\rho^f(y_0, \{x, y_0\}|f(k_i))} \right) = V(x|k_i).$$

If ρ^f is CbCA on $(K', \mathcal{D}'_{K'})$, then, for every $x \neq y_0$, and $i = 1, \dots, P$,

$$V(x|k_i) = V^f(x|f(k_i)) = u(x) + \sum_{t' \in f(k_i)} A(x, t').$$

Thus, for each $x \neq y_0$, the vector $(V(x|k_i))_{i=1}^P$ lies in $\mathcal{L}_f$. The normalized values rationalized under this alternative structure via f are therefore contained in

$$(\mathcal{L}_f)^n = \{(V(x|k_i))_{i,x \neq y_0} \in \mathbb{R}^{P \times n} : (V(x|k_i))_{i=1}^P \in \mathcal{L}_f \text{ for every } x \neq y_0\}.$$

Now, consider the case $|K'| < P - 1$. Then $\dim(\mathcal{L}_f) \leq |K'| + 1 < P$, so $(\mathcal{L}_f)^n$ is a proper linear subspace of $\mathbb{R}^{P \times n}$. There are only finitely many such subspaces: $\mathcal{L}_f$ depends only on incidence vectors in the finite set $\{0, 1\}^P$, and fewer than $P - 1$ of these vectors are selected. Hence the relevant set of CLIIA data is contained in a finite union of proper linear subspaces of $\mathbb{R}^{P \times n}$ and has Lebesgue measure zero.

Consider next an alternative contextual structure $(K', \mathcal{D}'_{K'})$ satisfying TF, with no restriction on $|K'|$. Let C_f be the $P \times (|K'| + 1)$ matrix whose columns are $\mathbf{1}$ and the incidence vectors $\{b_{t'}^f\}_{t' \in K'}$. Its column space is $\mathcal{L}_f$. We show that C_f has rank strictly below P .

Since contexts are non-empty, K' is non-empty. Fix $t' \in K'$. By TF, there exist distinct contexts $k, k' \in \mathcal{D}'_{K'}$ such that $t' \in k \cap k'$, $k \setminus \{t'\} \in \mathcal{D}'_{K'}$, and $k' \setminus \{t'\} \in \mathcal{D}'_{K'}$. Removing t' changes only the column of C_f corresponding to topic t' , and therefore

$$C_f(i(f^{-1}(k)), \cdot) - C_f(i(f^{-1}(k \setminus \{t'\})), \cdot) = C_f(i(f^{-1}(k')), \cdot) - C_f(i(f^{-1}(k' \setminus \{t'\})), \cdot),$$

with $i(f^{-1}(k))$ being the row index of $f^{-1}(k)$. The four contexts k , $k \setminus \{t'\}$, k' , and $k' \setminus \{t'\}$ are distinct. Since f is a bijection, the four corresponding rows of C_f are distinct. The equality is therefore a non-trivial linear dependence among rows of C_f , so $\text{rank}(C_f) < P$. Hence $\dim(\mathcal{L}_f) < P$, and the set of normalized values rationalized under this contextual structure is contained in a proper linear subspace of $\mathbb{R}^{P \times n}$.

It remains to let the contextual structure and f vary, subject to TF. For any such pair, $\mathcal{L}_f$ is generated by $\mathbf{1}$ and by the distinct incidence vectors $b_{t'}^f \in \{0, 1\}^P$. Since $\{0, 1\}^P$ is finite, only finitely many spans of this form can arise. By the preceding

paragraph, every span arising under TF is proper. Hence the normalized values rationalized by some contextual structure satisfying TF are contained in a finite union of proper linear subspaces of $\mathbb{R}^{P \times n}$, and therefore have Lebesgue measure zero. ■

Taken together, Proposition 3 and Proposition 4 clarify the role of the contextual structure in the model. Axioms CLIIA and AC test CbCA on a proposed contextual structure. When the data pass the test, Proposition 3 shows that the successful representation is unique up to relabeling under Richness on both domains, UPT and TNR. When they fail it, Proposition 4 shows that admissible changes in the contextual structure do not generically restore the representation.

C Prompting LLMs

Prompt. Using **Moral Foundations Theory**, analyze the text provided below. For each moral foundation, determine **which side of the moral pair is explicitly present in the text**, if any.

Decision rules:

- A foundation side should be marked as present **only if the text contains clear, explicit evidence**.
- If references are **weak, ambiguous, neutral, indirect, or merely implied**, select “**Neither**.”
- **Do not infer intent, motivations, or unstated values** beyond what is explicitly stated in the text.

Moral Foundations to Evaluate:

- Care vs. Harm
- Fairness vs. Cheating
- Loyalty vs. Betrayal
- Authority vs. Subversion
- Purity vs. Degradation

Output Format (use exactly this structure):

Care / Harm: Care | Harm | Neither
 Fairness / Cheating: Fairness | Cheating | Neither
 Loyalty / Betrayal: Loyalty | Betrayal | Neither
 Authority / Subversion: Authority | Subversion | Neither
 Purity / Degradation: Purity | Degradation | Neither

Text to Analyze: [PASTE TEXT HERE]

Purity Text. The absence of universal healthcare in the United States practically ensures that we will have unclean, infected, and diseased Americans walking among us. The uninsured often develop “diseases of poverty”, such as tuberculosis, hepatitis, toxocariasis, and other viruses or parasites that can easily spread throughout the population. These diseases are disgusting infestations that invade the human body and leech out needed nutrients to survive. Many of these diseases have grotesque symptoms like yellowing of the skin and eyes, coughing up bloody mucus, itchy rashes, and lesions. These diseases are contagious and spread through the population infecting many, including those who are not poor. The spread of these diseases, however, would be easily preventable if all Americans had healthcare. Individuals infected with these contagious diseases would become much less likely to spread their sickness to others, because doctors could provide them with medicine to cure or control the disease, and doctors could educate these sick patients on how to avoid infecting others. According to a 2009 Kaiser Commission on Medicaid there are 50 million Americans who lack any type of healthcare insurance. The 50 million uninsured are especially susceptible to contracting infectious diseases because they are often the poorest among us, suffering from malnutrition and stress, both of which lead to weakened immune systems. So, this means that the uninsured tend to serve as repositories that harbor and cultivate contagions which can then, after incubating, more easily spread to others in the population. Healthcare reform represents a major step towards the extermination of infectious diseases. With universal healthcare, those who are currently uninsured would be able to see doctors and become educated in how to avoid contracting diseases, and those who have contracted disease could get the medicine they need to cure them or at least making them less contagious. Overall, universal healthcare is a way of purifying America from some of its most infectious diseases, making it less and less likely that healthy individuals will ever encounter these diseases. Everyone should have healthcare.

Fairness Text. In its current state healthcare in the U.S. is inherently unfair and unjust. We need reform to ensure that everyone, not just the rich and the fortunate, has access to the doctors and the medicine they need. According to a 2009 Kaiser Commission on Medicaid an estimated 50 million American citizens are without adequate insurance coverage – including millions of children who have done nothing wrong. Because most American families obtain healthcare coverage from long term employers, the blue collar sector of the labor force is unfairly hit the hardest. These laborers typically work for hourly wages in temporary jobs which do not offer comprehensive health insurance. These are honest, law-abiding citizens who are being unfairly denied access to resources their wealthier countrymen are entitled to, simply because they make less money. To make matters worse, the poorest are often the ones who are at the highest risk for health problems. Individuals in the lower economic brackets suffer more diabetes, cancer, asthma, emphysema, hypertension, and heart disease largely due to the financial stress they experience on a day-to-day basis. They are the ones who need healthcare the most, but unjustly are the ones who have the least access to it. Not having healthcare often amounts to the poor getting poorer because when they or their children get sick, they have to pay the enormous costs for healthcare out-of-pocket. A 2001 study in five States found that medical debt contributed to 46.2% of all personal bankruptcies and in 2007, 62.1% of filers for bankruptcies claimed high medical expenses as the cause. Since then, health costs and the numbers of uninsured and underinsured have increased. It all comes down to simple questions: Shouldn't all people equally have the right to see a doctor and get medicine when they are sick? Is it fair to allow people to suffer

from illness that are easily curable, simply because they are poor? Is this justice for all? Access to healthcare is a basic human right not a luxury reserved for the rich. Everyone should have healthcare.

Moral Foundation	Purity Text	Fairness Text
Care / Harm	Care	Care
Fairness / Cheating	Neither	Fairness
Loyalty / Betrayal	Neither	Neither
Authority / Subversion	Neither	Neither
Purity / Degradation	Purity	Neither

Table 1: Moral Foundations Across Texts Retrieved using ChatGPT 6 Astra

References

- Aina, C. (2025). *Tailored stories*, Mimeo.
- Akerlof, G. A., & Kranton, R. E. (2000). Economics and identity. *The Quarterly Journal of Economics*, 115(3), 715–753.
- Alós-Ferrer, C., & Mihm, M. (2025). A characterization of the luce choice rule for an arbitrary collection of menus. *Journal of Economic Theory*, 223, 105941.
- Apesteguia, J., & Ballester, M. A. (2018). Monotone stochastic choice models: The case of risk and time preferences. *Journal of Political Economy*, 126(1), 74–106.
- Bartling, B., Weber, R. A., & Yao, L. (2015). Do markets erode social responsibility? *The Quarterly Journal of Economics*, 130(1), 219–266.
- Bénabou, R., Falk, A., & Tirole, J. (2018). *Narratives, imperatives, and moral reasoning*, National Bureau of Economic Research.
- Benjamin, D. J., Choi, J. J., & Strickland, A. J. (2010). Social identity and preferences. *American Economic Review*, 100(4), 1913–1928.
- Bernheim, B. D., & Rangel, A. (2004). Addiction and cue-triggered decision processes. *American Economic Review*, 94(5), 1558–1590.
- Bernheim, B. D., & Rangel, A. (2009). Beyond revealed preference: Choice-theoretic foundations for behavioral welfare economics. *The Quarterly Journal of Economics*, 124(1), 51–104.
- Bernheim, B. D., Braghieri, L., Martínez-Marquina, A., & Zuckerman, D. (2021). A theory of chosen preferences. *American Economic Review*, 111(2), 720–754.
- Bertrand, M., & Kamenica, E. (2023). Coming apart? cultural distances in the united states over time. *American Economic Journal: Applied Economics*, 15(4), 100–141.
- Besley, T., & Coate, S. (1997). An economic model of representative democracy. *The Quarterly Journal of Economics*, 112(1), 85–114.
- Bhatia, S. (2017). Associative judgment and vector space semantics. *Psychological Review*, 124(1), 1.
- Bhatia, S., Goodwin, G. P., & Walase, L. (2018). Trait associations for hillary clinton and donald trump in news media: A computational analysis. *Social Psychological and Personality Science*, 9(2), 123–130.

- Biderman, D., Blei, D., Cai, W., Ciccarone, A., Feder, A., & Prat, A. (2025). *Fixed effects topic model*, Mimeo.
- Bonomi, G. (2024). Divide and diverge: Polarization incentives. *arXiv preprint*.
- Bonomi, G., Gennaioli, N., & Tabellini, G. (2021). Identity, beliefs, and political conflict. *The Quarterly Journal of Economics*, 136(4), 2371–2411.
- Bordalo, P., Gennaioli, N., & Shleifer, A. (2022). Salience. *Annual Review of Economics*, 14(1), 521–544.
- Bordalo, P., Gennaioli, N., & Shleifer, A. (2013). Salience and consumer choice. *Journal of Political Economy*, 121(5), 803–843.
- Bordalo, P., Gennaioli, N., & Shleifer, A. (2012). Salience theory of choice under risk. *The Quarterly Journal of Economics*, 127(3), 1243–1285.
- Bordalo, P., Conlon, J., Gennaioli, N., Kwon, S., & Shleifer, A. (2026). How people use statistics. *Review of Economic Studies*, 93(1), 250–285.
- Bornschier, S., Häusermann, S., Zollinger, D., & Colombo, C. (2021). How “us” and “them” relates to voting behavior - social structure, social identities, and electoral choice. *Comparative Political Studies*, 54(12), 2087–2122.
- Buisseret, P., & Van Weelden, R. (2022). Polarization, valence, and policy competition. *American Economic Review: Insights*, 4(3), 341–352.
- Buser, T. (2025). Economic preferences, personality, and voting.
- Callander, S. (2008). Political motivations. *The Review of Economic Studies*, 75(3), 671–697.
- Caprini, G. (2024). *Visual bias*, University of Oxford.
- Chambers, C. P., Masatlioglu, Y., Natenzon, P., & Raymond, C. (2025). Weighted linear discrete choice. *American Economic Review*, 115(4), 1226–1257.
- Chong, D., & Druckman, J. (2007). A theory of framing and opinion formation in competitive elite environments. *Journal of Communication*, 57, 99–118.
- Compiani, G., Morozov, I., & Seiler, S. (2025). Demand estimation with text and image data. *arXiv preprint*.
- Conlon, J. J., & Kwon, S. Y. (2025). *Beliefs from cues*, Mimeo.
- Cooper, M. J., Dimitrov, O., & Rau, P. R. (2001). A rose. com by any other name. *The Journal of Finance*, 56(6), 2371–2388.
- De Pelsmacker, P., Driesen, L., & Rayp, G. (2005). Do consumers care about ethics? Willingness to pay for fair-trade coffee. *Journal of Consumer Affairs*, 39(2), 363–385.
- Desmet, K., & Wacziarg, R. (2021). The Cultural Divide. *The Economic Journal*, 131(637), 2058–2088.
- Djourelouva, M. (2023). Persuasion through slanted language: Evidence from the media coverage of immigration. *American Economic Review*, 113(3), 800–835.
- Druckman, J. (2001). Evaluating framing effects. *Journal of Economic Psychology*, 22, 91–101.
- Eliasz, K., & Spiegler, R. (2020). A model of competing narratives. *American Economic Review*, 110(12), 3786–3816.
- Enke, B. (2024). *The cognitive turn in behavioral economics*, Mimeo.
- Enke, B. (2020). Moral values and voting. *Journal of Political Economy*, 128(10), 3679–3729.
- Enke, B., Rodríguez-Padilla, R., & Zimmermann, F. (2023). Moral universalism and the structure of ideology. *The Review of Economic Studies*, 90(4), 1934–1962.

- Enke, B., Schwerter, F., & Zimmermann, F. (2020). *Associative memory and belief formation*, National Bureau of Economic Research.
- Feinberg, M., & Willer, R. (2015). From gulf to bridge: When do moral arguments facilitate political influence? *Personality and Social Psychology Bulletin*, 41(12), 1665–1681.
- Gayer, G., Minardi, S., Wang, F., & Gilboa, I. (2026). Walras–bowley lecture: Association rules—an axiomatic approach. *Econometrica*, *Forthcoming*.
- Gennaioli, N., & Tabellini, G. (2023). *Presidential lecture: Identity politics*, CESifo Working Paper.
- Gentzkow, M., & Shapiro, J. M. (2010). What drives media slant? Evidence from us daily newspapers. *Econometrica*, 78(1), 35–71.
- Gentzkow, M., Shapiro, J. M., & Taddy, M. (2019). Measuring group differences in high-dimensional choices: Method and application to congressional speech. *Econometrica*, 87(4), 1307–1340.
- Ghorbani, A., Wexler, J., Zou, J. Y., & Kim, B. (2019). Towards automatic concept-based explanations. *Advances in Neural Information Processing Systems*, 9273–9282.
- Gigerenzer, G. (2018). The bias bias in behavioral economics. *Review of Behavioral Economics*, 5(3-4), 303–336.
- Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. Pantheon Books.
- Hainmueller, J., Hiscox, M. J., & Sequeira, S. (2015). Consumer demand for fair trade: Evidence from a multistore field experiment. *Review of Economics and Statistics*, 97(2), 242–256.
- Hannan, M. T., Le Mens, G., Hsu, G., Kovács, B., Negro, G., Pólos, L., Pontikes, E., & Sharkey, A. J. (2019). *Concepts and categories: Foundations for sociological and cultural analysis*. Columbia University Press.
- He, J., & Natenzon, P. (2024). Moderate utility. *American Economic Review: Insights*, 6(2), 176–95.
- Iliev, R., Dehghani, M., & Sagi, E. (2015). Automated text analysis in psychology: Methods, applications, and future developments. *Language and Cognition*, 7(2), 265–290.
- Kahneman, D., & Tversky, A. (1982). On the study of statistical intuitions. *Cognition*, 11(2), 123–141.
- Kamada, Y., & Kojima, F. (2014). Voter preferences, polarization, and electoral policies. *American Economic Journal: Microeconomics*, 6(4), 203–236.
- Keller, K. (1993). Conceptualizing, measuring, and managing customer-based brand equity. *The Journal of Marketing*.
- Kovach, M., & Tserenjigmid, G. (2022). Behavioral foundations of nested stochastic choice and nested logit. *Journal of Political Economy*, 130(9), 2411–2461.
- Kunda, Z. (1999). *Social cognition: Making sense of people*. MIT press.
- Laibson, D. (2001). A cue-theory of consumption. *The Quarterly Journal of Economics*, 116(1), 81–119.
- Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. *Science*, 350(6266), 1332–1338.
- Lauderdale, B. E., & Blumenau, J. (2025). Polarization over the priority of political problems. *American Journal of Political Science*.
- Le Mens, G., Kovács, B., Hannan, M. T., & Pros, G. (2023). Uncovering the semantics of concepts using gpt-4. *Proceedings of the National Academy of Sciences*, 120(49), e2309350120.
- Li, X., & Camerer, C. F. (2022). Predictable effects of visual salience in experimental decisions and games. *The Quarterly Journal of Economics*, 137(3), 1849–1900.

- Liang, A. (2025). Using machine learning to generate, clarify, and improve economic models. *arXiv preprint*.
- Loomes, G., & Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal*, 92(368), 805–824.
- Loureiro, M. L., & Lotade, J. (2005). Do fair trade and eco-labels in coffee wake up the consumer conscience? *Ecological Economics*, 53(1), 129–138.
- Lu, Y., Kaushal, N., Huang, X., & Gaddis, S. M. (2021). Priming covid-19 salience increases prejudice and discriminatory intent against asians and hispanics. *Proceedings of the National Academy of Sciences*, 118(36), e2105125118.
- Luettgau, L., Tempelmann, C., Kaiser, L. F., Jocham, G., et al. (2020). Decisions bias future choices by modifying hippocampal associative memories. *Nature Communications*, 11, 3318.
- Mahieux, E., de-Wit, L., Otten, L. J., Devlin, J. T., & Wicha, N. Y. (2024). The n400 effect captures nuances in implicit political preferences. *Scientific reports*, 14(1), 16730.
- Matějka, F., & McKay, A. (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*, 105(1), 272–98.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, 26.
- Minardi, S., Wang, F., & Gilboa, I. (2025). Consumption of values. *Management Science*, 71(3), 2623–2634.
- Mullainathan, S. (2002). A memory-based model of bounded rationality. *The Quarterly Journal of Economics*, 117(3), 735–774.
- Murphy, G. L. (2002). *The big book of concepts*. Cambridge: MIT Press.
- Noor, J. (2021). *Intuitive priors*, Mimeo.
- Pennesi, D. (2024). Identity, information, and situations. *Journal of the European Economic Association*, 22(2), 751–780.
- Peters, E., & Levin, I. P. (2008). Dissecting the risky-choice framing effect: Numeracy as an individual-difference factor in weighting risky and riskless options. *Judgment and Decision making*, 3(6), 435–448.
- Petrocik, J. R. (1996). Issue ownership in presidential elections, with a 1980 case study. *American Journal of Political Science*, 825–850.
- Salant, Y., & Rubinstein, A. (2008). (A, f): Choice with frames. *The Review of Economic Studies*, 75(4), 1287–1296.
- Santavirta, S., Wu, Y., Suominen, L., & Nummenmaa, L. (2025). Gpt-4v shows human-like social perceptual capabilities at phenomenological and neural levels. *Imaging Neuroscience*, 3, 1–19.
- Schofield, N. (2007). The mean voter theorem: Necessary and sufficient conditions for convergent equilibrium. *The Review of Economic Studies*, 74(3), 965–980.
- Schwartzstein, J., & Sunderam, A. (2021). Using models to persuade. *American Economic Review*, 111(1), 276–323.
- Scott, D. (1964). Measurement structures and linear inequalities. *Journal of mathematical psychology*, 1(2), 233–247.
- Shiller, R. J. (2017). Narrative economics. *American Economic Review*, 107(4), 967–1004.

- Spector, D. (2000). Rational debate and one-dimensional conflict. *The Quarterly Journal of Economics*, 115(1), 181–200.
- Spiegler, R. (2016). Bayesian networks and boundedly rational expectations. *The Quarterly Journal of Economics*, 131(3), 1243–1290.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211(4481), 453–458.
- Vaeth, M. (2023). *Rational voter learning, issue alignment, and polarization* (Working paper). Paris School of Economics.
- Walgrave, S., Lefevere, J., & Tresch, A. (2012). The associative dimension of issue ownership. *Public Opinion Quarterly*, 76, 771–782.
- Wimmer, G. E., & Shohamy, D. (2012). Preference by association: How memory mechanisms in the hippocampus bias decisions. *Science*, 338(6104), 270–273.
- Wolsko, C., Ariceaga, H., & Seiden, J. (2016). Red, white, and blue enough to be green: Effects of moral framing on climate change attitudes and conservation behaviors. *Journal of Experimental Social Psychology*, 65, 7–19.
- Zhou, B., Sun, Y., Bau, D., & Torralba, A. (2018). Interpretable basis decomposition for visual explanation. *European Conference on Computer Vision*.