A Criterion of Model Decisiveness*

Jeffrey Yang

April 8, 2023†

Abstract

When faced with decision-relevant information, decision-makers are often exposed to a multiplicity of different models, or accounts of how information should be interpreted. This paper proposes a theory of model selection — an account of what models decision-makers find compelling, and ultimately adopt — based on the insight that individuals seek *decisive* models that provide clear guidance regarding the best course of action. The decisiveness criterion is characterized by a demand for extreme models, which generates inferential biases such as overprecision and confirmation bias, but predicts meaningful bounds on the extent of these biases. The dependence of the decisiveness criterion on the decision-maker's objectives can produce documented patterns of preference reversals, rationalize seemingly contradictory patterns of inferential attribution errors, and generate novel predictions as to how belief polarization can arise along heterogeneity in decision-makers' objectives. I discuss applications of the theory to financial decision-making, the provision of expert advice, and social learning through the exchange of models.

*I am indebted to Benjamin Enke, Matthew Rabin, and Joshua Schwartzstein for their excellent supervision and guidance. I also thank Peter Andre, Chiara Aina, Thomas Graeber, David Laibson, Shengwu Li, Gautam Rao, Andrei Shleifer, Tomasz Strzalecki, Andi Wang and Harvard PhD workshop participants for helpful comments and suggestions.

†Department of Economics, Harvard University. jeffrey_yang@g.harvard.edu.

1 Introduction

When faced with uncertainty over the correct course of action, decision-makers are often confronted with an abundance of potentially decision-relevant information. It is increasingly understood that in order to make sense of this information, decision-makers often reason through the lens of a model – an account of how information should be interpreted. When an investor decides to sell a security because of a recent price drop, she acts based on a model that relates price patterns to future performance. When a manager rejects a candidate on the basis of a botched interview in spite of an impressive track record, she appeals to a model specifying what information is most diagnostic of candidate productivity. A voter who bases her political views on the reporting of certain news outlets while remaining skeptical toward others operates according to a model that specifies the credibility of different sources. In each of these cases, the decision-maker is guided by a framework, whether implicit or explicit, that specifies what information is relevant to the decision at hand, and what conclusions should be drawn from that information.

Decision-makers are often exposed to multiple candidate models: there are many plausible ways to value a firm, evaluate a job candidate, or interpret the news, and these models may be supplied by experts, encountered through social exchange, or developed through introspection. In contrast to the rational expectations view, under which decision-makers interpret data according to a well-calibrated model, decision-makers may not know the true model that should guide their inference, and so must decide which model to adopt for a given situation. In many cases, decision-makers are drawn to models that lead them astray: professional forecasters overreact to market signals, hiring managers put too much stock in noisy interview measures, and voters are misled by baseless claims and misinformation. When individuals seek to interpret information, what kinds of models do they find compelling, and ultimately use to guide their decision-making?

One potential answer stems from the basic insight that people dislike indecision, an observation which draws support from research in psychology. First, research on cognitive dissonance has demonstrated that individuals seek to avoid the psychological discomfort that arises from holding cognitions that are inconsistent with each other (Festinger, 1957), and modern accounts of dissonance theory emphasize the role between dissonance and indecision, wherein dissonance is aroused when "cognitions with action implications are in conflict with each other, making it difficult to act" (Harmon-Jones et al., 2015). A second line of research studies individuals' need for cognitive closure, conceptualized as a "desire to have a definite answer to a question, as opposed to uncertainty, confusion and ambiguity" (Kruglanski and Fishman, 2009). Under this account, individuals desire closure because it affords "a base for action", and research has studied how the need for closure affects how individuals process information and revise their beliefs.[1]

[1] More recently, Proulx and Inzlicht (2012) proposes a psychological framework of sense-making to unify these two accounts, in addition to a broad range of related accounts and evidence in psychology. Under this account, events that violate *meaning* — defined as a set of expected relationships that serve as a "guide for action" — leads to a physiological state of aversive arousal, which individuals seek to reduce.

If individuals are indeed averse to indecision, they may seek models that are *decisive*: that provide clear guidance regarding the best course of action. In this paper, I propose a formal notion of decisiveness, under which a model is decisive to the extent it reduces the decision-maker's residual uncertainty over the optimal course of action; I then analyze the inferences and choices of a decision-maker who is drawn to such models in interpreting information. In my formal framework, a decision-maker chooses from a menu of actions A, and observes the realization of data s that may be informative about the payoff-relevant state θ. To take an example in which the decision-maker is evaluating a job candidate, A may be the choice of whether to hire or reject the candidate, s is the information about the candidate observed by the decision-maker, such as their qualifications and interview performance, and θ is the underlying productivity of the candidate. In interpreting this data, the decision-maker does not have access to the true data-generating process but instead entertains a set of models: likelihood functions that map data to posterior beliefs over states. For example, the decision-maker may entertain models under which interview performance is highly diagnostic of productivity, in addition to models that instead emphasize the diagnostic value of the candidate's track record.

A model is decisive to the extent it recommends an action that, *under the beliefs induced by the model*, is close to the ex-post optimal course of action. That is, given $p^m(\theta|s)$, the posterior over states θ induced by model m for the observed data s, the decisiveness of m is

$$I(m|s) = -\left[\sum_{\theta} \max_{a' \in A} u(a', \theta) p^m(\theta|s) - \max_{a \in A} \sum_{\theta} u(a, \theta) p^m(\theta|s)\right]$$

where $u(a, \theta)$ gives the payoff of action a if the state is θ. Note that decisiveness corresponds to a payoff-metric measure of residual uncertainty: $-I(m|s)$ is equal to the decision-maker's willingness to pay to resolve remaining uncertainty about the state, under the beliefs induced by m. In other words, a model is decisive if a decision-maker operating under that model ascribes little value — if such an alternative was available to her — to resolving uncertainty prior to making her decision, instead of choosing the course of action recommended by the model.

Decisive models make strong recommendations, but they may also lead the decision-maker astray: a model suggesting that stock returns are highly predictable on the basis of past returns will be decisive, but is also far from the truth. Decisive models need not coincide with the true model governing the data-generating process, and to the extent the decision-maker entertains such models, the decisiveness criterion will distort their beliefs and decision-making. Importantly, in my framework the decision-maker does not know the true model — she does not willingly commit to a model she knows is wrong. Instead, my framework corresponds to a cognitive process in which the DM entertains a set of candidate models and tries each on "for size", assessing each model according to its implications for her decision and ultimately adopting the model that feels the most compelling. As such, this framework departs from certain theories of motivated beliefs in which the decision-maker seeks to maintain a set of beliefs in spite of the errors that may arise when she acts on them; rather, the decision-maker in my framework seeks decisive models precisely because she is motivated to identify the correct course of action.

This paper analyzes the systematic distortions in inference and choice produced by model selection under the decisiveness criterion. In particular, I show that the decisiveness criterion is characterized by a demand for models that render the decision-maker's environment more predictable. This causes the decision-maker to exhibit forms of both overprecision and confirmation bias: when initially uncertain over the optimal course of action, the decision-maker favors models that overstate the diagnostic value of the data, and when sufficiently predisposed towards a course of action, the decision-maker favors models that minimize the diagnostic value of disconfirmatory data. I also show how the dependence of the decisiveness criterion on the decision-maker's objectives can produce documented patterns of preference reversals, rationalize seemingly contradictory patterns in social attribution, and generate novel predictions as to how belief polarization can arise along heterogeneity in decision-makers' objectives. Finally, I demonstrate how choice under the decisiveness criterion exhibits an aversion to various forms of hedging against uncertainty.

To illustrate the basic implications of the decisiveness criterion, consider a simple example. Suppose that a manager is deciding whether or not to make an offer to a potential internal hire to join her team. The candidate is either high-productivity (θ_h) or low-productivity (θ_l), and the manager believes either possibility is equally likely ex-ante. The manager's payoffs are as follows:

	θ_h	θ_l
hire	v	$-k$
reject	0	0

That is, the manager wishes to hire the candidate only if they are high-productivity.

The manager has access to two components of information on the candidate: the candidate's track record in their previous role, s^R, and the candidate's performance in a newly developed interview assessment tailored to the current role s^I. Suppose that performance in either component $c \in \{R, I\}$ can either be high ($s^c = 1$) or low ($s^c = 0$), and that the manager is uncertain over the diagnostic value of each piece of information. In particular, suppose the manager entertains three models: m_R, under which only the candidate's track record has diagnostic value, m_I, under which only the interview assessment has diagnostic value, and m_{RI}, under which both components hold equal diagnostic value. In particular, denoting $L_m(s^R, s^I) = \frac{m(s^R, s^I|\theta_h)}{m(s^R, s^I|\theta_l)}$ as the likelihood ratio associated with model m, we have

$$L_{m_R}(s^R, s^I) = \begin{cases} 4 & s^R = 1 \\ 1/4 & s^R = 0 \end{cases}$$

$$L_{m_I}(s^R, s^I) = \begin{cases} 4 & s^I = 1 \\ 1/4 & s^I = 0 \end{cases}$$

$$L_{m_{RI}}(s^R, s^I) = \begin{cases} 4 & s^I = s^R = 1 \\ 1 & s^I \neq s^R \\ 1/4 & s^I = s^R = 0 \end{cases}$$

Suppose that the candidate's track record demonstrates high performance in their previous role, but that they perform poorly in the interview assessment: $(s^R, s^I) = (1, 0)$. Which model does the manager adopt under the decisiveness criterion? Below, we consider how the manager acts under two different hiring regimes, each of which dictate the relative costs and benefits of hiring.

Growth regime: $v = 4$, $k = 1$. In this regime, the benefits of hiring a high-productivity candidate are greater than the costs of hiring a low-productivity candidate; the manager is therefore predisposed to hiring the candidate. Here, m_R is the most decisive model. The intuition is as follows: under m_R the data suggests that the candidate is likely high productivity, providing the manager with yet greater justification for hiring the candidate. On the other hand, under m_{RI} the data are inconclusive about candidate productivity, and so this model provides weaker justification for hiring compared to m_R. Under m_I, the data are bad news about the productivity of the candidate, an interpretation which, given the manager's decision problem, would result in her being maximally uncertain over whether to hire or reject the candidate. As such, m_R provides the most decisive recommendation of the three models.

Downsizing regime: $v = 1$, $k = 4$. In this regime, hiring costs are high, and so the manager is predisposed to rejecting the candidate. Here, m_I is the most decisive model. The intuition is analogous to the case above: while both m_I and m_{RI} recommend that the manager reject the candidate, m_I provides a stronger justification toward this course of action, whereas under m_R, the manager is maximally uncertain over whether to hire or reject.

This example illustrates two key properties of the decisiveness criterion. First, the criterion tends to privilege extreme models — formally, models that cannot be expressed as a mixture of other models the DM entertains. Note that in either case, the model m_{RI}, which ascribes some diagnostic value to both the candidate's track record and interview performance, and can be expressed as the mixture of m_R and m_I, is never selected. The intuition is as follows: if the manager prefers *reject* under m_{RI}, m_I provides yet stronger justification towards *reject*, whereas if the manager prefers *hire* under m_{RI}, m_R provides yet stronger justification towards that course of action; the manager must find m_{RI} less decisive than one of the extreme models, which will provide greater certainty regarding her optimal hiring decision. Second, the decisiveness criterion depends crucially on the decision-maker's objectives, as the difference between the two regimes demonstrate: in each regime, the model that supports the action that the manager is predisposed towards choosing is more decisive than the model that provides discomfirmatory evidence.

The remainder of the paper is organized as follows: Section 2 develops the general framework and introduces the decisiveness criterion. Section 3 characterizes the implications of the decisiveness criterion for model selection, holding fixed the action space and payoffs. I show how model selection under the decisiveness criterion is principally characterized by two key conditions, which formalize the notion in which the criterion privileges models that render the decision-maker's environment predictable: 1) the decision-maker has a tendency to adopt extreme models — those that cannot be expressed as a mixture of other models she entertains, and 2) the decision-maker has a preference for models that induce high certainty in a single state. I apply these results to show how the decisiveness criterion generates forms of two documented patterns of biased inference: overprecision and confirmation bias. In particular, the criterion predicts that if the decision-maker has greater initial uncertainty with respect to her available courses of action, she will tend to exhibit overprecision, seeking models that overstate the informational content of her signals. However, as the decision-maker's prior sufficiently favors one course of action, she will no longer uniformly exhibit overprecision, but instead seek models that explain away disconfirmatory information, producing a form of confirmation bias. Finally, I illustrate the relationship between my model selection criterion and related criteria, such as the Blackwell ordering on experiments and notions of model parsimony.

Section 4 studies how model selection under the decisiveness criterion varies with the payoffs and objectives of the decision-maker. First, I demonstrate that the decisiveness criterion generates a "sour grapes" effect: the addition of an unchosen action a will lead the decision-maker towards models that ascribe a low value to choosing a, as well as similar actions. I show how this sour grapes effect predicts context effects documented in Tversky and Shafir (1992), who show how the addition of an action to the choice set can reduce subjects' propensity to choose similar, competing actions, and instead induce them to choose dissimilar actions. Next, I show how under the decisiveness criterion, the attractiveness of a model is increasing in the relative attractiveness of the actions that the model recommends. I demonstrate how this property can account for seemingly contradictory patterns in social attribution in which individuals neglect the confounding influence of situational factors when inferring dispositional traits of others from their behavior, yet make the opposite error when interpreting poor behavior from ingroup members (Vonk and Konst, 1998). In another application, I show how the same force can generate belief polarization resulting from heterogeneity in decision-makers' objectives and preferences. Finally, I study how model selection varies with the payoff-relevance of states. In particular, I show that when a state has the interpretation of a *nuisance variable* — that is, when models that attribute the signal to that state are less informative about other states compared to models that neglect that state — the DM will tend to adopt models that neglect that state when it is not payoff-relevant. I demonstrate how this property can account for the observed sensitivity of various attribution errors in inference problems to decision-maker's inferential goals, as documented by Krull (1993) and Graeber (2022).

In Section 5, I discuss implications for choice. I show that choice under the decisiveness criterion is characterized by an aversion to hedging or diversification. In addition, I study comparative statics of choice under the decisiveness criterion, focusing on a notion of relative aversion to *C-diversified* actions — that is, actions whose payoffs fall short of the maximal payoff that can be achieved in each state by a constant. I show that the decision-maker will be more diversification-averse if she entertains a larger set of models, and that conversely, a greater level diversification-aversion reveals that the decision-maker entertains a larger set of models. In Appendix A.6, I provide a behavioral characterization of the model and discuss its identification properties, building directly on results in Stoye (2011), which studies a closely related model of min-max regret.

Section 6 discusses additional applications of the decisiveness criterion. Whereas the basic framework takes the set of models the decision-maker entertains as primitive, each of these applications considers a process that may give rise to this set. In the first application, I analyze a setting in which expert advisors supply models to the decision-maker, and apply the decisiveness criterion to shed light on why individuals are drawn to certain advice. I demonstrate that this force, in conjunction with competition between advisors, induces advisors to inflate the certainty of their recommendations. In the second application, I analyze a setting in which social learning occurs through the exchange of models, following the framework introduced in Schwartzstein and Sunderam (2022), and demonstrate that the decisiveness criterion predicts group polarization arising from such a process.

Section 7 discusses extensions to the basic framework. In the first extension, I address a shortcoming of the decisiveness criterion — that when the decision-maker entertains the full space of models, she will always come to adopt models that resolve all residual uncertainty, regardless of the plausibility of those models. I discuss two methods of constraining the space of models the DM entertains, one which assumes a cost of adopting far-fetched beliefs, the other which stipulates an entry condition on the models the DM can entertain based on the notion of model fit developed in Schwartzstein and Sunderam (2021). In the second extension, I outline an *ex-ante* version of the model in which the decision-maker adopts a model prior to the signal realization and evaluates models based on their expected decisiveness, averaged over signal realizations. I show that analogs of certain properties — such as the preference for extreme models, overprecision, and comparative statics involving the relative attractiveness of actions and the decision-relevance of states — continue to hold the ex-ante case, whereas others, such as confirmation bias and the sour grapes effect, do not. All proofs are collected in the Appendix.

Related Literature

This paper directly contributes to an active literature studying the implications of model selection, which has focused on several distinct model selection criteria. One criterion is model fit, corresponding to the notion that decision-makers are drawn to models that explain the data well. In particular, the formal framework employed in this paper draws directly from Schwartzstein and Sunderam (2021), who analyze "model persuasion" — in which a persuader influences the model the receiver uses to interpret the realized data — and assume that receivers employ a model selection criteria based on fit. Aina (2022) builds on this framework, analyzing a setting in which the persuader commits to a set of models prior to the realization of the data, whereas Schwartzstein and Sunderam (2022) adapt the framework to study social learning based on the exchange of models. In contrast, this paper focuses primarily on non-strategic contexts, in which the set of models the decision-maker entertains is primitive; however, Section 6 discusses applications of the decisiveness criterion to strategic and social learning contexts. Operating under different formal frameworks, Izzo et al. (2021) and Hong et al. (2007) study the implications of fit-based model selection, and focus on political persuasion and a non-strategic financial decision-making context, respectively. Another criterion that has received attention in the literature is optimism, corresponding to the notion that decision-makers are drawn to "hopeful narratives". Eliaz and Spiegler (2020) formalize narratives as causal models (directed acyclic graphs) to study political persuasion, and posits a model selection criteria based on optimism, whereas Caplin and Leahy (2019) studies the implications of model selection based on optimism for belief polarization, trading decisions, and the formation of asset bubbles.

The decisiveness criterion provides an account of what makes a model compelling that is complementary to these existing approaches, and in particular delivers predictions distinct from those of selection criteria based on fit or optimism. The fit criterion, for one, only concerns how well models can explain past data, and is silent on how individuals may be drawn to models due to their implications for future action. For example, one might imagine that employers continue to place high stock in unstructured interviews (Dana et al., 2013) not necessarily because such interviews have demonstrated explanatory power over the productivity of past hires, but because taken at face value, they provide strong recommendations for hiring decisions. Similarly, investors may find technical analysis appealing not just because it provides an explanation for realized price trends, but also because it often provides a clear recommendation on whether to buy or sell. Importantly, because a fit-based model selection criterion is backward-looking, without further structure and assumptions, such a criterion cannot account for the possibility that the model a decision-maker adopts can be shaped by their objectives. On the other hand, while an optimism-based criterion takes seriously the idea that a model's forward-looking implications may matter for model selection, it is clear that decision-makers also regularly adopt models that do not induce optimistic beliefs. For example, employers may overstate the diagnostic value of an interview even if it produces bad news about the productivity of the candidate, and voters are swayed by political spin often solely focused on casting the opposition in a negative light. Because a decisive model need not lead to optimistic beliefs, as illustrated in the motivating example, the decisiveness criterion can rationalize the adoption of such models.

Though the decisiveness criterion aims to capture a psychological notion distinct from regret — namely, that individuals are drawn to clear recommendations that resolve indecision — the criterion is formally equivalent to a formulation of expected regret, and therefore relates to a large literature studying the implications of regret aversion in decision-making (e.g. Bell, 1982; Loomes and Sugden, 1982; Hayashi, 2008; Sarver, 2008; Stoye, 2011). This literature studies how regret aversion may affect choices over actions or menus of actions, corresponding to the psychology that the desire to avoid regret may affect choices prior to the realization of uncertainty. This paper, on the other hand, focuses on the implications of a regret-based criterion for model selection and its subsequent implications for decision-making, motivated by the notion that individuals find models compelling to the extent that they guide future action. Formally, this theory has a tight connection to a model of multiple-priors min-max regret studied in Hayashi (2008) and subsequently Stoye (2011), in which the decision-maker chooses the action to maximize expected utility (minimize expected regret) under the regret-maximizing prior belief. My theory corresponds to a model in which the decision-maker instead chooses to maximize expected utility under the regret-minimizing prior belief. While the two models produce contrasting predictions for both model selection and choice — for instance, min-max regret would predict that the DM always selects m_{RI} in the hiring example analyzed above, whereas the decisiveness criterion predicts that the DM never selects m_{RI} — the axiomatic characterization of my model builds directly on results from Stoye (2011).

One paper that takes a related conceptual approach is Eyster et al. (2021), which develops a model of ex-post rationalization in which a decision-maker adopts preferences to minimize an *ex-post* notion of regret with respect to past choices, motivated by evidence for the sunk-cost fallacy. This approach is both formally distinct from and complementary to the approach taken in this paper, which develops a theory in which the decision-maker adopts models that minimize *ex-ante* regret. In the hiring example analyzed above, for instance, model selection under ex-post rationalization depends crucially on each model's implications for the optimality of actions the decision-maker has already taken, such as previous hiring decisions. On the other hand, model selection under the decisiveness criterion depends only on each model's implications for the decision problem the decision-maker currently faces.

2 Framework and Decisiveness Criterion

2.1 General Setup

Consider a decision-maker (the DM) with priors ρ over a finite set of payoff relevant states Θ, and updates her beliefs about the payoff relevant state given a signal $s \in S$. The DM entertains a set of models M, where each model $m \in M$ is a mapping from payoff relevant states to a probability distribution over signals: $m : \Theta \to \Delta(S)$. Let $\mathcal{M}$ denote the set of all models. Let m_T denote the true model that governs the data-generating process.

The DM faces a menu of actions A and the payoff function $u : A \times \Theta \to \mathbb{R}$; call $\mathcal{D} = (A, u)$ the DM's decision problem. Assume that the decision problem is well defined, in that for any beliefs there will be a set of expected-utility maximizing actions. Also assume that the set of models the DM entertains is closed.

Assumption 1. For all $p \in \Delta(\Theta)$, $\arg\max_{a \in A} \sum_\theta u(a, \theta)p(\theta)$ is nonempty.

Assumption 2. M is closed.

The timing is as follows: nature first draws the state θ and the signal s is generated according to m_T and is observed by the DM. The DM then adopts a model $m \in M$, and forms the posteriors $p^m(\cdot|s)$ according to Bayes rule, where $p^m(\theta|s) = m(s|\theta)\rho(\theta)/(\sum_{\theta'} m(s|\theta')\rho(\theta'))$. The DM then takes the action that maximizes the model-induced posterior expected utility; let

$$A^m_{\mathcal{D}}(s) \equiv \arg\max_{a \in A} \sum_{\theta} u(a,\theta)p^m(\theta|s)$$

denote the set of actions recommended by m, and say that m recommends a from $\mathcal{D}$ if $a \in A^m_{\mathcal{D}}(s)$. After the action is taken, the DM's payoffs are then realized according to u and the realized state θ.

We now specify the criterion that governs the DM's model selection. For a posterior belief p and decision problem $\mathcal{D}$, let

$$R_{\mathcal{D}}(p) = \sum_{\theta} \max_{a'} u(a',\theta)p(\theta) - \max_{a} \sum_{\theta} u(a,\theta)p(\theta) \tag{1}$$

denote the *residual uncertainty* associated with p: it is the average difference in utility between the ex-post optimal action in each state and the ex-ante optimal action, which is equal to the DM's willingness-to-pay to uncertainty prior to choosing an action given beliefs p.[2] Define the *decisiveness* of a model m given realized signal s as follows:

$$I_{\mathcal{D}}(m|s) = -R_{\mathcal{D}}(p^m(\cdot|s)) \tag{2}$$

The DM adopts a model that maximizes decisiveness, satisfying $m \in \arg\max_{m \in M} I_{\mathcal{D}}(m|s)$. Decisive models make strong prescriptions about the optimal course of action, where here a prescription is strong in the sense that it recommends an action that is likely close to the ex-post optimal action. That is, a model is decisive to the extent the DM ascribes little value — if such an alternative was available — to resolving uncertainty prior to making her decision, instead of choosing the course of action recommended by the model.

The decisiveness criterion captures a notion of what it means for a model to guide action. Two related notions of guidance that are not captured by this criterion are model consistency and parsimony. Individuals may seek models that provide *consistent* guidance – that is, models that recommend the same action under many realizations of the data. For example, consider a model of stock returns under which the price path follows a random walk with positive drift; for any history of returns, the model's recommendation is the same: the DM should buy and hold. Such a model produces a consistent recommendation, in contrast to models that recommend various strategies to time the market depending on past returns. As an application in Section 3.3 explores, the decisiveness criterion will tend to privilege models of the latter kind over the former: under this criterion, it is not the consistency of guidance that matters, but rather its strength as measured by residual decision uncertainty. Individuals may also be drawn to *parsimony* – that is, they may see a model as providing

[2] $R_{\mathcal{D}}$ is formally equivalent to a formulation of the expected regret under beliefs p.

better guidance to the extent it provides a simplified way of translating data into action, for instance, by directing focus to certain features of the data while ignoring others. In Section 3.2, I show that the decisiveness criterion does not universally privilege such simplified models, but rather generates predictions as to when such models will be selected by the DM.

Note that in this framework, model selection occurs ex-post: after the signal is realized, the DM evaluates each model based on its decisiveness, given the signal realization, and adopts the most decisive model. One might instead imagine an account in which the DM evaluates each model according to an *ex-ante* notion of decisiveness, and adopts a model prior to the signal realization. The ex-post notion may be more appropriate for analyzing situations in which the DM faces a decision problem with an idiosyncratic information structure, as in the motivating example, in which the manager seeks to make a one-off hire on the basis of information specific to that particular candidate. Furthermore, because the ex-post decisiveness of a model can be evaluated solely on the basis of its recommendation given the observed signal realization, whereas ex-ante decisiveness requires the DM to consider what the model would recommend for all possible signal realizations, ex-post decisiveness may be more reasonable criterion for studying settings in which the set of possible signal realizations may be large, unknown to the DM, or otherwise hard for the DM to imagine. The ex-ante perspective, on the other hand, may be more appropriate for situations in which the DM faces a sequence of decision problems that share a common information structure – for example, if the hiring manager instead sought to evaluate a collection of candidates on the basis of their performance in a standardized test. While this paper focuses primarily on the ex-post notion of decisiveness, in Section 7.2 I outline a formulation of ex-ante decisiveness, and compare the implications of the ex-ante and ex-post specifications of decisiveness for model selection.

Finally, the fact that the decisiveness criterion is defined with respect to a decision problem raises the question of how the theory can be applied to situations where the DM does not face a particular decision problem but nevertheless has the opportunity to learn from data. In such situations, one natural approach is to assume that the DM learns from data as if they face a prediction problem — that is, as if they are tasked with reporting a set of beliefs from $A = \Delta(\Theta)$ and paid according to a scoring function $u : \Delta(\Theta) \times \Theta \to \mathbb{R}$. In Appendix A.1, I characterize model selection under the class of prediction problems with incentive-compatible scoring rules.

2.2 Examples

The following set of examples illustrate the basic mechanics of the decisiveness criterion. In each of these examples, we take the realized signal as given and work directly with model-implied posteriors $p^m(\cdot|s)$.

Example (Hiring Decision). Consider the hiring decision from the introduction. Here, the payoff relevant states are the productivity of the candidate, $\Theta = \{\theta_l, \theta_h\}$, and the DM's decision problem is given by the actions $A = \{\mathit{hire}, \mathit{reject}\}$ and the utility function

$$u(a, \theta) = \begin{cases} v & a = \mathit{hire},\ \theta = \theta_h \\ -k & a = \mathit{hire},\ \theta = \theta_l \\ 0 & a = \mathit{reject} \end{cases}$$

Fix a model m, and let $p_h^m = p^m(\theta_h|s)$ denote the model-implied posterior belief in θ_h. For $p_h^m \geq \frac{k}{v+k}$ the DM chooses *hire*, and chooses *reject* otherwise. The residual uncertainty associated with m is then

$$R_{\mathcal{D}}(p^m(\cdot|s)) = \begin{cases} k(1 - p_h^m) & p_h^m \geq \frac{k}{v+k} \\ vp_h^m & \text{otherwise} \end{cases}$$

That is, models that induce beliefs toward either extreme tend to be more decisive, and the range for which decisiveness is increasing in p_h^m is increasing in the value of hiring a high-productivity worker v, and decreasing in the cost of hiring a low-productivity worker k.

Example (Prediction Market). The DM participates in a prediction market based on a binary state $\Theta = \{\theta_A, \theta_B\}$. There are two assets, A and B, which pay 1 if the corresponding state is realized and 0 otherwise. The assets are priced at q_A and q_B, respectively, with $q_A + q_B = 1$.

The DM is endowed with wealth w and chooses how much of each asset to purchase, (x_A, x_B). Without loss of generality we can restrict attention to choices $x_A, x_B \geq 0$, with $x_A x_B = 0$, and so the DM's choice can be summarized by $x = x_A - x_B$. Therefore, the DM's wealth will be $w + x(1 - q_A)$ if θ_A is realized and $w - xq_A$ if θ_B is realized. Assume the DM is risk averse with log utility. The DM's decision problem can then be described by the action space $A = \mathbb{R}$ and utility

$$u(a, \theta) = \begin{cases} \ln(w + x(1 - q_A)) & \theta = \theta_A \\ \ln(w - xq_A) & \theta = \theta_B \end{cases}$$

Fix a model m, and denote $p_A^m = p^m(\theta_A|s)$ as the model-induced posterior belief in θ_A. Given this model, the DM chooses x to maximize expected utility $U = p_A^m \ln(w + x(1 - q_A)) + (1 - p_A^m) \ln(w - xq_A)$. First-order conditions imply the the optimum $x = \frac{(p_A^m - q_A)}{(1-q_A)q_A} w$.

Note that under state θ_A, the DM's ex-post optimal choice is to invest all of her wealth into asset A, which yields a payoff of $\ln(w_i/q_A)$, and under state θ_B, the DM's ex-post optimal choice is to invest all of her wealth into asset B, which yields a payoff of $\ln(w_i/(1 - q_A))$. The residual uncertainty associated with m is then

$$R_{\mathcal{D}}(p^m(\cdot|s)) = -\ln(1 - p_A^m)(1 - p_A^m) - \ln(p_A^m)p_A^m$$

That is, the decisiveness criterion predicts that for this decision problem, the agent will select the model that minimizes posterior entropy, given the data.

2.3 Additional Maintained Assumptions

Throughout this paper we will maintain two additional assumptions:

Assumption 3. ρ has full support over Θ.

Assumption 4. S has at least two elements.

Assumption 3 is without loss, as we can restrict attention to states Θ' on which the DM's prior places positive probability. Assumption 4 amounts to a non-triviality assumption that guarantees that learning can occur. Assumptions 3 and 4 guarantee that fixing a signal realization $s \in S$, for every $p \in \Delta(\Theta)$ there exists a model m (not necessarily unique) such that implements beliefs p, with $p^m(\cdot|s) = p$.

3 Properties of Selection Under Decisiveness

In this section, I show that model selection under the decisiveness criterion is principally characterized by a a preference for extreme models — those that cannot be expressed as a mixture of other models she entertains. In a series of applications, I discuss how this extremeness property can predict both existing and novel forms of overprecision and confirmation bias. I then discuss the relationship between the decisiveness criterion and related model selection criteria.

3.1 Characterization of Model Selection Rule

For a decision problem $\mathcal{D}$ and signal s, let $C_{\mathcal{D}}(M|s) \equiv \arg\max_{m \in M} I_{\mathcal{D}}(m|s)$ denote the model choice correspondence under the decisiveness criterion. Because the regret function is concave in beliefs for any decision problem, $C_{\mathcal{D}}$ satisfies the following extremeness property:

Proposition 1. For $m, m' \in M$, if $m, m' \notin C_{\mathcal{D}}(M|s)$, then for any $\lambda \in (0,1)$, $\lambda m + (1 - \lambda)m' \notin C_{\mathcal{D}}(M|s)$. Furthermore, if $A^m_{\mathcal{D}}(s)$ and $A^{m'}_{\mathcal{D}}(s)$ are disjoint, then $\lambda m + (1 - \lambda)m' \notin C_{\mathcal{D}}(M|s)$.

In words, under the decisiveness criterion, the DM tends to adopt extreme models — models that cannot be expressed as a mixture of other models in M — because the decisiveness of a composite model obtained by averaging two models cannot exceed the decisiveness of both of those models. Furthermore, if those two models recommend different actions, then the composite model delivers strictly lower decisiveness.

An immediate consequence of Proposition 1 is that the decisiveness criterion "biases" the DM against model averaging: if the true model m_T is an average over models in M, Proposition 1 implies that m_T will never deliver the strictly highest decisiveness among the models $M \cup \{m_T\}$, and that so long as at least two models in M recommend different actions, the DM will never select the true model.

For intuition as to why the DM finds such composite models unattractive, consider the following example: the DM is trying to predict whether it will rain in an hour, and entertains two models. m_1 says that current humidity is the key predictor of rain, and m_2 says that cloud cover is the key predictor. Now consider a composite model $\overline{m} = 1/2m_1 + 1/2m_2$ that says that some mixture of humidity and cloud cover predict rain — the composite model can be interpreted as saying that on any given day, humidity is the key predictor of rain with 50% probability, and otherwise cloud cover is the key predictor of rain. Both m_1 and m_2 provide a specific lens through which to interpret the data, whereas $\overline{m}$ leaves the DM uncertain over which of these two interpretations to use. As such, the DM will find that one of m_1 or m_2 provides more compelling guidance than the composite model $\overline{m}$.

Proposition 1 also demonstrates a sense in which the decisiveness criterion generates "coarse thinking" — the tendency for individuals to represent the full space of possibilities as a coarse set of "categories" (Mullainathan, 2002; Mullainathan et al., 2008), and provides predictions over what the categories individuals use. To illustrate this, following the framework in Mullainathan (2002), suppose that the true information structure is governed by a set of underlying models $\tilde{M}$ (the underlying types in Mullainathan (2002))[3], and suppose that the DM entertains the set of models $M = Co(\tilde{M})$, where $Co(\tilde{M})$ denotes the convex hull of $\tilde{M}$ — this choice of M ensures that model selection *per se* does not generate coarse thinking, since the DM is free to interpret the data according to any mixture over the underlying models. Proposition 1, however, implies that the DM will only ever interpret the data according to a subset of M; in particular, the DM will only ever select from the extreme models in $\tilde{M}$. While this subset of M is taken to be a primitive in Mullainathan (2002) — the set of "categories", in their terminology — the decisiveness criterion endogenizes the set of categories the DM considers as a function of the underlying models $\tilde{M}$. Note the key feature of this set of categories is that it contains only the "extremes": in inferring effort from output, a worker is categorized as either low or high-productivity, rather than of moderate productivity; in inferring skill from returns, a fund manager's performance is categorized as the product of either luck or skill, rather than some mix of the two. Just as Mullainathan (2002) motivates categorical thinking as a heuristic that streamlines decision-making, the categories that the DM entertains under the decisiveness criterion are those that provide strong decision-making guidance.

Finally, note that a model m can be expressed as a mixture of m' and m'' if and only if the model induced belief p^m can be expressed as a mixture of $p^{m'}$ and $p^{m''}$. This implies that the preference for extreme models established in Proposition 1 results in the DM adopting similarly extreme beliefs.

We now turn to characterizing $C_{\mathcal{D}}$. The extremeness property from Proposition 1 is a key axiom in this characterization, along with an axiom that states that any model that

[3] To formalize this setting within my framework, consider a setting where nature draws the model m from $\tilde{M}$ according to $\pi \in \Delta(\tilde{M})$, where π is assumed to have full support. In this case, given signal s, a Bayesian would use the model $m_T = \sum_{m \in \tilde{M}} m\pi(m|s)$, where $\pi(m|s) = \sum_\theta m(s|\theta)p(\theta)\pi(m) / \left(\sum_{m \in \tilde{M}} \sum_\theta m(s|\theta)p(\theta)\pi(m)\right)$ is the Bayesian posterior belief in model m given the signal s.

induces certainty over the state must always be weakly preferred to other models.

Proposition 2. A model choice correspondence C satisfies

1. Sen's α, β: If $m \in M \subseteq M'$ and $m \in C(M')$, then $m \in C(M)$. Also, if $m, m' \in C(M)$, $M \subseteq M'$ and $m' \in C(M')$ then $m \in C(M')$.

2. Continuity: For all $m \in \mathcal{M}$, $\{m' \in \mathcal{M} : m' \in C(\{m, m'\})\}$ and $\{m' \in \mathcal{M} : m \in C(\{m, m'\})\}$ are closed.

3. Scale Invariance: For any $m \in \mathcal{M}$, if m' satisfies $m'(s|\theta) = \lambda m(s|\theta) \; \forall \; \theta, \lambda > 0$, then $C(\{m, m'\}) = \{m, m'\}$.

4. Extremeness: For $m, m' \in M$, if $m, m' \notin C(M)$, then for any $\lambda \in (0, 1)$, $\lambda m + (1 - \lambda)m' \notin C(M)$.

5. Certainty Preference: If $m(s|\theta) = 1$ for any $\theta \in \Theta$ and $m \in M$, then $m \in C(M)$.

if and only if there exists a decision problem $\mathcal{D}$ and signal s such that $C(M) = C_{\mathcal{D}}(M|s)$ for all $M \subseteq \mathcal{M}$.

Sen's α, β reflects the fact that the decisiveness criterion induces a total order over $\mathcal{M}$ for a given decision problem, while Continuity is a technical condition. Scale Invariance ensures that models that induce the same posterior over outcomes are equivalent under the model selection criteria. Scale Invariance has the following interpretation: if we think of $m(s|\theta)$ as a measure of the strength with which θ explains the realized signal under m, the axiom implies that that the attractiveness of a model depends only on relative comparisons between these measures of explanatory strength across states θ.

See the discussion of Proposition 1 for an interpretation of Extremeness. Certainty Preference says that any model that induces certainty in an outcome must always be adopted, as such a model eliminates residual uncertainty. Together, Extremeness and Certainty Preference formalize the notion in which the decisiveness criterion privileges models that render the decision-maker's environment predictable. In what follows, I discuss a series of applications illustrating how these properties of the decisiveness criterion can account for documented patterns of belief updating, as well as produce new predictions.

3.2 Applications

3.2.1 Overprecision

A large experimental literature has documented that individuals exhibit overprecision — the tendency to be excessively confident in the accuracy of one's beliefs (Moore et al., 2015). In a common experimental paradigm, subjects are asked to predict a binary outcome, and state their belief that their prediction is correct. Across a variety of domains, subjects' estimates of their own accuracy exceed the true hit rates. Similarly, studies eliciting subjects' subjective confidence intervals around their estimates of continuous outcomes find that these confidence intervals tend to be too narrow.

A natural consequence of Proposition 1 is that the DM will exhibit overprecision in the following sense — the DM will tend to overstate the informational content of her information. Formally, fix any decision problem $\mathcal{D}$ and the realization of the signal s. Let m_T denote the true model that describes the DM's actual experiment, and let $m_\varnothing$ denote an uninformative model for which $m(s|\theta)$ is constant in θ. Proposition 1 implies the following result:

Corollary 1. Consider a model m satisfying $\lambda m + (1-\lambda)m_\varnothing = m_T$ for $\lambda \in (0,1)$. If $I_\mathcal{D}(m_T|s) > I_\mathcal{D}(m_\varnothing|s)$, then $I_\mathcal{D}(m|s) > I_\mathcal{D}(m_T|s)$.

Here, m can be interpreted as a model that overstates the informational content of the true experiment m_T, which is a mixture of m and noise. Corollary 1 states that to the extent the DM entertains such models that overstate the informativeness of the their experiment, those models will be favored over the true model, so long as the DM finds the true model more decisive than her prior.

Example (Overprecision in Hiring). Suppose a hiring manager wishes to learn whether a candidate is high-productivity (θ_h) or low productivity (θ_l), and holds a prior belief ρ. The employer observes a signal s of productivity and chooses from a set of actions (e.g. hiring/rejecting the candidate, bringing the candidate to a second interview, etc). Suppose that $I_\mathcal{D}(m_T|s) > I_\mathcal{D}(m_\varnothing|s)$ — that is, the manager's signal reduced her residual uncertainty regarding which action to take. Suppose the manager also entertains a model m that overstates the decisiveness of her signal in the sense described earlier: m satisfies $\lambda m + (1-\lambda)m_\varnothing = m_T$ for $\lambda \in (0,1)$; corollary 1 implies that the hiring manager will adopt m over the true model m_T. Given the binary state setting, the implications for the manager's posterior beliefs are straightforward: if under the true model, s is good news about productivity, with $\rho(\theta_h) < p^{m_T}(\theta_h|s)$, then m induces yet a higher belief in productivity, with $p^m(\theta_h|s) > p^{m_T}(\theta_h|s)$; similarly if under the true model s is bad news about productivity, then m induces yet a lower belief in productivity. In other words, the manager overstates her ability to discern high-productivity candidates from low-productivity candidates.

To relate this example to the two-alternative forced-choice paradigm used in experimental tests for overprecision, suppose that the manager's decision problem consists of a binary prediction j regarding the productivity of the candidate, as well as an estimate q of the probability that the prediction will be correct. Formally, we have $A = \{h, l\} \times [0,1]$, with $u((j,q),\theta) = u_1(j,\theta) + u_2((j,q),\theta)$, where

$$u_1(j,\theta) = \begin{cases} 1 & \theta_j = \theta \\ 0 & \text{otherwise} \end{cases}$$

$$u_2((j,q),\theta) = \begin{cases} -(1-q)^2 & \theta_j = \theta \\ -q^2 & \text{otherwise} \end{cases}$$

That is, given belief p, the manager predicts the type she believes is more likely, and reports the probability that her prediction will be correct.

Now suppose that $\rho(\theta_h) = 0.5$, and that as before, the DM entertains the true model m_T, as well as the model m as defined above. Here, we have $I_\mathcal{D}(m_T|s) \geq I_\mathcal{D}(m_\varnothing|s)$ for all s, which implies that $I_\mathcal{D}(m|s) \geq I_\mathcal{D}(m_\varnothing|s)$. As before, the manager overreacts to her signal about productivity.

In particular, assume that some signal s is good news about productivity — the manager predicts productivity to be $j = h$ after observing s. Under the true model, the probability that manager's prediction is correct is given by $p^{m_T}(\theta_h|s)$, but since the manager finds m more decisive than m_T, the manager instead reports a confidence level of $q = p^m(\theta_h|s) > p^{m_T}(\theta_h|s)$. Analogously, when s is bad news about productivity, the manager reports an inflated confidence level in her prediction that $j = l$ of $q = p^m(\theta_l|s) > p^{m_T}(\theta_l|s)$. Therefore, when faced with with a population of candidates, the manager is systematically overconfident in her predictions about productivity. ▲

Note that a key condition for overprecision is that $I_\mathcal{D}(m_T|s) > I_\mathcal{D}(m_\varnothing|s)$: under the true model, the realized signal s must (weakly) reduce the DM's residual uncertainty over the optimal course of action. In other words, the DM is more likely to exhibit overprecision in situations where she is initially uncertain over the optimal course of action. For instance, if we suppose that the hiring manager in the previous example has binary actions $\{$*hire*, *don't hire*$\}$, the manager will tend exhibit overprecision when her prior beliefs leave her close to indifferent between hiring and not hiring.

In contrast, consider the case where $I_\mathcal{D}(m_T|s) < I_\mathcal{D}(m_\varnothing|s)$ the signal increases the manager's residual uncertainty over the optimal course of action. For instance, it may be the case that the manager's prior beliefs strongly recommend that she hire the candidate, but she receives a signal to the contrary that would push her to greater uncertainty over whether to hire, under m_T. In this case, the decisiveness criterion instead predicts that the manager may seek models that *minimize* the informational content of her signal, a form of confirmation bias. The next section will study exactly this implication of the decisiveness criterion.

3.2.2 Confirmation Bias

Consider a setting with binary states: $\Theta = \{\theta_l, \theta_h\}$; fix the decision problem $\mathcal{D}$ and the signal realization s. Say that m is in favor of θ_h if $m(s|\theta_h) \geq m(s|\theta_l)$, and that m is in favor of θ_l if $m(s|\theta_h) \leq m(s|\theta_l)$. Also, say that a model m is *interior* if $p^m(\cdot|\theta) \neq \delta_\theta$ for any state θ. The DM then exhibits confirmation bias in the following sense:

Corollary 2. For any interior m in favor of θ_h, there exists a threshold prior belief $\overline{p} > 0$ such that if $\rho(\theta_h) \leq \overline{p}$, $I_\mathcal{D}(m'|s) \geq I_\mathcal{D}(m|s)$ for any m' in favor of θ_l. Furthermore, for $\rho(\theta_h) \leq \overline{p}$ and any m'' satisfying $m''(s|\theta_h)/m''(s|\theta_l) \leq m(s|\theta_h)/m(s|\theta_l)$, we also have $I_\mathcal{D}(m'|s) \geq I_\mathcal{D}(m''|s)$ for any m' in favor of θ_l.

In words: given model m providing evidence towards one conclusion, there exists a range of prior beliefs under which the DM prefers to adopt an alternative model that interprets the signal as evidence for the opposing conclusion. Furthermore, this range of prior beliefs is decreasing in the strength of the evidence under m.

Example (Prior-Based Polarization). Following the classic experimental setup of Darley and Gross (1983), suppose that there are two subjects whose priors disagree about the state $\Theta = \{\theta_h, \theta_l\}$ (in their experiment, the reading level of a student), but who are otherwise identical. In particular, we have $\rho_A(\theta_h) > \rho_B(\theta_h)$: subject A believes θ_h is more likely whereas subject B believes θ_l is more likely.

Suppose that both subjects receive the same signal s (results from an ability test), where there is uncertainty over how to interpret s (e.g. uncertainty over how difficult or diagnostic the test is): both subjects entertain models in favor of θ_h and in favor of θ_l. Assume that the evidence does not completely resolve the debate; all models the DMs entertain are interior.

Corollary 2 states that there exists $\overline{p}, \underline{p} \in (0,1)$ such that if $\rho_A(\theta_h) > \overline{p}$ and $\rho_B(\theta_h) < \underline{p}$, subject A will adopt a model in favor of θ_h and subject B will adopt a model in favor of θ_l — that is, after observing the *same* common signal, beliefs become increasingly polarized. Darley and Gross (1983) and similar studies (Lord et al., 1979; Plous, 1991) document exactly this kind of belief polarization along prior beliefs. ▲

The decisiveness criterion not only rationalizes the evidence for confirmation bias discussed above, but also produces predictions regarding the when the bias is likely to arise, as the example below illustrates.

Example (Confirmation Bias in Hiring). Consider again the hiring manager example from the discussion of overprecision. Suppose that the hiring manager faces the decision problem

	θ_h	θ_l
hire	v	$-k$
reject	0	0

and that under the true model, $r \equiv m_T(s|\theta_h)/m_T(s|\theta_l) > 1$; under the true model, s is good news about the productivity of the candidate, where r denotes the diagnostic strenght of this signal. Note that residual uncertainty is increasing in $p(\theta_h)$ for $p(\theta_h) \in [0, k/(k+v)]$. Bayes' rule implies that so long as $\rho(\theta_h) < \frac{k}{rv+k}$, $p^{m_T}(\theta_h) < k/(k+v)$, and so for $\rho(\theta_h) < \frac{rk}{v+rk}$, the hiring manager prefers the uninformative model $m_\varnothing$ over the true model m_T. Intuitively, for $\rho(\theta_h)$ low enough, the hiring manager is confident that rejecting the candidate is the right call, and so will ignore a disconfirmatory signal that induces greater uncertainty over whether she should hire. ▲

This analysis demonstrates two key predictions that the decisiveness criterion makes regarding confirmation bias. First, the decisiveness criterion predicts limits to confirmation bias: if under the true model the disconfirmatory signal is sufficiently informative — that is, if r is large — the DM will find the true model more decisive than the uninformative model, and so will not exhibit confirmation bias. In fact, the analysis on overprecision implies that in such cases the DM will instead exhibit over-inference — to the extent there are interpretations of the data that overstate the informational content of the signal, the DM will select these interpretations. These limits to confirmation bias depend in an intuitive way on the DM's prior: the more certainty afforded by the DM's prior, the greater scope of disconfirmatory models the DM will find less decisive than the uninformative model.

The second prediction is that the DM's decision problem determines the regions of prior beliefs where confirmation bias operates. This is clearly illustrated in the hiring manager example discussed above: fixing the signal strength, parameterized by r, the DM exhibits confirmation bias for priors in the region $\rho(\theta_h) \in [0, \frac{k}{rv+k}]$. We can see that as v increases — that is, as hiring the candidate becomes a more attractive option — greater certainty in the DM's prior is required to sustain confirmation bias. Contrast this to approaches to modeling confirmation bias, such as Rabin and Schrag (1999), in which the region of priors in which the bias is operates is exogenously determined — typically $\rho \in [0, 0.5)$. Under my model, a disconfirmatory signal is not necessarily one that moves the DM's beliefs towards 50-50, but rather one that reduces DM's confidence in the course of action they would have chosen in the absence of the signal.

Taken together, the results on overprecision and confirmation bias imply that the DM's beliefs will exhibit the following pattern, to the extent the DM entertains models that both overstate and understate the decisiveness of her signal: if the DM is initially uncertain (with respect to her available courses of action), she will exhibit overreaction to information. However, if the DM's prior is sufficiently concentrated in one state, she will tend to underreact to disconfirmatory news while continuing to overreact to confirmatory news. As such, the decisiveness criterion predicts the tendency of individuals to both "jump to conclusions" — evaluators form strong impressions based on unreliable diagnostic tools, noise traders act on illusory correlations in price data, and forecasters are overstate the accuracy of their estimates — and, once strong convictions have been formed, to explain away disconfirmatory evidence: stereotypes tend to persist even in the face of counterexamples, and initial disagreement can become increasingly polarized in light of new information.

3.2.3 Investor Sentiment

The DM is trying to predict whether a stock price will increase (θ_h) or decrease (θ_l). The DM holds the prior belief $\rho(\theta_h)$, corresponding to the DM's belief regarding the drift the stock price exhibits on average, and observes the history of price changes, $s = (s_{t-1}, s_{t-2}, ..., s_0)$ where for simplicity we assume each s_t is binary, with $s_t = 1$ corresponding to a price increase and $s_t = 0$ corresponds to a price decrease. Denote $\tilde{s}_t = (s_t, s_{t-1}, ..., s_0)$ as the history of price changes leading up to date t. Suppose that under the true model, stock prices follow a random walk: for all s, $m_T(s|\theta_h) = m_T(s|\theta_l)$.

Suppose that the DM entertains the true model m_T, as well as models under which the stock price is predictable on the basis of previous returns. For instance, following Barberis et al. (1998), the DM may entertain a "mean-reverting" model in which price increases are more likely to be followed by price decreases: $m((s_{t-1} = 1, \tilde{s}_{t-1})|\theta_h) < m((s_{t-1} = 1, \tilde{s}_{t-1})|\theta_l)$ for all $\tilde{s}_{t-1}$, or an "extrapolative" model in which price increases are more likely to be followed by further increases: $m((s_{t-1} = 1, \tilde{s}_{t-1})|\theta_h) > m((s_{t-1} = 1, \tilde{s}_{t-1})|\theta_l)$ for all $\tilde{s}_{t-1}$. When does the decisiveness criterion predict that these latter models, in which prices are predictable on the basis of the past, will be selected over the true model?

Proposition 1 implies that regardless of the DM's decision problem, so long as M contains models m, m' for which $p^m(\theta_h|s) < p^{m_T}(\theta_h|s) < p^{m'}(\theta_h|s)$, the DM will not adopt the true model (ignoring ties in the decisiveness ranking). In particular, this implies that if the DM entertains both mean-reverting and extrapolative models of stock prices, she will not adopt the true model over either of these two classes of models, regardless of the price history and the decision problem. For intuition behind this prediction, suppose that the DM chooses between actions $A = \{buy, sell\}$, where *buy* delivers a higher payoff than *sell* for $\theta = \theta_h$, and vice versa for $\theta = \theta_l$, and suppose that in the previous period a gain was realized (the same intuition holds for the case for $s_{t-1} = 0$). Note that if the DM preferred *buy* under the true model, the extrapolative model provides yet stronger evidence in favor of *buy*, and if the DM instead preferred *sell* under the true model, the mean-reverting model provides yet stronger evidence in favor of *sell*; in either case, the DM adopts an alternative model over the true model. Proposition 1 states that this same intuition extends to an arbitrary set of actions and payoffs.

Whereas Barberis et al. (1998) defend their assumption that the DM does not entertain the true random walk model by appealing to constraints on the speed of learning or errors in Bayesian inference over the space of models, my framework highlights a distinct force that drives the DM to seek other models over the true model — in particular, that the DM perceives the alternative models as providing better guidance than the true model.

What does the decisiveness criterion have to say about when the DM adopts the mean-reverting vs. the extrapolative model? A set of stylized facts reviewed in Barberis et al. (1998) suggest that prices underreact to news in normal times, but tend to overreact to news following periods of sustained positive or negative news, consistent with investors adopting an extrapolative model following sustained price increases or decreases, and a mean-reverting model otherwise. The decisiveness criterion can rationalize these patterns if we expand the set of extrapolative models to include those where the informational content of a price change s_{t-1} is increasing in the number of consecutive previous price changes that agree with s_{t-1}. In other words, under such an model, the DM extrapolates more in response to a price change when it follows a series of price changes in the same direction. Suppose the DM entertains such an extrapolative model m_E, as well as a mean-reverting model m_R, satisfying

$$p^{m_R}(\theta_h|(s_{t-1} = 1, \tilde{s}_{t-2})) < p^{m_T}(\theta_h|(s_{t-1} = 1, \tilde{s}_{t-2})) < p^{m_E}(\theta_h|(s_{t-1} = 1, \tilde{s}_{t-2}))$$
$$p^{m_R}(\theta_h|(s_{t-1} = 0, \tilde{s}_{t-2})) > p^{m_T}(\theta_h|(s_{t-1} = 0, \tilde{s}_{t-2})) > p^{m_E}(\theta_h|(s_{t-1} = 0, \tilde{s}_{t-2}))$$

for all $\tilde{s}_{t-2}$. Suppose that $s_{t-1} = 1$. By Extremeness and Certainty Preference, residual uncertainty $R_{\mathcal{D}}(\cdot)$ is single-peaked in $p^m(\theta_h|s)$. This implies that there does not exist a decision problem for which an increase in $p^{m_E}(\theta_h|s)$ results in the DM switching models from m_R to m_E, whereas there exists decision problems where the contrary is true. Therefore, the DM will tend to adopt the extrapolative model following a price increase when it follows a series of price increases, and the same logic extends to price decreases.

As Barberis et al. (1998) demonstrate, a fit-based model selection criteria can rationalize these patterns in model selection: the extrapolative model becomes more likely in light of sustained price increases and decreases, and the mean-reverting model is becomes more likely when prices are not trending. My framework provides an alternative channel that could drive these patterns: in times of sustained price changes, the extrapolative model provides stronger guidance to investors, whereas when prices oscillate, the extrapolative model is less conclusive and so investors adopt the mean-reverting model.

3.3 Comparison to Alternative Selection Criteria

Here, I illustrate the differences between the decisiveness criterion and related model selection criteria.

3.3.1 Decisiveness vs. Parsimony

Given the complexity of information decision-makers are often confronted with, it is plausible that individuals may seek parsimonious models of the world to organize data and make decisions, consistent with the view that individuals are "cognitive misers" (Kahneman and Tversky, 1973). Does the decisiveness criterion generate a demand for parsimonious models? Given the difficulty of formalizing a general notion of model complexity, I instead outline key intuitions regarding how the decisiveness criterion relates to a demand for model parsimony through a simple setting, first analyzed in Hong et al. (2007).

The DM is trying to predict whether the stock price will increase (θ_h) or decrease (θ_l) in the following period, and observes two sources of news: $s = (s_A, s_B)$, where $s_i \in \{0, 1\}$ for $i = A, B$. Motivated by the view that individuals are "cognitive misers" and so seek simplified models of the world, Hong et al. (2007) analyze a situation in which the DM selects one of two models, each of which uses only a single news source (either A or B) to predict the stock price. In this example, I study when the decisiveness criterion generates demand for such parsimonious models.

Formally, say that model m ignores news source i if for any realization of s_{-i}, posterior beliefs are constant in the value of s_i: this holds only if for all s_{-i},

$$\frac{m(s_i = 1, s_{-i}|\theta_h)}{m(s_i = 0, s_{-i}|\theta_h)} = \frac{m(s_i = 1, s_{-i}|\theta_l)}{m(s_i = 0, s_{-i}|\theta_l)}$$

First, note that under the unrestricted model space $\mathcal{M}$, given any signal realization, a model that ignores source i can implement any posterior belief, as can a model utilizing both news sources. As a result, without further structure on the model space, both classes of models are equally privileged under the decisiveness criterion. As such, we impose the following structure on M: suppose that M contains the set of models m satisfying the following conditions:

1. Under m, the s_A, s_B are independent conditional on θ. Let $m(s_i|\theta) = \sum_{s_{-i}} m(s_i, s_{-i}|\theta)$.
2. $m(s_i|\theta) \in [\underline{q}_i, \overline{q}_i] \subset [0, 1]$ for all θ, i.
3. $m(s_i = 1|\theta_h) \geq m(s_i = 1|\theta_l)$ for $i = A, B$.

Condition 2 ensures that that no model completely eliminates residual uncertainty, and Condition 3 implies an ordering on news: $s_i = 1$ implies good news about the stock price, whereas $s_i = 0$ implies bad news. Under these conditions, the signal realization crucially determines whether the DM selects a model that ignores a news source, as the following cases illustrates.

Case 1: Mixed signals. Suppose that $s_A = 1, s_B = 0$ (the analysis is identical for the case where $s_A = 0, s_B = 1$). In this case, the set of models that the DM adopts, for any decision problem, are models that ignore either source A or B. To see this, note that any m that maximizes $p^m(\theta_h)$ must ignore source B, the source delivering bad news; similarly, any m minimizing $p^m(\theta_h)$ must ignore source A, the source delivering good news. Proposition 1 implies that only models implementing these extreme beliefs will be selected under the decisiveness criterion (ignoring ties in the decisiveness ranking) for any decision problem.

Case 2: Aligned signals. Suppose that $s_A = 1, s_B = 1$ (the analysis is identical for the case where $s_A = 0, s_B = 0$). In this case, the set of models the DM adopts, for any decision problem, are models that either ignore *both* sources, or take both sources into account. The logic is similar to the case above: any m that maximizes $p^m(\theta_h)$ must take both sources into account, and any m minimizing $p^m(\theta_h)$ must ignore both sources.

Under the decisiveness criterion, the DM does not value parsimony *per se*, as the contrast between the two cases illustrates: the DM adopts the simplified models considered in Hong et al. (2007) only when the two sources produce conflicting news. Under this account, the DM dislikes incorporating additional dimensions into her model to the extent these dimensions increase her residual decision uncertainty. As such, to the extent that news sources are correlated under the true model and therefore are more likely to agree, the DM will more often adopt models that take these multiple sources into account. However, to the extent news sources are independent and therefore more likely to disagree, the DM will tend to adopt models that selectively attend to certain sources.[4]

3.3.2 Decisiveness vs. Blackwell Ordering

Note that in my framework, each model m is a Blackwell experiment (Blackwell, 1953), and so can be ranked according to the Blackwell ordering, a measure of the informativeness of an experiment. Formally, m dominates m' in the Blackwell order if for all priors ρ and decision problems $\mathcal{D} = (A, u)$,

$$\sum_s \sum_\theta \max_{a \in A} u(a, \theta) p^m(\theta|s) p^m(s) \geq \sum_s \sum_\theta \max_{a \in A} u(a, \theta) p^{m'}(\theta|s) p^{m'}(s)$$

[4]Note that such models need not be "simple"; a model that assigns differing diagnostic weights to each news source is arguably more complex than a model that treats each source as identical, and yet will tend be more decisive in the case of mixed signals.

where $p^m(s) = \sum_\theta m(s|\theta)\rho(\theta)$ gives the likelihood of observing s under model m.

While both the decisiveness criteria and the Blackwell order capture notions of the informativeness of a model, note that the decisiveness criteria does not respect the Blackwell order. To see why not, as Section 3.2.2 on confirmation bias demonstrates, a completely uninformative model $m_\varnothing$ may be more decisive than an informative model m, whereas m necessarily dominates $m_\varnothing$ in the Blackwell order. The key distinction driving the discrepancy is that the decisiveness criterion is an ex-post notion of informativeness evaluated for a given signal realization, whereas the Blackwell order is an ex-ante notion. Indeed, it can be shown that if m dominates m' in the Blackwell order, then the *average* decisiveness of m must be greater than that of m': $\sum_s I_\mathcal{D}(m|s)p^m(s) \geq \sum_s I_\mathcal{D}(m'|s)p^{m'}(s)$.[5] As such, even if m is more decisive than m' on average, for some signal realizations it may be the case that m' is more decisive than m.

4 Selection as a Function of Payoffs and Objectives

In this section, I study how model selection under the decisiveness criterion varies with the objectives of the DM. In a series of applications, I discuss how these comparative statics can generate documented context effects and attribution errors, as well as predict novel forms of belief polarization.

4.1 Maximal Payoff Profile Improvements

Consider two decision problems $\mathcal{D} = (A, u), \mathcal{D}' = (A', u)$, where $A' = A \cup \{a'\}$. Say that $\mathcal{D}'$ improves the maximal payoff profile of $\mathcal{D}$ if $u(a', \theta) > \max_{a, \in A} u(a, \theta)$ for some θ — that is, $\mathcal{D}'$ is formed from $\mathcal{D}$ by adding an action a' that improves the maximal payoff for at least one outcome. For such payoff profile improvements, let $\Delta u_\theta = \max_{a \in A'} u(a, \theta) - \max_{a \in A} u(a, \theta)$ denote the improvement in the maximal payoff associated with state θ when moving from $\mathcal{D}$, to $\mathcal{D}'$.

The following gives a condition on how model selection must respond to such payoff profile improvements, in the case where the payoff profile-improving action a' is not chosen:

Proposition 3. Fix the menu of models M. Suppose $\mathcal{D}'$ improves the maximal payoff profile of $\mathcal{D}$ via a'. Then, for any $m \in C_\mathcal{D}(M|s)$ and $m' \in C_{\mathcal{D}'}(M|s)$, if m' does not recommend a' from $\mathcal{D}'$, then m', m must satisfy $\sum_\theta p^m(\theta|s)\Delta u_\theta \leq \sum_\theta p^m(\theta|s)\Delta u_\theta$.

To parse this condition, consider the case where a' improves the maximal payoff associated with a single state θ. In this case, the proposition states that $p^m(\theta|s) < p^{m'}(\theta|s)$: any change in models must result in lower posterior beliefs in the state for which a' improves the maximal payoff. When a' improves the maximal payoff associated with multiple states, the proposition states that any change in models must result in a lower payoff-weighted average

[5] In Section 7.2, I analyze precisely this formulation of ex-ante decisiveness, and contrast its predictions with the ex-post formulation.

of posterior beliefs in the improved states. For intuition behind this result, recall that the decisiveness criterion favors models that concentrate beliefs in states under which the recommended action will be ex-post optimal. As a result, the addition of an unchosen action that increases the maximal payoff achievable for a set of states reduces the decisiveness of models that induce high beliefs in those states.

This comparative static on the DM's choice of model selection has implications for the DM's choice of action: the addition of an unchosen action a leads the decision-maker to select models that recommend against choosing a — as well as similar actions. This implication is most clear in the case of binary states, as summarized in the corollary below:

Corollary 3. Consider a binary state setting, where $\Theta = \{\theta_h, \theta_l\}$. Suppose $\mathcal{D}'$ improves the maximal payoff associated with θ_h of $\mathcal{D}$ via a'. If a' is not among the actions chosen in $\mathcal{D}'$, then for any action a chosen from $\mathcal{D}$ and any action a'' chosen from $\mathcal{D}'$, $u(a, \theta_h) \geq u(a'', \theta_h)$ and $u(a, \theta_l) \leq u(a'', \theta_l)$.

This prediction can account for findings in the experimental literature in which the addition of an unchosen action reduces subjects' propensity to choose similar actions, or equivalently increases their propensity to choose dissimilar actions.

Example (Reason-Based Choice). Consider the following experimental finding from Tversky and Shafir (1992). Subjects decide whether to buy a CD player. In the first treatment (the low-conflict treatment), subjects can choose to either buy a mid-range CD player or to defer the purchase. In the second treatment (the high-conflict treatment), subjects can choose to either buy a mid-range CD player, buy a top-of-the-line CD player, or defer the purchase. Their finding is that the proportion of subjects choosing to defer the purchase is greater in the high-conflict treatment vs. the low-conflict treatment. Tversky and Shafir (1992) interpret these findings as suggesting that in the high-conflict treatment, subjects looking to buy a CD player need to weigh the lower price of the midrange model against the higher quality of the top-of-the-line player, a difficult tradeoff that makes deferring the purchase an easier decision to justify.

To translate this setting into the framework, normalize the DM's payoff from deferring to 0, and suppose that the DM's payoff from purchasing a CD player is $\theta q - k$, where q and k are the quality and price of the player, and θ parameterizes how much the DM weighs quality over price. Suppose that the DM is uncertain about the value of θ, and suppose that this state is binary: $\theta \in \{0, 1\}$. Let (q_l, p_l) and let (q_h, p_h) denote the price and quality of the mid-range and top-of-the-line CD player, respectively, and suppose that $k_h > k_l$ and $v_h > v_l > 0$, where $v_h \equiv q_h - k_h$ and $v_l \equiv q_l - k_l$ denote the net quality of the products. The decision problems corresponding to the low-conflict and high conflict treatments are

$\mathcal{D}_{low}$: Low-conflict	$\theta = 1$	$\theta = 0$
midrange	v_l	$-k_l$
defer	0	0

$\mathcal{D}_{high}$: High-conflict	$\theta = 1$	$\theta = 0$
top-of-line	v_h	$-k_h$
midrange	v_l	$-k_l$
defer	0	0

Suppose further that $k_l/(v_l + k_l) < k_h/(v_h + k_h)$ — that is, conditional on choosing to purchase a CD player, the choice is not obvious — there exists a range of beliefs for which the DM prefers *midrange* over *top-of-line*, and vice versa.

To begin, fix a signal realization s; we will first show there exists some set of models M such that if the DM entertains M, she will choose *midrange* in the low-conflict treatment and *defer* in the high conflict treatment[6]. First note that by Proposition 1, we can restrict attention to the extreme models $\overline{m}$, $\underline{m}$ satisfying $\overline{m} \in \arg\max_{m\in M} p^m(\theta = 1|s)$ and $\underline{m} \in \arg\min_{m\in M} p^m(\theta = 1|s)$, respectively; Let $\overline{p} = \max_{m\in M} p^m(\theta = 1|s)$ and $\underline{p} = \min_{m\in M} p^m(\theta = 1|s)$ denote the respective extreme posteriors.

Suppose that M is such that $\underline{p} < k_l/(v_l + k_l)$, $k_l/(v_l + k_l) < \overline{p}$; the DM's choice of model is material for whether she buys a CD player or not. Proposition 3 then implies that if $\overline{p} < k_h/(v_h + k_h)$ — that is, no model recommends *top-of-line* — the addition of *top-of-the-line* must increase the decisiveness of $\underline{m}$ relative to $\overline{m}$. It can be shown that for some M, this leads the DM to switch from $\overline{m}$ in the low-conflict treatment to $\underline{m}$ in the high-conflict treatment, which in turn induces a switch from *midrange* to *defer*[7].

Therefore, adding an unchosen alternative, the top-of-the-line player, can cause the DM to switch from the choosing the midrange player to deferring the decision via a change in models from $\overline{m}$ to $\underline{m}$. The addition of the top-of-the-line player makes $\overline{m}$ a less satisfying justification for the DM's decision, as $\overline{m}$ now leaves the DM uncertain about which player to buy, causing the switch to $\underline{m}$, which makes a comparatively more decisive recommendation that the DM should defer the purchase.

Now, note that if the DM chooses *defer* under in the low-conflict treatment, she must also choose *defer* in the high-conflict treatment. To see this, let m be the model selected under $\mathcal{D}_{low}$ and take any $m' \in M$ that does not recommend *defer* from $\mathcal{D}_{high}$; let $p^m = p^m(\theta = 1|s)$, $p^{m'} = p^{m'}(\theta = 1|s)$ denote the respective model-induced posteriors. By assumption, we have $I_{\mathcal{D}_{low}}(m|s) > I_{\mathcal{D}_{low}}(m'|s)$ which implies $p^m v_l < (1 - p^{m'})k_l$; this, along with the maintained assumptions on k and l, implies that $I_{\mathcal{D}_{high}}(m|s) > I_{\mathcal{D}_{high}}(m'|s)$. In line with the intuition behind Proposition 3, the addition of *top-of-line* penalizes models that place higher weight on the improved state $\theta = 1$, which precludes a switch from midrange to defer.

Therefore, for any signal realization, the DM may switch from *midrange* in the low-conflict treatment to *defer* in the high conflict treatment, but must choose *defer* in the high-conflict treatment if she chose *defer* in the low-conflict treatment. As such, fixing any probability distribution governing over signals, the proportion of subjects choosing *defer* in the high-conflict treatment is higher compared to that of the low-conflict treatment, in line

[6] An underlying signal structure in this application could involve, for example, the models specifying different interpretations of an advertisement the DM received regarding the CD players.

[7] The set of $\overline{p}$ and $\underline{p}$ that yield this choice pattern is characterized by the condition $\underline{p} \in \left[\frac{(1-\overline{p})k_l}{v_h}, \frac{(1-\overline{p})k_l + \overline{p}(v_h - v_l)}{v_h}\right]$ which is non-empty given the maintained assumptions $\underline{p} < k_l/(v_l + k_l)$, $k_l/(v_l + k_l) < \overline{p} < k_h/(v_h + k_h)$

with the experimental findings.[8] ▲

4.2 Reductions in Action Value

Consider two decision problems $\mathcal{D} = (A, u), \mathcal{D}' = (A, u')$. Say that action $a^* \in A$ is uniformly worse in $\mathcal{D}'$ relative to $\mathcal{D}$ if $u'(a^*, \theta) \leq u(a^*, \theta)$ for all θ and $u'(a, \theta) = u(a, \theta)$ for all θ, $a \neq a^*$. The following gives a condition for how the model must change when an action is made uniformly worse:

Proposition 4. Suppose a^* is uniformly worse in $\mathcal{D}'$ relative to $\mathcal{D}$. If m recommends a^* from $\mathcal{D}'$ and m' does not recommend a^* from $\mathcal{D}$ then $m \notin C_{\mathcal{D}}(M|s), m' \in C_{\mathcal{D}}(M|s) \implies m \notin C_{\mathcal{D}'}(M|s)$.

In words, the decisiveness of any model that recommends a^* must decrease relative to a model that did not recommend a^* when a^* is made uniformly worse. Intuitively, as a^* is made uniformly worse, a model that recommends a^* provides a weaker justification toward its recommended action than does any model recommending a different action.

Proposition 4 implies that the more predisposed the DM is to taking an action, the more they will tend to adopt interpretations of the data that recommend that action, as opposed to interpretations recommending other actions. Note the key distinction between this prediction and that of a model selection criterion based on optimism, in which individuals seek interpretations of the data that increase the perceived value of their chosen course of action: a model can produce a recommendation for an action a by increasing the perceived value of a *relative* to other actions, without increasing the perceived *absolute* value of a. As such, a DM who is predisposed toward a will find a model decisive if it reduces the attractiveness of competing actions, even if it does not result in a higher perceived value of the chosen action a. In this manner, the decisiveness criterion can predict patterns that are puzzling from the perspective of a optimism-based criterion, such as individuals adopting interpretations that downplay the credibility of evidence suggesting the safety of a new vaccine, or point to the ineptitude of a political candidate. Such models do not induce optimism but may induce a sense of certainty over the correct course of action, if for instance the decision-maker is choosing whether to take the vaccine or choosing who to vote for. The following example illustrates the differences between the decisiveness criterion and selection criterion based on optimism.

Example (Wishful Thinking vs. Decisivness). Consider the following experimental results from Bastardi et al. (2011). They study a group of soon-to-be parents with similar priors — all believe that home care is superior to day care. However, parents face different incentives: some intended to use home care, whereas others intended to use day care. When shown with

[8] To formally adapt the model to the analysis of choice probabilities, let $\mu(s)$ denote the objective distribution of signals and let $S^a_{\mathcal{D}}$ denote the set of signals in s under which the DM chooses a in decision problem $\mathcal{D}$; to avoid dealing with ties in the definition of choice probabilities, assume that $A^m_{\mathcal{D}}$ and $\arg\max_{m \in M} I_{\mathcal{D}}(m|s)$ are singleton sets. The probability of choosing a in decision problem $\mathcal{D}$ is then $Q(a, \mathcal{D}) \equiv \sum_{s \in S^a_{\mathcal{D}}} \mu(s)$. My results indicate that that $Q(\textit{defer}, \mathcal{D}_{low}) \leq Q(\textit{defer}, \mathcal{D}_{high})$ for any set of models M, where the inequality is strict for some M.

a study providing evidence in favor of the effectiveness of day care, parents who intended to use home care rated the credibility of the study as low, whereas parents who intended to use day care rated the credibility of the study as high. Can wishful thinking explain these results?

Here, the state is given by $\theta = (\theta^h, \theta^d)$, where $\theta^h = 1$ ($\theta^h = 0$) corresponds to home care being effective (not effective), and likewise for θ^d, which corresponds to the effectiveness of day care. For simplicity, suppose that DMs believe that the quality of day care and home care is independent. In line with the study, suppose that DMs hold common priors over the state. DMs face the menu $A = \{a_h, a_d\}$, a choice between home care and day care. Assume that the payoff from choosing each type of care is greater if that type of care is effective:

$$u(a_h, (\theta^h = 1, \theta^d)) > u(a_h, (\theta^h = 0, \theta^d)) \qquad \text{for all } \theta_d$$
$$u(a_d, (\theta^h, \theta^d = 1)) > u(a_d, (\theta^h, \theta^d = 0)) \qquad \text{for all } \theta_h$$

and that the payoff from choosing a given type of care is independent of whether the alternative is effective or not: $u(a_h, (\theta^h, \theta^d))$ is constant in θ_d and $u(a_d, (\theta^h, \theta^d))$ is constant in θ_h. To rationalize the baseline heterogeneity in choice observed in the experiment, suppose that the DMs who initially choose home care find day care uniformly more more costly than DMs who initially choose day care.

Now suppose, as in the experiment, that the DM sees a study s providing evidence that day care is effective. The DM entertains two models; either the study is credible (m_1) or not (m_2), where $p^{m_1}(\theta^d = 1|s) > p^{m_2}(\theta^d = 1|s) = \rho(\theta^d = 1|s)$; assume that the study is uninformative about the effectiveness of home care under both m_1 and m_2.

First consider the inferences of DMs who chose day care at baseline. Under either model, they would choose day care, but m_1 induces lower regret associated with that choice, and so these DMs select m_1; they rate the study as credible. Note that a model of wishful thinking would make identical predictions in this case; an agent committed to choosing day care would want to believe that day care is effective.

Now consider the inferences of the DMs who chose home care at baseline. According to Proposition 4, these DMs find m_1 less decisive relative to m_2 than do DMs who choose day care at baseline. These DMs are therefore inclined to select m_2; they rate the study as not credible. Note that wishful thinking *cannot* explain this finding: an agent who simply wishes to be optimistic would have no motive to downplay the credibility of good news regarding an option they don't plan to choose (and would in fact avoid doing so insofar as there are costs to warping one's beliefs, as in Caplin and Leahy (2019)); the decisiveness criterion, however, generates exactly this kind of motive. ▲

4.2.1 Application: Social Attributions

Consider a set of stylized facts from the psychology literature: when making inferences about others' dispositional traits on the basis of their behavior, individuals tend to ignore the confounding role of situational factors in determining behavior, thus committing the so-called fundamental attribution error (Ross, 1977). However, there are exceptions to this behavior — in particular, when forming inferences about ingroup members on the basis of poor behavior, individuals commit the opposite error — they explain away the behavior as the result of situational factors, as opposed to dispositional traits (Vonk and Konst, 1998).

The decisiveness criterion can rationalize these patterns as a product of differences in the decision-makers inferential goals when making inferences about a stranger vs. an ingroup member. Concretely, suppose that the DM is uncertain whether an actor is low-type (θ_l) or high-type (θ_h). The actor is either a stranger or ingroup member, corresponding to the decision problems $\mathcal{D}_{st}$ and $\mathcal{D}_{in}$ respectively, and in either case, the DM must decide whether to interact with the actor: $A = \{interact, don't\ interact\}$, where where *interact* delivers higher payoffs in θ_h and *don't interact* delivers higher payoffs in θ_l. Suppose that the DM has a preference for interacting with an ingroup member as opposed to a stranger – that is, *interact* is uniformly improved in $\mathcal{D}_{in}$ relative to $\mathcal{D}_{st}$, whereas the payoffs of *don't interact* are identical across both decision problems.

In either case, the DM observes a signal s, the actor's behavior, and entertains the following models: under m, behavior is informative of the actor's type; under $m_\varnothing$, behavior is explained away by situational factors, and is thus uninformative about type; and the true model $m_T = \lambda m + (1-\lambda)m_\varnothing$ reflects uncertainty over whether the actor's behavior was the result of situational factors or their underlying type.

First consider the DM's inferences over the stranger. Here, Corollary 1 naturally generates a tendency towards the fundamental attribution error: so long as so long as $I_{\mathcal{D}_{st}}(m_T|s) > I_{\mathcal{D}_{st}}(m_\varnothing|s)$ — that is, as long as the DM is initially uncertain over whether or not to interact with the stranger, the DM will select m, thereby attributing the stranger's behavior entirely to their type, and neglecting the confounding role of the situation.

In contrast, consider the DM's inferences over the ingroup member, in the situation where the ingroup member displayed poor behavior — $p^m(\theta_h|s) < \rho(\theta_h)$, and that for this signal realization, m recommends that *don't interact*, whereas $m_\varnothing$ recommends *interact*. Here, Proposition 4 states that the DM's stronger preference for interacting with the ingroup member increases the decisiveness of $m_\varnothing$ relative to m when faced with an ingroup member as compared to a stranger; if this difference in preferences is sufficiently large, the DM will select $m_\varnothing$, and explain away the behavior with situational factors, thus committing a "reversal" of the fundamental attribution error. Here, the DM's predisposition towards interacting with the ingroup member increases the decisiveness of models that reinforce this course of action, as opposed to models that suggest otherwise. The decisiveness criterion also predicts an asymmetry in attribution documented in (Vonk and Konst, 1998), where the fundamental attribution error only reverses for inferences over negatively-valenced behavior of ingroup members. In particular, if the ingroup member instead exhibited positive behavior, with $p^m(\theta_h|s) > \rho(\theta_h)$, the DM will instead find m more decisive.

While the attribution patterns discussed above are consistent with a form of ingroup bias in which the DM seeks to interpret information in a way that is favorable towards the ingroup, the decisiveness criterion makes additional predictions that cannot be rationalized by such a bias. Under the decisiveness criterion, the DM seeks to explain away poor behavior by ingroups not because she has an innate preference for holding favorable beliefs towards the ingroup, but because this interpretation helps reduce her uncertainty over who to interact with; as such, the DM's inference will be sensitive to changes to the payoffs from interacting with the ingroup member.

For example, consider a decision problem $\mathcal{D}'_{st}$ in which the cost of not interacting with the stranger is higher than in $\mathcal{D}_{st}$ — that is, *don't interact* is uniformly worse in $\mathcal{D}'_{st}$ relative to $\mathcal{D}_{st}$.[9] Proposition 4 predicts that such a shift in payoffs can induce the DM to explain away negative behavior in $\mathcal{D}'_{st}$ just as she does in $\mathcal{D}_{in}$, thereby bringing inferences in both decision problems in line with each other.[10]

This analysis can be similarly applied to study discrimination in hiring decisions, and in this setting, predicts a novel channel through which taste-based discrimination can drive belief-based biases against minority applicants. The logic is analogous to the example above — if a hiring manager incurs greater costs to hiring a minority applicant due to discriminatory tastes, they will have a greater inclination against adopting positive interpretations of the data that would recommend a decision to hire — such interpretations lead the manager to greater decision uncertainty. This logic similarly suggests channels for interventions in such a setting — since managers' inferential biases are not due to the minority status of applicants per se, but rather due to the increased costs of hiring such applicants, the model predicts that providing minority hiring incentives can reverse this particular inferential bias.

4.2.2 Application: Belief Polarization

As alluded to in the previous section, Proposition 4 highlights a channel for differences in DMs' objectives to lead to belief polarization. To illustrate, consider a society of DMs who have identical priors over the state, $\Theta = \{\theta_h, \theta_l\}$, and choose from the same set of actions A. There are two types of DMs, type-1 and type-2, characterized by their utility functions u_1, u_2. For type-1 DMs, denote by $\overline{a}$ and $\underline{a}$ the actions that deliver the highest payoffs under θ_h and θ_l, respectively, and assume that $\overline{a} \neq \underline{a}$, so that the DM does not have a dominant course of action. Denote $(\overline{v}, -\overline{k}) = (u_1(\overline{a}, \theta_h), u_1(\overline{a}, \theta_l))$ and $(\underline{v}, -\underline{k}) = (u_1(\underline{a}, \theta_h), u_1(\underline{a}, \theta_l))$. Suppose that the utility of type-2 DMs satisfies $(u_2(\overline{a}, \theta_h), u_2(\overline{a}, \theta_l)) = (\overline{v} - c, -\overline{k} - c)$, with $u_2 = u_1$ otherwise. In words, type-2 DMs find it more costly to take $\overline{a}$ relative to type-1 DMs.

Suppose that both types of DMs entertain the same set of models M and recieve a common signal s. Proposition 1 implies that only two models may ultimately be chosen: the model that induces the maximum belief in θ_h and the model that induces the minimum

[9]For example, $\mathcal{D}_{st}$ might concern a stranger who lives three blocks down from the DM, whereas $\mathcal{D}'123_{st}$ concerns a stranger who happens to be the DM's new next-door neighbor.

[10]Note that while a model of ingroup bias may predict that inferences towards outgroups may become more favorable as the DM finds it less costly to interact with outgroups due to the increased costs of holding such distorted beliefs, such a model cannot easily rationalize why such a change in incentives would induce the DM to interpret the behavior of outgroup members in an overly positive light, as in the example.

belief; let $\overline{m}$ and $\underline{m}$ denote those models, respectively, and let $p^{\overline{m}}, p^{\underline{m}}$ denote the beliefs those models induce in θ_h given the signal. In this setting, say that belief polarization occurred if one type of DM adopts $\overline{m}$ to interpret the data, whereas the other type of DM adopts $\underline{m}$.

To take a concrete example, the states could encode whether a new disease is dangerous (θ_h) or benign (θ_l), where A denotes the level of precautions the DM takes against the disease, with $\overline{a}$ and $\underline{a}$ indicating the highest and lowest levels of precaution, respectively. Relative to type-1 DMs, type-2 DMs face a higher cost to taking high precautions in either state[11]. The signal s could be composed of opposing information from two news sources — one of which claiming that the disease is likely dangerous, the other of which claiming the disease is likely benign, where the models in M specify the credibility of the two sources.

The following result characterizes conditions under which belief polarization occurs in such a setting.

Corollary 4. Let $\overline{c}$ denote the maximum value of c such that for type-2 DMs, $\overline{a}$ maximizes expected utility for some belief. Then, for $c \in (0, \overline{c})$, there exists a set of models M for which belief polarization occurs, and furthermore, any such M exhibits

$$p^{\underline{m}} \in \left(\underline{L}(p^{\overline{m}}), \underline{U}(p^{\overline{m}}, c)\right)$$
$$p^{\overline{m}} \in \left(\overline{L}(p^{\underline{m}}), \overline{U}(p^{\underline{m}}, c)\right)$$

where the bounds $\underline{L}, \underline{U}$ are decreasing in $p^{\overline{m}}$ and $\overline{L}, \overline{U}$ are decreasing in $p^{\underline{m}}$, and $\underline{U}, \overline{U}$ are increasing in c.

The fact that that $\underline{U}, \overline{U}$ are increasing in c implies that the set of M for which polarization occurs is expanding in the magnitude of preference heterogeneity. In other words, the more a society disagrees over their objectives, the greater the scope for polarization. Also, that the bounds on $p^{\overline{m}}$ are decreasing in $p^{\underline{m}}$, and vice versa for the bounds on $p^{\underline{m}}$, reflects the fact that for polarization to occur, $\overline{m}$ and $\underline{m}$ must provide interpretations of comparable strength. This implies a key limitation to polarization as predicted by the decisiveness criterion: if one model provides strong enough evidence toward its favored state relative to other models, polarization will disappear as all DMs adopt that model.

Like the decisiveness criterion, a model selection criterion based on optimism can also generate belief polarization through heterogeneity in DMs' objectives, as Caplin and Leahy (2019) demonstrate. Note, however, that wishful thinking cannot generate polarization in the above setting if $\overline{v}, \underline{v}, \overline{k}, \underline{k} > 0$ across all actions. In this case, utility is always lower in θ_h than under θ_l, and so DMs engaged in wishful thinking have no motive to adopt $\overline{m}$ over $\underline{m}$ — in the example above, there is no reason for a wishful thinker to discredit news that the disease is likely benign, which is strictly positive news. In many settings,

[11] This variation in costs could result from differences in "real" costs — for example, DMs employed in essential jobs may have higher costs to staying in lockdown. This variation could also be due to "social" costs — for example, if taking the vaccine results in the DM being viewed as a non-conformer by her social group.

wishful thinking struggles to generate polarization in this same sense — for example, it seems difficult to argue that a belief that an impending influx of immigrants will reduce employment opportunities and lower wages, or a belief that the opposing administration is running the economy into the ground, represents optimistic thinking. Wishful thinking similarly struggles to explain the apparent effectiveness of mudslinging and negative ad campaigns in generating polarization. The decisiveness criterion, however, can support the adoption of such pessimistic interpretations of the data, so long as they guide the DM's decision-making.

4.3 Decision-Relevance of States

Suppose that the state space is given by the product space $\Theta = \Theta_1 \times \Theta_2 \times ... \times \Theta_K$. For a decision problem $\mathcal{D} = (A, u)$, say that Θ_k is not *decision relevant* if for all $a \in A$, $u(a, \theta)$ is constant in θ_k.

Under the decisiveness criterion, if a state Θ_k is not decision relevant, then under certain conditions on M, the decision-maker will favor models that neglect the role of Θ_k in determining the signal—that is, models for which $m(s|\theta)$ is constant in θ_k. In particular, as formalized in Proposition 9 in Appendix A.2, this comparative static holds when Θ_k has the interpretation of a *nuisance variable* — that is, when any model that accounts for Θ_k is less informative about the remaining states compared to a model that neglects Θ_k. As the example below illustrates, this comparative static is consistent with experimental results documenting that errors in attribution are sensitive to the prediction goals of the subject.

Example (Inferential Goals in Attribution). As discussed in Section 4.3.1, the decisiveness criterion can generate the fundamental attribution error as a consequence of overprecision: an individual seeking to learn about another's type θ will ignore the role of situational factors in inferring from observed behavior s, to the extent that her prior over θ has low decisiveness. The decisiveness criterion can also account for another finding from the psychology literature – that when individuals seek to infer situational factors rather than personality traits from behavior, they tend to ignore the role of personality traits in generating behavior, thus committing a reversal of the fundamental attribution error. For example, Krull (1993) finds that when the target of inference is the situation rather than the actor's dispositional traits, observers are less likely to attribute behavior to dispositional traits.

To see how the decisiveness criterion can account for these results, consider a setting in which the DM is uncertain over both an actor's type $\Theta_1 = \{\theta_{1,h}, \theta_{1,l}\}$ and the situation $\Theta_2 = \{\theta_{2,h}, \theta_{2,l}\}$, and that Θ_1 and Θ_2 are independent under her prior, with $\rho(\theta_{1,h}) = \rho(\theta_{2,h}) = 1/2$. Consider a signal realization s, which can be attributed to the DM's type, the situation, or some combination of the two. In particular, consider the models m_1 and m_2, where for $q > 0.5$

$$m_1(s|\theta_{1,h}, \theta_2) = 1/2q,\ m_1(s|\theta_{1,l}, \theta_2) = 1/2(1-q) \qquad \text{for all } \theta_2$$
$$m_2(s|\theta_1, \theta_{2,h}) = 1/2q,\ m_1(s|\theta_1, \theta_{2,l}) = 1/2(1-q) \qquad \text{for all } \theta_1$$

That is, m_1 and m_2 fully attribute the signal to θ_1 and θ_2, respectively. The true model is given by

$$m_T(s|\theta_{1,h}, \theta_{2,h}) = \phi^2,\ m_T(s|\theta_{1,l}, \theta_{2,l}) = (1-\phi)^2$$
$$m_T(s|\theta_{1,h}, \theta_{2,l}) = m_T(s|\theta_{1,l}, \theta_{2,h}) = \phi(1-\phi)$$

for $\phi < q$. That is, under the true model, the signal is attributed to both θ_1 and θ_2, but is less informative about θ_1 than it would be under m_1, and likewise less informative about θ_2 than it would be under m_2. Let $M = \{m_1, m_2, m_T\}$.

Consider a decision problem $\mathcal{D}$ under which the DM's prior is less decisive than some model in m: $I_{\mathcal{D}}(m_\varnothing|s) < I_{\mathcal{D}}(m|s)$ for some $m \in M$. We then have, by Proposition 9, that if Θ_2 is not decision-relevant to $\mathcal{D}$, then $C_{\mathcal{D}}(M) = \{m_1\}$ – that is, if the DM is only concerned with learning about the actor's type, then the DM will adopt the model that ignores the role of situational factors in explaining behavior, consistent with the fundamental attribution error. Intuitively, a model that ignores the potentially confounding role of situational factors in attributing behavior to the actor's type provides stronger recommendations than a model that also attributes behavior to situational factors, when only the actor's type is payoff-relevant. By the same logic, if Θ_1 is not decision-relevant – that is, if the DM is concerned only with learning about situational factors – we must have $C_{\mathcal{D}}(M) = \{m_2\}$. Consistent with experimental evidence, if the DM's focus is instead on learning about the situation, the role of the actor's type in explaining behavior is ignored.

Note that when both Θ_1 and Θ_2 are decision-relevant, the DM may adopt the true model m_T. This is broadly consistent with experimental evidence from Graeber (2022), who finds that in an abstract belief updating task where the signal is a function of both Θ_1 and Θ_2, subjects neglect the role of Θ_2 in making inferences when they are incentivized to predict only Θ_1, but adjust for Θ_2 in inference when incentivized to predict both Θ_1 and Θ_2. As the abive example illustrates, the *nuisance neglect* that Graeber (2022) identifies – that is, the tendency for individuals to ignore payoff-irrelevant variables when making inferences, can be generated by the decisiveness criterion.

Furthermore, the decisiveness criterion makes a sharp prediction over when nuisance neglect will tend to occur: payoff-irrelevant variables will tend to be neglected in inference when accounting for them results in greater uncertainty over the prediction target. As such, the criterion predicts that the DM may adopt models that account for payoff-irrelevant variables if doing so increases her certainty over the prediction target: for instance, in the example considered in Section 3.3.1, the DM accounts for the trustworthiness of the news sources — a variable that is not directly payoff-relevant — because incorporating this auxiliary variable helps her draw stronger conclusions from the data. In contrast, an account based solely on the premise that the DM directs greater attention to payoff-relevant states would predict neglect of such auxiliary variables regardless of their implications for inference. ▲

Contrast the predictions of the decisiveness criterion in the above example to those made by models of rational inattention (see Maćkowiak et al. (2021) for a review) – a class of models that predicts that individuals may neglect variables when they are not decision-relevant due to attentional costs. Note that standard models of rational inattention cannot account for the predictions above: to the extent that correctly accounting for otherwise payoff-irrelevant variables helps individuals form more accurate inferences about the prediction target, such models would not predict a systematic neglect of such variables. Moreover, while models of subjectively rational inattention (Schwartzstein, 2014; Gagnon-Bartsch et al., 2021) demonstrate how attentional costs can prevent individuals who hold misspecified models that neglect the role of certain variables from noticing their error in the face of feedback — so long as such those variables are not directly payoff-relevant — such models do not explain why individuals come to adopt such erroneous models in the first place, nor how the adoption of these models can itself be determined by which variables are payoff-relevant.

5 Choice Under the Decisiveness Criterion

In this section, I study the implications of model selection under the decisiveness criterion on choice. Fixing a signal realization s: given a utility function u and a set of models M, let $\mathcal{C}(A)$ denote the actions recommended by the decisiveness-maximizing models in M:

$$\mathcal{C}(A) = \bigcup_{m \in \arg\max_{m \in M} I_{\mathcal{D}}(m|s)} \arg\max_{a \in A} \sum_{\theta} u(a, \theta) p^m(\theta|s)$$

in which case we say that $\mathcal{C}$ is *represented* by (u, M) given signal s. I relegate discussion of a full behavioral characterization of $\mathcal{C}$ and its identification properties to Appendix A.6, and focus on a key property of $\mathcal{C}$ – an aversion to hedging or diversification, and also study a key comparative static: that if the DM entertains a larger set of models, the DM will be more averse to diversification.

5.1 Attitudes Towards Diversification

Recall that a key property characterizing model selection under the informativeness criterion is a preference for extreme models, as summarized by Propositions 1 and 2. An immediate consequence of the DM's preference for extreme models is that the DM exhibits a preference for extreme actions, or an aversion to hedging.

Proposition 5. For $a, a' \in A$, let a'' satisfy $u(a'', \theta) = \lambda u(a, \theta) + (1 - \lambda) u(a', \theta)$ for some $\lambda \in (0, 1)$. If $a'' \in \mathcal{C}(A \cup \{a''\})$, then either $a \in \mathcal{C}(A \cup \{a''\})$ or $a' \in \mathcal{C}(A \cup \{a''\})$.

Proposition 5 states that the DM exhibits a form of *mixture aversion*: she cannot strictly prefer a hedge between two actions to both of those actions[12]. That is, choice under the decisiveness criterion is averse to a specific form of diversification aversion corresponding to mixtures of acts in utility space.

[12] In Appendix A.6, I show that mixture aversion is a key axiom in the behavioral characterization of choice under the decisiveness criterion

In addition to predicting this form of diversification aversion, the decisiveness criterion produces the following comparative static on $\mathcal{C}$: that if the DM entertains a larger set of models, the DM will be more averse to a different notion of diversification, which I define below.

For the remainder of this section, consider a family of $\mathcal{C}$ represented by the same utility u. For a menu A, let $\overline{u}^A(\theta) = \max_{a \in A} u(a, \theta)$ for all θ — that is, $\overline{u}^A(\theta)$ gives the maximal payoff in each state among actions in A. We will now introduce an order on $\mathcal{C}$ that corresponds to a level of aversion to a specific notion of diversification. Note that we should expect any notion of diversification to stipulate that a diversified action does not deliver payoffs that exceed the maximal payoff profile of the menu, $\overline{u}^A(\theta)$. The following definition is more restrictive, in that it requires that the action delivers payoffs that differ from $\overline{u}^A(\theta)$ by a constant.

Definition (C-Diversified Action). Say that a is a *C-diversified action* with respect to A if there exists $k > 0$ such that $u(a, \theta) = \overline{u}^A(\theta) - k$ for all θ. Let H_A collect acts that are C-diversified with respect to A.

Definition (Relative Diversification Aversion). Say that $\mathcal{C}'$ is *more diversification-averse* than $\mathcal{C}$ if for any menu A, $a \in H_A$, $a \notin \mathcal{C}(A \cup \{a\}) \implies a \notin \mathcal{C}'(A \cup \{a\})$.

That is, $\mathcal{C}'$ is more diversification-averse than $\mathcal{C}$ if $\mathcal{C}'$ never chooses a C-diversified action from a menu whenever $\mathcal{C}$ does not. As the proposition below demonstrates, an appropriately defined set inclusion order on the model-implied posteriors the DM entertains characterizes this notion of relative diversification aversion. For a set of models M, let $P_{M|s} = \{p^m(\cdot|s) : m \in M\}$ denote the corresponding set of model-implied posteriors given the signal realization. Say that $P_{M|s}$ is *interior* if no model-implied posterior in $P_{M|s}$ induces certainty in a state.

Proposition 6. Suppose $\mathcal{C}, \mathcal{C}'$ are represented by (u, M), (u, M'), respectively, given signal s. If $co(P_{M|s}) \subseteq co(P_{M'|s})$, then $\mathcal{C}'$ is more diversification-averse than $\mathcal{C}$. Furthermore, if $\mathcal{C}'$ is more diversification-averse than $\mathcal{C}$ and $P_{M'|s}$ is interior, then $co(P_{M|s}) \subseteq co(P_{M'|s})$.

Proposition 6 is a natural consequence of Proposition 1, which states that under the decisiveness criterion, the DM selects extreme models. If the the set of models the DM entertains expands, she will select yet more extreme models and as a result, chose yet more extreme actions. Proposition 6 also states that the converse is true, so long as the DM does not entertain models that eliminate all residual uncertainty. One immediate consequence of Proposition 6 is a comparative static on the DM's valuation of costly but perfectly revealing information.

Example (Value of Information). Consider a DM who chooses to acquire costly but perfectly revealing information about the state, whose utility is quasilinear and separable in money. Consider the action space $Z \times W$, where Z corresponds to a set of state-contingent prizes and $W \subset \mathbb{R}$ corresponds to money. Suppose that the DM's utility is quasilinear and separable in money: for any action (z, w), $u((z, w), \theta) = v(z, \theta) + w$.

For any menu A, and cost of information $\kappa < 0$ let a^A_κ denote the action that corresponds to the continuation payoff of acquiring perfectly revealing information about the state at monetary cost. Given the quasilinearity assumption, we have $u(a^A_\kappa, \theta) = \overline{u}^A(\theta) - \kappa$: by learning the state prior to choosing from A, the DM achieves the maximal utility achievable in each state, less the cost of the information. Note that a^A_κ delivers precisely the payoffs associated with a C-diversified action. Proposition 6 then implies that if the DM entertains a larger set of models, the DM will have a lower valuation for the information: formally, for $\mathcal{C}, \mathcal{C}'$ represented by $(u, M), (u, M')$, respectively, if $co(P_M) \subseteq co(P_{M'})$, $a^A_\kappa \notin \mathcal{C}'(A \cup \{a^A_\kappa\}) \implies a^A_\kappa \notin \mathcal{C}(A \cup \{a^A_\kappa\})$, and furthermore if $P_{M'}$ is interior, the converse holds. Intuitively, if the DM entertains a larger set of models, she will tend to hold more extreme beliefs due to Proposition 1; this in turn reduces the value of information. ▲

Although Proposition 6 only characterizes a notion of relative diversification aversion relating to C-diversified action, the result has implications for more general notions of diversification. In particular, consider a less restrictive notion of a diversified action that only stipulates that such an act deliver payoffs that do not exceed the maximal payoff profile of a menu:

Definition (Diversified Action). Say that a is a *diversified action* with respect to A if $u(a, \theta) \leq u(a^A_{max}, \theta)$ for all θ. Say that a is a *strictly diversified action* with respect to A if $u(a, \theta) < u(a^A_{max}, \theta)$ for all θ.

A consequence of Proposition 6 is that so long as the set of models the DM entertains includes sufficiently extreme (but interior) models, the DM will never choose a strictly diversified action.

Corollary 5. If a is strictly diversified with respect to A, then there exists a set of interior models $\overline{M}$ such that for any $M \supseteq \overline{M}$ and $\mathcal{C}$ represented by (u, M) given signal s, $a \notin \mathcal{C}(A \cup \{a\})$.

Example (Underdiversification). A robust finding is that households tend to hold underdiversified portfolios — that is, they hold fewer securities than are needed to eliminate idiosyncratic risk (see, e.g., Blume and Friend 1975, Kelly 1995, Odean 1999, Vissing-Jorgensen 1999, Polkovnichenko 2005, and Goetzmann and Kumar 2004). A common assumption in modeling under-diversification is that the return distributions of securities are known, and that under-diversification results from non-standard preferences such as cumulative prospect theory (Barberis and Huang 2005) or skewness preferences (Mitton and Vorkink 2007). My framework highlights an additional force that can generate under-diversification: when investors face uncertainty over the return distributions of securities, they favor models that pick out "winners" over models that recommend diversification, as diversification is necessarily likely to be ex-post suboptimal for any realization of returns.

To take a stylized example, consider a DM who chooses how much to allocate between two securities $i = 1, 2$, which deliver monetary payoffs in $\mathbb{R}$. Suppose the DM has utility over money $v : \mathbb{R} \to \mathbb{R}$; assume only that v is strictly increasing in money. Let Θ describe the possible joint payoff realizations of two securities, and let $r_{i,\theta}$ denote the monetary payoff that security i delivers in state θ. Let a_i denote the action corresponding to allocating the full portfolio share to security i, and for $\alpha \in (0,1)$, let a_α denote the action that corresponds to allocating portfolio weight α to security 1; we have $u(a_i,\theta) = v(r_{i,\theta})$ for $i = 1,2$, and $u(a_\alpha,\theta) = v(\alpha r_{1,\theta} + (1-\alpha)r_{2,\theta})$. Here, I impose one assumption on the structure of payoffs: that for each θ, $f_1(\theta) \neq f_2(\theta)$ — that is, with probability 1, the two securities deliver different returns.

Note that under these assumptions, a_α is strictly diversified with respect to $A = \{a_1, a_2\}$. Corollary 5 then implies that so long the DM entertains a sufficiently extreme set of models, she will choose either a_1 or a_2 over a_α.

What characterizes a "sufficiently extreme" model? Here, I give a sufficient condition. Let Θ_1 denote the states where security 1 delivers higher payoffs than security 2, and let Θ_2 denote the states where security 2 delivers higher payoffs. As the proof of Corollary 5 shows, there exists $\epsilon > 0$ such that if $p^{m^*}(\Theta_1|s) > 1-\epsilon$, then for any M containing m^*, the DM will not choose the diversified action. Intuitively, a model that picks out a security as a winner, such as the m^* defined above, will in general be decisive: such a model necessarily induces a belief that investing in that security will likely be ex-post optimal. On the other hand, models that recommend diversification are in general not decisive, since diversification will necessarily be ex-post suboptimal. ▲

5.2 Attitudes Towards Delay of Decision-Making

Consider again a family of $\mathcal{C}$ represented by the same utility u, and fix a menu A. Associate each C-diversified action $a \in H_A$ with $k_a \equiv \overline{u}^A(\theta) - u(a,\theta)$, the utility difference between a and the maximal payoff profile.[13]

Definition (Value of Delay). For a set of models M and signal s, let the *value of delay* $K_A(M|s)$ denote the minimum k_a such that $a \in \mathcal{C}(A \cup \{a\})$ for $\mathcal{C}$ represented by (u, M), given signal s.

In words, the value of delay $K_A(M|s)$ is the highest amount the DM would be willing to pay to learn the state before choosing her action. The following lemma demonstrates that the value of delay is a behavioral measure of the DM's residual uncertainty under the model she selects from M.

Lemma 1. For $\mathcal{D} = (u, A)$, $K_A(M|s) = \min_{m\in M} R_{\mathcal{D}}(p^m(\cdot|s))$.

This section will contrast $K_A(M|s)$, the value of delay of a decisiveness-maximizing DM who entertains models M, from $K_A(\{m_T\}|s)$, the value of delay for a Bayesian — that is, a DM who chooses according to the true model.

[13]Recall that by definition, $\overline{u}^A(\theta) - u(a,\theta)$ is constant in θ for all $a \in H_A$.

An immediate consequence of Lemma 1 is that so long as $m_T \in M$, $K_A(M|s) \leq K_A(\{m_T\}|s)$: the decisiveness-maximizing DM will have a lower value of delay than a Bayesian, so long as she entertains the true model. This may be at odds with evidence suggesting that individuals appear to excessively delay decision-making, such as in work on choice overload (Chernev et al., 2015). Note however, that while a decisiveness-maximizing DM places too low a value on delay in an absolute sense, her value of delay may be less responsive to *changes* in information that would reduce the value of delay for a Bayesian, as the following example illustrates.

Example (Value of Delay: Hiring Example). Consider motivating example in the introduction. The DM has uniform priors over the productivity of the candidate $\{\theta_h, \theta_l\}$, faces the decision problem

$$\begin{array}{lcc} & \theta_h & \theta_l \\ \textit{hire} & v & -k \\ \textit{reject} & 0 & 0 \end{array}$$

and observes the signal $(s^R, s^I) \in \{0,1\} \times \{0,1\}$. Suppose $M = \{m_R, m_I, m_{RI}\}$ characterized by the likelihood functions

$$L_{m_R}(s^R, s^I) = \begin{cases} 4 & s^R = 1 \\ 1/4 & s^R = 0 \end{cases}$$

$$L_{m_I}(s^R, s^I) = \begin{cases} 4 & s^I = 1 \\ 1/4 & s^I = 0 \end{cases}$$

$$L_{m_{RI}}(s^R, s^I) = \begin{cases} 4 & s^I = s^R = 1 \\ 1 & s^I \neq s^R \\ 1/4 & s^I = s^R = 0 \end{cases}$$

where $L_m(s) = \frac{m(s|\theta_h)}{m(s|\theta_l)}$. Consider the growth regime from the motivating example, where $v = 4, k = 1$, and compare the value of delay for two signal realizations $s = (1,0)$ and $s' = (1,1)$; under s, the DM receives mixed signals about the candidate's productivity, whereas whereas under s', the DM receives aligned signals indicating that the candidate is high productivity.

Note that $K_A(\{m_T\}|s) > K_A(\{m_T\}|s')$; a Bayesian has a lower value of delay after receiving aligned signals, as opposed to receiving mixed signals. Note, however, that since the decisiveness-maximizing DM adopts the model m_R under s, thereby focusing only on the positive component of the mixed signal s, the posterior of the decisiveness-maximizing DM is identical for both s and s', and so $K_A(\{m_T\}|s) = K_A(\{m_T\}|s')$. In other words, whereas the Bayesian's value of delay decreases from observing additional information in favor of hiring the candidate, the decisiveness-maximizing DM's value of delay does not. Intuitively, the decisiveness-maximizing DM was already operating according to a model that reduced the indecisiveness of the mixed signal s, which blunts the response to receiving an aligned signal s'. Appendix A.4 demonstrates how this logic generalizes. ▲

6 Applications

I now discuss two additional applications of the decisiveness criterion. Whereas the preceding analysis took as primitive M, the set of models the DM entertains, in each of these applications I extend the framework by considering two distinct forces that may shape M: the supply of models by expert advisors, and the social exchange of models.

6.1 Certainty in Expert Advice

Evidence suggests that individuals are drawn to advisors who provide more certain advice. As Kahneman (2011) writes,

> Experts who acknowledge the full extent of their ignorance may expect to be replaced by more confident competitors, who are better able to gain the trust of clients. An unbiased appreciation of uncertainty is a cornerstone of rationality—but it is not what people and organizations want. (page 263)

Consistent with this view, research in psychology has documented that individuals have a more favorable view toward advisors who make more certain forecasts as opposed to moderate ones. In a stark demonstration, Price and Stone (2004) provide subjects with probabilistic forecasts of a series of events (the likelihood that the price of a stock will increase) made by two advisors, coupled with the realization of each event; the forecasts of one advisor are designed to be more extreme (closer to 0% or 100%) but poorly calibrated compared to those of the other advisor[14]. The majority of subjects prefer the extreme advisor, despite the presence of outcome data indicating the superior accuracy of the moderate advisor. Studies using related experimental paradigms corroborate this finding (Yates et al., 1996; Gaertig and Simmons, 2018).

An immediate consequence of the decisiveness criterion is that individuals will be drawn to more certain advice. To see this, consider a setting in which the DM entertains models proposed by a set of advisors. Proposition 1 states that the DM will tend to adopt extreme models; in a binary state case, this amounts to the DM adopting models that place the likelihood of the event close to 0% or 100%. As Kahneman (2011) points out, this tendency for individuals to adopt certain advice can lead to a proliferation of overly certain advice in a competitive market for advisors, even if advisors do not inherently have incentives to provide biased advice. Formally, consider a binary state setting with $\Theta = \{\theta_h, \theta_l\}$. A receiver and a set of senders (advisors) share a common prior over the state and observe a common signal realization s. Each sender i proposes a single model m_i to the receiver, who adopts the most decisive model from $M = \bigcup_i m_i$ given her decision problem $\mathcal{D}$ (in case of a tie, suppose the receiver adopts a random model among the decisiveness-maximizing models in M). Each

[14]In particular, the forecasts of the moderate advisor are constructed to be accurately calibrated, and the forecasts of the extreme advisor are constructed by adding 15% to each forecast made by the moderate advisor greater than 50%, and vice versa subtracting 15% from each forecast less than 50%.

sender i knows the true model m_T, and their payoffs are given by

$$u_i(m_i) = \begin{cases} v - (p^{m_i} - p^{m_T})^2 & \text{if } m_i \text{ is adopted} \\ 0 & \text{otherwise.} \end{cases}$$

where $p^m \equiv p^m(\theta_h|s)$ denotes the posterior induced by model m. Senders earn a payoff if their model is adopted, and conditional on this have incentives to provide accurate advice. First, note that in absence of competition between senders, the sender maximizes utility by proposing the true model m_T, which in turn maximizes the utility of the receiver, who adopts the true model. In what follows, consider the (pure strategy) Nash equilibrium of the game in which senders simultaneously propose models. Let M^* denote the model(s) that the receiver adopts in such an equilibrium.

Proposition 7. Suppose that $R_{\mathcal{D}}(p)$ is non-constant in the neighborhood of $p = p^{m_T}$. In any Nash equilibrium with at least two senders, for any $m^* \in M^*$,the receiver's equilibrium posterior belief p^{m^*} must be either $\max\{0, p^{m_T} - \sqrt{v}\}$ or $\min\{1, p^{m_T} + \sqrt{v}\}$.

Proposition 7 states that at an interior equilibrium posterior belief, the model adopted by the sender must satisfy a "zero-profit" condition – it is the most extreme model the receiver can propose that still yields weakly positive payoffs. Competition causes receivers to ratchet up the extremeness of their proposed models, even though receivers have no inherent incentives to push biased models.[15]

Consider further the special case in which $0 < p^{m_T} - \sqrt{v}$, $p^{m_T} + \sqrt{v} < 1$, and in which the receiver's decision problem $\mathcal{D}$ is *symmetric*: that is, $R_D(p)$ is symmetric around $p = 1/2$. In this case, it can be shown that if $p^{m_T} \neq 1/2$, the equilibrium posterior belief is unique, and is given by

$$p^{m^*} = \begin{cases} p^{m_T} + \sqrt{v} & p^{m_T} > 1/2 \\ p^{m_T} - \sqrt{v} & p^{m_T} < 1/2 \end{cases}$$

In this sense, the models the senders propose in equilibrium are an exaggerated version of the truth: if the sender believes that θ_h is more likely than not to occur, she will propose a model that exaggerates the likelihood of θ_h, and vice versa if she believes that θ_l is more likely than not to occur.

6.2 Shared Models and Group Polarization

As discussed in Sections 3.2 and 4.2, the decisiveness criterion can generate belief polarization in response to information — that is, receiving identical information can cause the opinions of individuals to diverge along differences in prior beliefs and/or objectives. In this section, I explore another source of polarization: the exchange of models, or interpretations

[15]The condition that $R_{\mathcal{D}}(p)$ is non-constant around p^{m_T} rules out the case in which the receiver finds any model that would deliver weakly positive profits to the sender to be equally decisive.

of the data, within a group.

Consider evidence of so-called *group polarization* — the phenomena in which attitudes following group discussion are more extreme than the attitudes held by members of the group prior to discussion. In one such experiment, Schkade et al. (2000) provide. subjects with identical case information and ask subjects to individually rate the severity of punishment appropriate for the defendant. Mock juries are then formed and given the task of deliberating as a group on the punishment severity. The key finding is that group deliberation increases the extremity of punishment ratings: for juries in which individual punishment ratings were high, deliberation tends to increase the group rating relative to the median individual rating within the jury, and vise versa when individual punishment ratings were low. Similar evidence of group polarization has been documented in group judgements along a variety of other dimensions, such as the appropriate level of risk-taking (Myers and Lamm, 1976) or the degree of racial prejudice (Myers and Bishop, 1976).[16]

What can account for these effects of group deliberation? One possibility is that deliberation allows for the aggregation of private information, which can lead to group polarization (Roux and Sobel, 2015). One tension with this explanation, as Roux and Sobel (2015) note, is that subjects are given identical information from which to form their judgements in the standard experimental paradigm used to study group polarization. Glaeser and Sunstein (2009) echo this critique, noting evidence for large shifts in beliefs due to group polarization in settings in which individuals likely have little new knowledge or information to bring to the table, such as debates surrounding climate change or affirmative action. How can group polarization arise in settings where individuals lack private information? Glaeser and Sunstein (2009) analyze a model of "credulous Bayesians" in which individuals overstate the informational content of others' beliefs — in their model, initial heterogeneity in opinions does not reflect differences in private information but rather noise, and group polarization arises because individuals misattribute this noise to private information. I propose an alternative account of group polarization, in which initial heterogeneity in opinions does not reflect differences in private information but rather in individuals' *interpretations* of public information — that is, individuals entertain different models — and that group polarization is driven by the exchange of models.

The decisiveness criteria provides a simple intuition for why exchanging models can result in more extreme judgements: extreme models tend to be decisive. Here, I adopt the formal framework introduced in Schwartzstein and Sunderam (2022) to study the social exchange of models. Consider a setting with binary states $\Theta = \{\theta_h, \theta_l\}$ (e.g. whether the defendant committed a serious or a mild offense) over which a group of individuals share a common prior ρ, a common decision problem $\mathcal{D}$ (e.g. deciding the severity of punishment to inflict on the defendant), and observe a public signal s. Suppose that each individual i initially entertains a single model m_i with which to interpret the data, and after group delibera-

[16]Importantly, group deliberation has been found to polarize not only group judgements in these settings but also the private judgements of individuals in the group, suggesting that group polarization is not purely driven by distortions caused by the group decision-making process such a social desirability or the diffusion of responsibility.

tion is exposed to the models entertained by the group, entertaining $M_g \equiv \bigcup_i m_i$.[17] Let p^{m_i} denote individual i's posterior belief before group deliberation, where for any model m, $p^m \equiv p^m(\theta_h|s)$.

Consider the effects of deliberation on beliefs. First, note that group deliberation expands the set of models that each individual entertains (from $\{m_i\}$ to M_G); this leads to the adoption of more extreme models (Proposition 1), and in turn, more extreme actions (Proposition 6). Second, under the assumption that the group shares a common prior and decision problem, group deliberation causes the beliefs of individual in the group to converge, as each adopts the same model from M_g after deliberation. That is, exchanging models leads individuals in a group to adopt the extreme interpretations of the data found within the group.

In this manner, the exchange of models can amplify initial differences in individual judgements between groups, generating group polarization. To see this, suppose that the decision problem $\mathcal{D}$ is symmetric, and let $\underline{p} = \min_{m \in M_G} p^m$, $\overline{p} = \max_{m \in M_G} p^m$ denote the extreme viewpoints held by members of the group prior to deliberation. Let p^* denote the viewpoint held by the group after deliberation. We have

$$p^* = \begin{cases} \overline{p} & \overline{p} + \underline{p} > 1 \\ \underline{p} & \overline{p} + \underline{p} < 1 \end{cases}$$

That is, if the initial viewpoints of a group are skewed in favor of θ_h, the exchange of models within that group will cause the viewpoints of the the group to be increasingly skewed in favor of θ_h, and vice versa if initial viewpoints are skewed towards θ_l.[18]

Note that a key prediction of this account of group polarization is that following group deliberation, individuals arrive at beliefs that are not only more extreme, but that also provide greater certainty over the optimal course of action. Importantly, this latter prediction need not hold under an account based on Bayesian information aggregation. Under the Bayesian account, group polarization towards a certain conclusion is purely the consequence of individuals obtaining a greater balance of evidence in favor of that conclusion after deliberation, and in particular occurs regardless of the action implications of that conclusion. Therefore, as demonstrated in Appendix A.3, in certain situations the Bayesian account predicts that group deliberation will push individuals to hold *greater* uncertainty over the optimal course of action. This outlines a key distinction between the predictions of the two accounts: under a Bayesian account, deliberation can cause individuals to become increasingly convinced in the conclusion that the case facts are inconclusive over whether the defendant should be convicted or acquitted, or that there is insufficient evidence to determine whether or not

[17] The assumption that each individual entertains only a single model prior to group discussion is not central to the analysis. Without affecting the analysis, one can instead assume that each individual entertains a set of models M_i and that each individual shares M_i with the group, letting m_i denote the model that individual i finds most decisive from M_i.

[18] Analogous results hold if we relax the assumption that $\mathcal{D}$ is symmetric. In the general case, there exists a $p' \in (0,1)$ and increasing functions f_1, f_2, such that $p^* = \overline{p}$ if $f_1(\overline{p} - p') > f_2(p' - \underline{p})$ and $p^* = \underline{p}$ if $f_1(\overline{p} - p') < f_2(p' - \underline{p})$

a certain climate protection policy should be adopted, whereas under the present account, group deliberation systematically pushes individuals away from such conclusions.

7 Extensions

7.1 Plausibility Constraints on Model Selection

As discussed in Section 3, the decisiveness criterion favors extreme models, which induce extreme beliefs. At the same time, however, it is not the case that decision-makers hold extreme beliefs in every decision context – empirically, the models individuals adopt and their resulting conclusions do seem to be constrained by some plausibility criteria. This section discusses two plausibility-based constraints on model choice; the first imposes a "cost" to deviating from a default belief, while the second imposes an "entry condition" based on the fit of a default model.

7.1.1 Cost of Deviating from Default Beliefs

Here, I develop a constraint on model selection that incorporates a cost to selecting models that induce deviations from default beliefs, following Hansen and Sargent (2001) and Caplin and Leahy (2019). Let $p_d \in \Delta(\Theta)$ denote the DM's default belief over outcomes. Some candidate values of p_d are the DM's prior, the posterior belief induced by a default model, or the posterior beliefs of a Bayesian observer. The DM adopts a model m to maximize decisiveness, subject to an additional cost of deviating from the default belief, given by the relative entropy from p_d to model-implied posteriors $p^m(\cdot|s)$:

$$\frac{1}{\gamma}\sum_{\theta\in\Theta} p^m(\theta|s)\ln\frac{p^m(\theta|s)}{p_d(\theta)}$$

where γ governs the weight of this cost relative to the regret criterion. As before, given the DM's choice of model, the DM selects an action that maximizes model-implied posterior expected utility. Denoting $\overline{u}^A(\theta) = \max_{a\in A} u(a,\theta)$ as the maximum payoff achievable for state θ, the DM's problem can be written as

$$V = \max_{a\in A, m\in M} \sum_{\theta\in\Theta} \left(u(a,\theta) - \overline{u}^A(\theta)\right) p^m(\theta|s) - \frac{1}{\gamma}\sum_{\theta\in\Theta} p^m(\theta|s)\ln\frac{p^m(\theta|s)}{p_d(\theta)}$$

Full Model Space. Consider the case where the DM's entertains the full set of models – here, the DM can implement any posterior belief by adopting some model. In this case, the DMs problem can be rewritten as

$$V = \max_{a\in A, p\in\Delta(\Theta)} \sum_{\theta\in\Theta} \left(u(a,\theta) - \overline{u}^A(\theta)\right) p(\theta) - \frac{1}{\gamma}\sum_{\theta\in\Theta} p(\theta)\ln\frac{p(\theta)}{p_d(\theta)}$$

Note that the similarity between this formulation and that of Caplin and Leahy (2019): it is identical to their model of wishful thinking, except the DM's utility function is replaced by a *normalized* utility function $v(a,\theta) = u(a,\theta) - \overline{u}^A(\theta)$, which normalizes the payoff from taking an action in state θ by the maximal utility attainable in θ. The model is as if the agent seeks to maximize anticipatory normalized utility, given a cost of deviating from the default belief.

Fixing an action a, first order conditions for p imply that an interior solution p must satisfy

$$p(\theta) = \frac{\exp\left[\gamma(u(a,\theta) - \overline{u}^A(\theta))\right] p_d(\theta)}{\sum_{\theta' \in \Theta} \exp\left[\gamma(u(a,\theta') - \overline{u}^A(\theta'))\right] p_d(\theta')}$$

This expression tells us that the DM's posterior will be a "tilting" of their default posterior that puts less weight on states associated with higher regret given the chosen action a. Substituting the above expression into DM's problem yields

$$V = \max_{a \in A} \frac{1}{\gamma} \ln \left(\sum_{\theta \in \Theta} p_d(\theta) \frac{\exp(\gamma u(a,\theta))}{\exp\left(\gamma \overline{u}^A(\theta)\right)} \right)$$

giving a description of the DM's choice. To interpret this expression, note that the return to taking an action that pays off in state θ is increasing in $p_d(\theta)$, the DM's default belief in θ, but also is also decreasing in $\overline{u}^A(\theta)$, the maximal payoff achievable in state θ. The intuition for the latter effect is as follows: all else equal, as the maximal payoff associated with state θ increases, the residual uncertainty associated with that state increases, and so the DM will tend to select models that induce lower posteriors over θ. Note that this comparative static is related that of Proposition 3.

Example (Fund Manager). To illustrate this modified criterion, consider the following example. The DM wants to learn whether a fund manager is high-skilled (θ_h) or low-skilled (θ_l), and initially believes either possibility is equally likely: $\rho(\theta_h) = 1/2$. The DM's signal is the return of the manager's fund, which is either high or low: $\{r_h, r_l\}$. The DM's decision problem is given by

	θ_h	θ_l
hire	v	$-k$
don't hire	0	0

Suppose that $k > v$, so that given the DM's prior belief, the DM chooses not to hire the manager. Suppose further that the DM's default belief is $p_d(\theta_h) = 1/2$, which corresponds to a belief that returns are uninformative about manager skill. Consider the case that a low return is realized. Consider two cases:

Case 1: Full Model Space.
Suppose that the DM entertains the full space of models. Applying the expression above, the optimal model-implied posterior will be

$$p^m(\theta_h) = \frac{1}{1 + \exp(\gamma v)}$$

Note that this model-implied posterior $p^m(\theta_h)$ is weakly lower than 1/2, and for $\gamma \to \infty$, $p^m(\theta_h) \to 0$, corresponding to the case where the DM faces no costs of deviating from the default belief, whereas for $\gamma \to 0$, $p^m(\theta_h) \to 1/2$, the default belief.

Case 2: Discrete Model Space.
Suppose the DM entertains two models:

$$m_1 : m_1(r_l|\theta_l) = \phi,\ m_1(r_l|\theta_h) = \phi$$
$$m_2 : m_2(r_l|\theta_l) = q,\ m_2(r_l|\theta_h) = (1-q),\ q > 1/2$$

that is, under m_1, returns are uninformative about skill, whereas under m_2, returns are decisive. Given the DM's problem, the DM selects m_2 over m_1 iff

$$f(q) \equiv v(q - 1/2) - \frac{1}{\gamma}(q \ln q + (1-q)\ln(1-q) - \ln 1/2) \geq 0$$

Note that $f(1/2) = 0$, $f'(1/2) > 0$,and $f''(q) < 0$, which implies that for some range of q above 1/2, the DM will select the informative model m_2, but for q too large, the DM will instead select m_1. Intuitively, the DM finds m_2 decisive because it provides evidence in favor not hiring the manager, the action the DM is predisposed toward. At the same time, the DM will not select m_2 if it induces too extreme of a posterior, due to the costs of deviating from her dfefault belief. ▲

7.1.2 Entry Condition on Models

Here, I develop a constraint on model selection operating through an entry condition on the set of models the DM entertains. Let $m_d \in \mathcal{M}$ denote the DM's default model. Following Schwartzstein and Sunderam (2021), define the fit of a model given realized signal s, as

$$P(m|s) = \sum_{\theta \in \Theta} m(s|\theta)\rho(\theta)$$

The DM adopts a model that minimizes regret, subject to an entry condition that the chosen model must deliver weakly higher fit than the default model. The DM's problem is then

$$\min_{m \in M \cup \{m_d\}} \sum_{\theta \in \Theta} (u(a,\theta) - \overline{u}^A(\theta))p^m(\theta|s) \quad \text{s.t. } P(m|s) \geq P(m_d|s)$$

Proposition 1 in Schwartzstein and Sunderam (2021) shows that the above entry condition is equivalent to the following constraint:

$$P(m|s) \geq P(m_d|s) \iff p^m(\theta|s) \leq \frac{\rho(\theta)}{P(m_d|s)} \forall \theta$$

that is, the entry condition bounds the extent to which models are allowed to move the DM's beliefs away from the prior, where the bounds are tighter the better the fit of the default model.

Let $P_{m_d} = \left\{ p \in \Delta(\Theta), p(\theta) \leq \frac{\rho(\theta)}{P(m_d|s)} \forall \theta \right\}$ denote the set of feasible posteriors given the entry condition, and let $M_{m_d} = \{m \in M : p^m \in P_{m_d}\}$ denote the set of models in $M \cup \{m_d\}$ that survive the entry condition. The DM's model selection problem can be rewritten as

$$\min_{m \in M_{m_d}} \sum_{\theta \in \Theta} (u(a,\theta) - \overline{u}^A(\theta)) p^m(\theta|s)$$

Note that the results from Section 3 and 4 have analogs that continue to hold in this extension, once the entry condition on models is accounted for.

Example (Fund Manager Revisited). Consider again the fund manager setting. The DM wants to learn whether the fund manager is high-skilled (θ_h) or low-skilled (θ_l), and holds priors $\rho(\theta_h) = \rho$. The DM's signal is the return of the manager's fund, which is either high or low: $\{r_h, r_l\}$. The DM's decision problem is again given by

	θ_h	θ_l
hire	v	$-k$
don't hire	0	0

where that $k > v$, so that given the DM's prior belief, the DM chooses not to hire the manager. Suppose that the DM's default model m_d satisfies $m_d(r_l|\theta_l) = m_d(r_l|\theta_h) = \phi$. Suppose again that a low return is realized.

Which models survive the entry condition? The inequality above implies that only models that satisfy $p^m(\theta_h) \in [1 - (1-\rho)/\phi, \rho/\phi]$ will have greater fit than the default model. In particular, consider the case where $\phi = 1$: here, the default model states that the realized signal of a low return was inevitable, regardless of the skill of the manager. in this case, the only model that survives the entry condition is the default model itself. On the other hand, as $\phi \to 0$ — that is, as the realized signal becoems increasingly unlikely under the default model — the set of surviving models expands to the entire model space. ▲

7.2 Ex-Ante Decisiveness

Recall that in the basic framework, model selection occurs ex-post: after the signal is realized, the DM evaluates each model based on its decisiveness, given the signal realization, and adopts the most decisive model. One might instead imagine an account in which the DM evaluates each model according to an *ex-ante* notion of decisiveness, and adopts a model prior to the signal realization. In this section, I provide a formulation of an ex-ante notion of decisiveness, and compare its properties to the ex-post formulation.

Consider the following formulation of ex-ante decisiveness. As in the basic framework, the DM entertains a set of models M and faces a decision problem $\mathcal{D}$. Letting $p^m(s) = \sum_\theta m(s|\theta)\rho(\theta)$ denote the likelihood of signal s under model m, let

$$I^E_{\mathcal{D}}(m) = \sum_s I_{\mathcal{D}}(m|s) p^m(s)$$

denote the expected decisiveness of m prior to the signal realization. Consider a DM who, prior to the realization of the signal, selects a model from M that maximizes expected decisiveness; refer to this account of model selection as the *ex-ante* case and the account developed in Section 2 as the *ex-post* case. In what follows, I will outline which of the core results of the ex-post case have analogs in the ex-ante case, relegating formal details to Appendix A.5.

7.2.1 Properties of Model Selection under Ex-Ante Decisiveness

Extremeness. Recall that a key-property of the ex-post criterion is that it favors extreme models: in particular, as Proposition 1 implies, if $m = \lambda m' + (1-\lambda)m''$, then for any s, $I_{\mathcal{D}}(m|s) \leq \max\{I_{\mathcal{D}}(m'|s), I_{\mathcal{D}}(m''|s)\}$. The ex-ante criterion satisfies an analogous property: if $m = \lambda m' + (1-\lambda)m''$, then $I^E_{\mathcal{D}}(m) \leq \max\{I^E_{\mathcal{D}}(m'), I^E_{\mathcal{D}}(m'')\}$. The intuition for this extremeness result is similar to that of the ex-ante case: the composite model m'' corresponds to a case where the DM is uncertain over how to interpret the information, which results in lower average decisiveness.

Overprecision and Confirmation Bias. Recall that in the ex-post case, the extremeness property generates a form of overprecision. In particular, letting $m_\varnothing$ denote an uninformative model, Corollary 1 states that if $I_{\mathcal{D}}(m_\varnothing|s) < I_{\mathcal{D}}(m|s)$, then for any m' satisfying $m = \lambda m' + (1-\lambda)m_\varnothing$, it must be that $I_{\mathcal{D}}(m'|s) \geq I_{\mathcal{D}}(m|s)$. That is, under the ex-post criterion, if the DM finds a model m more decisive than her prior for a given signal realization, she will find a model that overstates the informativeness of m yet more decisive. The ex-ante criterion also generates overprecision, but does so *unconditionally*: for any m' satisfying $m = \lambda m' + (1-\lambda)m_\varnothing$, we have $I^E_{\mathcal{D}}(m') \geq I^E_{\mathcal{D}}(m)$: that is, the DM exhibits overprecision irrespective her prior. In fact, the ex-ante criterion produces a more general form of overprecision: if m' dominates m in the Blackwell order, then $I^E_{\mathcal{D}}(m') \geq I^E_{\mathcal{D}}(m)$ — that is, the ex-ante criterion respects the Blackwell order.[19]

An immediate consequence of the fact that overprecision holds unconditionally in the ex-ante case is that unlike ex-post decisiveness, model selection under ex-ante decisiveness does not generate confirmation bias. Recall that in the ex-post case, Corollary 2 states that if the DM's prior is sufficiently concentrated in one state, the DM will find an uninformative model more decisive than a model providing evidence against that state. In contrast, the ex-ante decisiveness criterion predicts that the DM will never find an uninformative model more decisive than an informative model, regardless of her priors.

7.2.2 Selection as a Function of Objectives under Ex-Ante Decisiveness

Maximal Payoff Improvements. Selection under the ex-post criterion exhibits the following comparative static, formalized in Proposition 3: adding an action to the menu that increases the maximal payoff associated with a set of states but is ultimately not chosen can only result in the DM adopting a model that places lower likelihood on those states. Note that one implication of this property is that the ex-ante model violates IIA – the addition of an unchosen action can induce a change in the model the DM adopts, and in turn lead to

[19]It can be shown that if $m = \lambda m' + (1-\lambda)m_\varnothing$, then m is a garbling of m' and so is dominated by m' in the Blackwell order.

a change in the action taken by the DM. The ex-ante model, on the other hand, does not allow for such violations of IIA – the addition of unchosen actions cannot induce a change in model selection. This precludes an analog of Proposition 3 from holding in the ex-ante case.

Reductions in Action Value. As formalized in Proposition 4, in the ex-post case, reducing the payoffs of an action a recommended by model m reduces the decisiveness of m relative to any model that does not recommend a. A similar comparative static holds for the ex-ante model. In particular, suppose model m recommends action a for some signal realization; it must be that reducing the payoffs of a reduces the decisiveness of m relative to any model that does not recommend a for any signal realization. In other words, both versions of the model share the following feature: he more predisposed the DM is to taking an action, the more they will tend to adopt interpretations of the data that recommend that action, as opposed to interpretations recommending other actions. As discussed in Appendix A.5, this property can generate belief polarization along differences in objectives in the ex-ante model, just as in the ex-post model.

Relationship to Optimism. Recall that ex-post decisiveness delivers distinct predictions from a model selection criterion based on optimism: in particular, the ex-post decisiveness of a model is evaluated only based on its implications for the relative, as opposed to absolute, values of actions – a model that induces pessimistic beliefs can still be decisive, so long as it produces a strong recommendation toward one course of action over its alternatives.

This distinction no longer holds in the ex-ante case, however. The ranking over models induced by ex-ante decisiveness is equivalent to a model selection criteria based on ex-ante optimism. To see this, note that the expression for expected decisiveness, for $\mathcal{D} = (A, u)$, can be expressed as

$$I^E_{\mathcal{D}}(m) = \sum_s \left[\max_{a \in A} \sum_\theta u(a.\theta) p^m(\theta|s) \right] p^m(s) - \sum_\theta \max_{a' \in A} u(a', \theta)\rho(\theta)$$

As the second term on does not depend on m, selecting the model that maximizes ex-ante decisiveness-maximizing model is equivalent to selecting the model that maximizes expected utility.

Decision-Relevance of States. Consider the setting analyzed in Section 4.3, in which the state space can be expressed as $\Theta = \Theta_1 \times \Theta_2 \times ... \times \Theta_K$. As discussed in Section 4.3, the ex-post model exhibits the following property: if a state Θ_k is a nuisance variable – that is, if any model in M that accounts for Θ_k in explaining the signal is less informative about the remaining states compared to some model in M that neglects Θ_k – then if Θ_k is not decision-relevant, the DM will adopt a model that neglects Θ_k. As this result is based on the fact that ex-post decisiveness generates over-precision, a feature shared by the ex-post case, an analogous result holds in the ex-ante case, as detailed in Appendix A.5.

8 Discussion

This paper presents a theory of model selection and inference based on the insight that individuals seek decisive models, or models that provide clear decision-making guidance, and studies the implications of this model selection criterion on inference and choice. I conclude by discussing potential extensions and additional applications of the theory.

One limitation of the theory is that it studies model selection as a one-shot procedure. In reality, the set of models decision-makers entertain is often in flux as they are exposed to new models, and decision-makers may revise their working model in light of new information. Section 6 studies one such setting, in which the set of models the decision-maker entertains expands as a result of the social exchange of models. A more complete extension of the theory that considers these dynamics could shed light on how the models individuals adopt change over time, and which models tend to survive the realization of uncertainty. An additional set of applications of the theory is to study its implications for model persuasion (Schwartzstein and Sunderam, 2021), in which senders influence receivers' beliefs by proposing models to interpret known data – under the assumption that receivers select models that they find decisive. While Section 6 analyzes a special case of model persuasion in which senders' preferences are aligned with those of the receiver, a more complete analysis of model persuasion under the decisiveness criterion could shed further light on what models we should expect decision-makers to be exposed to in the presence of persuaders.

References

Aina, C. (2022). Tailored stories.

Anscombe, F. J. and Aumann, R. J. (1963). A definition of subjective probability. *Annals of Mathematical Statistics*, 34(1):199–205.

Barberis, N., Shleifer, A., and Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics*, 49(3):307–343.

Bastardi, A., Uhlmann, E. L., and Ross, L. (2011). Wishful thinking: Belief, desire, and the motivated evaluation of scientific evidence. *Psychological Science*, 22(6):731–733.

Bell, D. E. (1982). Regret in decision making under uncertainty. *Operations Research*, 30(5):961–981.

Blackwell, D. (1953). Equivalent comparisons of experiments. *The Annals of Mathematical Statistics*, 24(2):265–272.

Caplin, A. and Leahy, J. V. (2019). Wishful thinking.

Chernev, A., Böckenholt, U., and Goodman, J. (2015). Choice overload: A conceptual review and meta-analysis. *Journal of Consumer Psychology*, 25(2):333–358.

Dana, J., Dawes, R., and Peterson, N. (2013). Belief in the unstructured interview: The persistence of an illusion. *Judgment and Decision-making*, 8(5):512–520.

Darley, J. M. and Gross, P. H. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology*, 44(1):20–33.

Dunford, N. and Schwartz, J. T. (1957). *Linear Operators, Part 1: General Theory.* John Wiley Sons.

Eliaz, K. and Spiegler, R. (2020). A model of competing narratives. *American Economic Review*, 110(12):3786–3816.

Eyster, E., Li, S., and Ridout, S. (2021). A theory of ex post rationalization.

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford university press.

Frankel, A. and Kamenica, E. (2019). Quantifying information and uncertainty. *American Economic Review*, 109(10):3650–3680.

Gaertig, C. and Simmons, J. P. (2018). Do people inherently dislike uncertain advice? *Psychological Science*, 29(4):504–520.

Gagnon-Bartsch, T., Rabin, M., and Schwartzstein, J. (2021). Channeled attention and stable errors.

Glaeser, E. L. and Sunstein, C. R. (2009). Extremism and social learning. *Journal of Legal Analysis*, 1(1):263–324.

Graeber, T. (2022). Inattentive inference. *Journal of the European Economic Association*, 19(3):340–348.

Hansen, L. and Sargent, T. J. (2001). Robust control and model uncertainty. *American Economic Review*, 91(2):60–66.

Harmon-Jones, E., Harmon-Jones, C., and Levy, N. (2015). An action-based model of cognitive-dissonance processes. *Current Directions in Psychological Science*, 24(3):185–189.

Hayashi, T. (2008). Regret aversion and opportunity dependence. *Journal of Economic Theory*, 139(1):242–268.

Hong, H., Stein, J. C., and Yu, J. (2007). Simple forecasts and paradigm shifts. *The Journal of Finance*, 62(3):1207–1242.

Izzo, F., Martin, G. J., and Callander, S. (2021). Ideological competition.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kahneman, D. and Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80(4):237–251.

Kruglanski, A. W. and Fishman, S. (2009). The need for cognitive closure. *Handbook of Individual Differences in Social Behavior*, pages 343–353.

Krull, D. S. (1993). Does the grist change the mill? the effect of the perceiver's inferential goal on the process of social inference. *Personality and Social Psychology Bulletin*, 19(3):340–348.

Loomes, G. and Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal*, 92(368):805–824.

Lord, C. G., Ross, L., and Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, 37(11):2098–2109.

Maćkowiak, B., Matějka, F., and Wiederholt, M. (2021). Rational inattention: A review.

Moore, D. A., Tenney, E. R., and Haran, U. (2015). Overprecision in judgment. *The Wiley Blackwell Handbook of Judgment and Decision-making*, 2:182–209.

Mullainathan, S. (2002). Thinking through categories.

Mullainathan, S., Schwartzstein, J., and Shleifer, A. (2008). Coarse thinking and persuasion. *The Quarterly Journal of Economics*, 132(2):577–619.

Myers, D. G. and Bishop, G. D. (1976). Discussion effects on racial attitudes. *Science*, 169(3947):778–779.

Myers, D. G. and Lamm, H. (1976). The group polarization phenomenon. *Psychological Bulletin*, 83(4):602–627.

Plous, S. (1991). Biases in the assimilation of technological breakdowns: Do accidents make us safer? *Journal of Applied Social Psychology*, 21(13):1058–1082.

Price, P. C. and Stone, E. R. (2004). Intuitive evaluation of likelihood judgment producers: Evidence for a confidence heuristic. *Journal of Behavioral Decision Making*, 17(1):39–57.

Proulx, T. and Inzlicht, M. (2012). The five "a" s of meaning maintenance: Finding meaning in the theories of sense-making. *Psychological Inquiry*, 23(4):317–335.

Rabin, M. and Schrag, J. L. (1999). First impressions matter: A model of confirmatory bias. *The Quarterly Journal of Economics*, 114(1):37–82.

Rockafellar, R. T. (2016). *Convex Analysis*. Princeton University Press.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. *Advances in Experimental Social Psychology*, 10:173–220.

Roux, N. and Sobel, J. (2015). Group polarization in a model of information aggregation. *American Economic Journal: Microeconomics*, 7(4):202–32.

Sarver, T. (2008). Anticipating regret: Why fewer options may be better. *Econometrica*, 76(2):263–305.

Schkade, D., Sunstein, C. R., and Kahneman, D. (2000). Deliberating about dollars: The severity shift. *Columbia Law Review*, 100:1139–1175.

Schwartzstein, J. (2014). Selective attention and learning. *Journal of the European Economic Association*, 12(6):1423–1452.

Schwartzstein, J. and Sunderam, A. (2021). Using models to persuade. *American Economic Review*, 111(1):276–323.

Schwartzstein, J. and Sunderam, A. (2022). Shared models in networks, organizations, and groups.

Stoye, J. (2011). Axioms for minimax regret choice correspondences. *Journal of Economic Theory*, 146(6):2226–2251.

Tversky, A. and Shafir, E. (1992). Choice under conflict: The dynamics of deferred decision. *Psychological Science*, 3(6):358–361.

Vonk, R. and Konst, D. (1998). Intergroup bias and correspondence bias: People engage in situational correction when it suits them. *British Journal of Social Psychology*, 37(3):379–385.

Yates, J. F., Price, P. C., Lee, J.-W., and Ramirez, J. (1996). Good probabilistic forecasters: The 'consumer's' perspective. *International Journal of Forecasting*, 12(1):41–56.

Appendix

A.1 Model Selection for Prediction Problems

As noted in Section 3, a natural approach to applying the theory in situations where the DM does not face a particular decision problem is to assume that the DM learns from data as if they face a prediction problem. Formally, a *prediction problem* is a decision problem $\mathcal{D} = (A, u)$ where the actions $A = \Delta(\Theta)$ consist of a belief report. For a prediction problem $\mathcal{D} = (\Delta(\theta), u)$, refer to u as a *scoring rule.* Say that a scoring rule u, as well as its associated decision problem $\mathcal{D}$, is *proper* if for all $p, q \in \Delta(\Theta)$, $\sum_\theta u(p, \theta)p(\theta) \geq \sum_\theta u(q, \theta)p(\theta)$; that is, a prediction problem is proper if the DM maximizes expected utility by truthfully reporting their beliefs.

Example (Proper Prediction Problems).
Quadratic Loss. Suppose $\Theta = \{0, 1\}$. In this case, we can represent $A = \Delta(\Theta)$ with the probability that $\theta = 1$; consider the scoring rule $u(a, \theta) = (a - \theta)^2$. It is well known that this scoring rule is proper. The decisiveness of a model m in this prediction problem is given by

$$I_{\mathcal{D}}(m|s) = -p^m(1 - p^m)$$

where $p^m \equiv p^m(\theta = 1|s)$. For the prediction problem associated with quadratic loss, the DM adopts the model that minimizes her posterior variance.

Logarthimic Loss. Consider the scoring rule $u(p, \theta) = -\log p(\theta)$. This scoring rule is proper, and the decisiveness of a model m in this prediction problem is

$$I_{\mathcal{D}}(m|s) = -\sum_\theta p^m(\theta|s) \ln(p^m(\theta|s))$$

That is, logarithmic loss leads the DM to adopt the model that minimizes posterior entropy. ▲

Here, I to characterize model selection under the decisiveness criterion in the case where $\mathcal{D}$ is a proper prediction problem. In particular, I ask whether the restriction to proper prediction problems imposes further properties on model selection beyond the necessary and sufficient properties given in Proposition 2. As Proposition 8 states, the answer turns out to be negative:

Proposition 8. A model choice correspondence C satisfies

1. Sen's α, β: If $m \in M \subseteq M'$ and $m \in C(M')$, then $m \in C(M)$. Also, if $m, m' \in C_{\mathcal{D}}(M)$, $M \subseteq M'$ and $m' \in C(M')$ then $m \in C(M')$.
2. Continuity: For all $m \in \mathcal{M}$, $\{m' \in \mathcal{M} : m' \in C(\{m, m'\})\}$ and $\{m' \in \mathcal{M} : m \in C(\{m, m'\})\}$ are closed.
3. Scale Invariance: For any $m \in \mathcal{M}$, if m' satisfies $m'(s|\theta) = \lambda m(s|\theta)\ \forall\ \theta, \lambda > 0$, then $C(\{m, m'\}) = \{m, m'\}$.

4. Extremeness: For $m, m' \in M$, if $m, m' \notin C(M)$, then for any $\lambda \in (0,1)$, $\lambda m + (1 - \lambda)m' \notin C(M)$.

5. Certainty Preference: If $m(s|\theta) = 1$ for any $\theta \in \Theta$ and $m \in M$, then $m \in C(M)$.

if and only if there exists a proper prediction problem $\mathcal{D}$ and signal s such that $C(M) = C_{\mathcal{D}}(M|s)$ for all $M \subseteq \mathcal{M}$.

In other words, the class of proper prediction problems is sufficiently rich so as not to rule out any patterns of model selection that would obtain under the decisiveness criterion.

A.2 Decision-Relevance of States: Details

As in Section 4.3, suppose that the state space can be expressed as $\Theta = \Theta_1 \times \Theta_2 \times ... \times \Theta_K$, where under the DM's prior ρ, the Θ_k are independent. Let ρ_k denote the marginal of the DM's prior over Θ_k, and for model m, let $m_{-k} : \Theta_{-k} \to \Delta(S)$ denote the likelihood function over states Θ_{-k} denoted by integrating m over θ_k, with $m_{-k}(s|\theta_{-k}) = \sum_{\theta_k} m(s|\theta)\rho_k(\theta)$. For a decision problem $\mathcal{D} = (A, u)$, say that Θ_k is not *decision relevant* if for all $a \in A$, $u(a, \theta)$ is constant in θ_k.

Say that m neglects state Θ_k if $m(s|\theta)$ is constant in θ_k. Say that Θ_k is a *nuisance variable* with respect to M if for all $m \in M$, there exists m^* that neglects k s.t. for some $\lambda \in [0,1]$, $\lambda m^*_{-k} + (1-\lambda)m_\varnothing = m_{-k}$. That is, a state is a nuisance variable if any model attributing the signal to that state renders the model less informative about other states.

Proposition 9. Fix a signal realization s, and suppose Θ_k is a nuisance variable with respect to M. If Θ_k is not decision relevant under $\mathcal{D}$ and additionally $I_{\mathcal{D}}(m|s) > I_{\mathcal{D}}(m_\varnothing|s)$ for some $m \in M$, then $C_{\mathcal{D}}(M)$ must contain a model that neglects k.

A.3 Shared Models and Group Polarization: Details

Here, I discuss an example illustrating that Bayesian information aggregation need not result in greater certainty over the state.

Example (Diagnositicity of Evidence). Suppose that a group of N individuals have a common, uniform prior over the state $\Theta = \{\theta_l, \theta_h\}$ (e.g. not guilty vs. guilty) and receive a public signal $s_h \in \{s_h, s_l\}$. Individuals are uncertain over how to interpret s_h, and in particular, entertain two different models that describe the data-generating process: under m_1, the signal is uninformative over the state, and under m_2, the signal provides evidence towards θ_h:

$$m_1(s_h|\theta_h) = 1 - m_1(s_h|\theta_h) = q > 1/2$$
$$m_2(s_h|\theta_h) = m_2(s_h|\theta_l) = 1/2$$

Suppose that individuals share a common, uniform prior over the two models, and each individual i obtains a private iid signal $\psi_i \in \{1, 2\}$ over the model space, where

$$Pr(\psi_i = 1|m_1) = Pr(\psi_i = 2|m_2) = p > 1/2$$

where the ψ_i are also independent of θ and s. It follows that if a greater number of individuals in the group receive the signal $\psi_i = 2$ than receive $\psi_i = 1$, then information aggregation causes individuals to become increasingly certain that m_2 describes the data-generating process, and therefore increasingly uncertain over the state. ▲

A.4 Value of Delay: Details

Consider a binary state setting where the DM has access to a vector of binary signals $s \in S_1 \times S_2 \times ... \times S_K$, where each $S_k = \{h, l\}$. Let m_k denote the model characterized by the likelihood ratio

$$L_{m_k}(s) \equiv \frac{m_k(s|\theta_h)}{m_k(s|\theta_l)} = \begin{cases} \lambda & s_k = h \\ 1/\lambda & s_k = l \end{cases}$$

for $\lambda > 1$. Suppose the true model $m_T = \sum_k \alpha_k m_k$ for weights $\alpha_k > 0$, $\sum_k \alpha_k = 1$, and let $M = \{m : m = \sum_k \beta_k m_k, \sum_k \beta_k = 1\}$. That is, M contains all possible mixtures of the m_k.

The following proposition states that if a Bayesian receives a signal s' that moves her beliefs further from the prior and reduces her value of delay relative to a signal s, a decisiveness-maximizing DM will exhibit a lower reduction in the value of delay.

Proposition 10. Fix any decision problem $\mathcal{D} = (u, A)$. Take any two signals $s, s' \in S$ such that either $\rho(\theta_h) \leq p^{m_T}(\theta_h|s) < p^{m_T}(\theta_h|s')$ or $\rho(\theta_h) \geq p^{m_T}(\theta_h|s) > p^{m_T}(\theta_h|s')$, and suppose that $K_A(\{m_T\}|s') < K_A(\{m_T\}|s)$. Then $K_A(\{m_T\}|s) - K_A(\{m_T\}|s') \leq K_A(M|s) - K_A(M|s')$.

A.5 Ex-Ante Decisiveness: Details

Let $C^E_{\mathcal{D}}(M) = \arg\max_{m \in M} I^E_{\mathcal{D}}(m)$ denote the models the DM adopts under the ex-ante decisiveness criterion. Begin by establishing that the extremeness property holds for the ex-ante criterion.

Proposition 11. For $m, m' \in M$, if $m, m' \notin C^E_{\mathcal{D}}(M)$, then for any $\lambda \in (0, 1)$, $\lambda m + (1 - \lambda)m' \notin C^E_{\mathcal{D}}(M)$.

We now show that overprecision holds for the ex-ante criterion. As in the main text, let m_T denote the true model and $m_\varnothing$ denote an uninformative model.

Proposition 12. For any m satisfying $\lambda m + (1 - \lambda)m_\varnothing = m_T$ for some $\lambda \in (0, 1)$, $I^E_{\mathcal{D}}(m) \geq I^E_{\mathcal{D}}(m_T)$

Now, we state analog of Proposition 4 for the ex-ante case. Consider two decision problems $\mathcal{D} = (A, u), \mathcal{D}' = (A, u')$. Say that action $a^* \in A$ is uniformly worse in $\mathcal{D}'$ relative to $\mathcal{D}$ if $u'(a^*, \theta) \leq u(a^*, \theta)$ for all θ and $u'(a, \theta) = u(a, \theta)$ for all θ, $a \neq a^*$.

Proposition 13. Suppose a^* is uniformly worse in $\mathcal{D}'$ relative to $\mathcal{D}$. If $m \in M$ recommends a^* from $\mathcal{D}'$ for some $s \in S$ and $m' \in C^E_{\mathcal{D}}(M)$ does not recommend a^* from $\mathcal{D}$ for any $s \in S$, then $m \notin C^E_{\mathcal{D}}(M) \implies m \notin C^E_{\mathcal{D}'}(M)$.

We now state the analog of Proposition 9 for the ex-ante case. Consider the setting studied in Appendix A.2. We have the following result:

Proposition 14. Suppose Θ_k is a nuisance variable with respect to M. If Θ_k is not decision relevant under $\mathcal{D}$ and additionally $I^E_{\mathcal{D}}(m) > I^E_{\mathcal{D}}(m_\varnothing)$ for some $m \in M$, then $C^E_{\mathcal{D}}(M)$ must contain a model that neglects k.

A.6 Behavioral Characterization and Identification Results

A key primitive in the framework is the set of models the DM entertains, M. In many situations, however, the set of models the DM entertains is difficult or impossible to directly observe. This raises the following questions: if we are unwilling to make a-priori restrictions on M, does the theory nevertheless make meaningful restrictions on choice, and can M be deduced from choice data?

I take up both of these questions, working in an extended environment that includes objective lotteries over outcomes following Anscombe and Aumann (1963). In particular, I build directly on results from Stoye (2011), which characterize min-max regret choice correspondences, to provide characterization and identification results for the theory.

A.6.1 Extending the Environment

We extend the DM's decision environment as follows. There is a finite set of prizes Z. The DM chooses from menus of acts, $A \subseteq X \equiv \Delta(Z)^\Theta$, where each act $f \in X$ is a mapping from states to objective lotteries over prizes. Let $\mathcal{A}$ denote the collection of finite subsets of X. Take as data the choice correspondence $\mathcal{C} : \mathcal{A} \rightrightarrows X$, which satisfies $\mathcal{C}(A) \subseteq A$ for all $A \in \mathcal{A}$.

Because multiple models can induce the same posterior belief for a given signal realization, it will not in general be possible to identify the set of models the DM entertains; the representation we consider will therefore focus on the set of posteriors P induced by the models the DM entertains. Recall that under Assumptions 3 and 4, which will be maintained in this section, there exists a model inducing any posterior $p \in \Delta(\Theta)$. Recall also that Assumption 2 ensures that the set of models M the DM entertains is closed; carry this assumption into this environment and restrict attention to closed sets of model-induced posteriors in the representation. Let $\mathcal{P}$ denote the collection of all closed subsets of $\Delta(\Theta)$. Let $\mathcal{U}$ denote the set of utility functions from Z to $\mathbb{R}$, and extend these to $\Delta(Z)$ by taking expectations[20].

Definition (Decisiveness-Maximizing Representation). A choice correspondence $\mathcal{C}$ has a

[20] In particular, for $u \in \mathbb{R}^Z$, for each $q \in \Delta(Z)$ let $u(q) = \sum_{z \in Z} u(z)q(z)$.

decisiveness-maximizing representation if there exists $(u, P) \in \mathcal{U} \times \mathcal{P}$ such that for all $A \in \mathcal{A}$,

$$\mathcal{C}(A) = \bigcup_{p \in \mathcal{I}(P|A)} \arg\max_{f \in A} \sum_{\theta \in \Theta} u(f(\theta))p(\theta)$$

where

$$\mathcal{I}(P|A) = \arg\max_{p \in P} \left\{ \max_{f \in A} \sum_{\theta \in \Theta} \left[u(f(\theta)) - \max_{f' \in A} u(f'(\theta)) \right] p(\theta) \right\}$$

If the above holds, say that (u, P) *represents* $\mathcal{C}$.

In words, if $\mathcal{C}$ is represented by a utility function and a set of model-implied posteriors, $\mathcal{C}$ chooses the acts that maximizes expected utility with respect to a posterior that maximizes decisiveness. Now, introduce two additional assumptions that will be needed for identification:

Assumption 5: There exists $z, z' \in Z$ such that $u(z) > u(z')$.

Assumption 6: P does not contain δ_θ for any $\theta \in \Theta$.

Assumption 5 amounts to a non-triviality assumption that ensures that the DM is not indifferent between all acts. Assumption 6 is substantive, and rules out models that induce certainty in a state. Assumptions 5 and 6 are crucial for the identification result that follows, but partial identification of the set of extreme models is still possible when Assumption 6 is relaxed, as I will subsequently discuss.

A.6.2 Relationship to Min-Max Regret Models

This model has a tight relationship with the *min-max regret* model, characterized in Hayashi (2008) and subsequently Stoye (2011), which has the representation

$$\mathcal{C}(A) = \arg\min_{f \in A} \max_{p \in P} \sum_{\theta \in \Theta} \left[\max_{f' \in A} u(f'(\theta)) - u(f(\theta)) \right] p(\theta)$$

Here, the DM chooses to minimize the worst-case expected regret, taken with respect to the set of beliefs P. As Lemma 2 in the Appendix shows, the decisiveness-maximizing representation is equivalent to a *min-min regret* model, where the DM chooses to minimize the *best-case* expected regret. As such, my characterization result directly follows the axiomatization in Stoye, with the appropriate adjustment in axioms to replace the max operator with a min operator. In particular, as I discuss below, while the min-max model is characterized by a preference for hedging, my model is instead characterized by an aversion to hedging.

A.6.3 Characterization Result

To state the axioms, first introduce some notation. For any act f, with some abuse of notation let $f(\theta)$ refer to a constant that yields lottery $f(\theta)$ in all states; where appropriate, let

z denote the constant act yielding prize $z \in Z$.

Say that f *improves on* A if there exists some θ for which for all $f' \in A$, $f(\theta) = \mathcal{C}(\{f(\theta), f'(\theta)\})$, and say that menu A' improves on A if there exists some $f \in A'$ that improves on A. Also, for menu A, act f, $\lambda \in [0,1]$ let $\lambda A + (1-\lambda)f = \{\lambda f' + (1-\lambda)f : f' \in A\}$. Finally, for acts f, g and menu A, let $f \oplus g = \frac{1}{2}f + \frac{1}{2}g$ and $A \oplus g = \frac{1}{2}A + \frac{1}{2}g$.

Consider the following axioms:

Axiom 1 (IINIA & IISIA).

- IINIA: Take $A \subseteq A'$, where A' does not improve on A. If $f \in A \subset A'$ then $f \in \mathcal{C}(A') \implies f \in \mathcal{C}(A)$. Also if $f, f' \in \mathcal{C}(A)$ then $f \in \mathcal{C}(A') \implies f' \in \mathcal{C}(A')$.
- IISIA: For $f_a, f_b \in X_c$, if A does not improve B and $A \oplus f_a$ does not improve $B \oplus f_b$, and additionally $\mathcal{C}(A \cup B) \subseteq A$ and $\mathcal{C}((A \oplus f_a) \cup (B \oplus f_b)) \subseteq A \oplus f_a$, then $f \in \mathcal{C}(A \cup B) \implies f \oplus f_a \in \mathcal{C}((A \oplus f_a) \cup (B \oplus f_b))$.

Axiom 2 (Monotonicity). If $f'(\theta) \in \mathcal{C}(f(\theta), f'(\theta))$ for all θ, then $f \in \mathcal{C}(A \cup \{f\}) \implies f' \in \mathcal{C}(A \cup \{f'\})$. Also, if $f'(\theta) = \mathcal{C}(f(\theta), f'(\theta))$ for all θ, then $f' \in A \implies f \notin \mathcal{C}(A)$.

Axiom 3 (Mixture Continuity). For $f, g, h \in X$ and $A \in \mathcal{A}$, the sets

$$\{\alpha \in [0,1] : \alpha f + (1-\alpha)g \in \mathcal{C}(\{\alpha f + (1-\alpha)g\} \cup \{h\} \cup A)\}$$
$$\{\alpha \in [0,1] : h \in \mathcal{C}(\{\alpha f + (1-\alpha)g\} \cup \{h\} \cup A)\}$$

are closed.

Axiom 4 (Mixture Independence). For $\lambda \in (0,1)$, $g \in X$, $f \in \mathcal{C}(A) \implies \lambda f + (1-\lambda)g \in \mathcal{C}(\lambda A + (1-\lambda)g)$.

Axiom 5 (Mixture Aversion). For $f, g \in X, \lambda \in (0,1)$ s.t. $\{f, g\} \subseteq \mathcal{C}(A)$, $f \in \mathcal{C}(A \cup \{\lambda f + (1-\lambda)g\})$.

Axiom 6 (Non-Triviality). There exists $z, z' \in Z$ such that $\mathcal{C}(\{z, z'\}) = z$.

Axiom 7 (No Certainty). For $f_a, f_b, f_c \in X_c$ s.t. $f_a = \mathcal{C}(f_a, f_b)$, $f_b = \mathcal{C}(f_b, f_c)$, for all θ^*, there exists $\lambda^* \in (0,1)$ such that for all $\lambda \in (0, \lambda^*)$, act g with $g(\theta) = \lambda f_a(\theta) + (1-\lambda) f_b(\theta)$ for $\theta = \theta^*$, $g(\theta) = f_c(\theta)$ for $\theta \neq \theta^*$, and any A that neither improves nor is improved by $g(\theta^*)$, $g \notin \mathcal{C}(A)$ whenever $f_b \in A$.

Axiom 1 is a weakening of the standard IIA axioms. This weakening reflects the fact that model selection under the decisiveness criterion, and therefore the evaluation of a given action, depends on the decision problem that the DM faces; Axiom 1 places testable restrictions on the nature of this dependence. In particular, IINIA states that choice satisfies an IIA property with respect to acts that do not improve the maximal payoff profile — that is, only the addition of acts that improve the maximal payoff profile can induce a change in choice among existing alternatives. On the other hand, IISIA states that scaling the payoffs of irrelevant maximal actions cannot induce preference reversals among non-maximal actions.

Example (Violation of IINIA). Suppose the prize space contains two elements, z_h, z_l, where $\mathcal{C}(\{z_h, z_l\}) = z_h$. Identify each act by the probability it places on prize z_h in each state. Suppose we have the acts f, g, h with

	θ_1	θ_2
f	1	0
g	0	1
h	0	0.9

and suppose that $\mathcal{C}(\{f, g\}) = f$ and $\mathcal{C}(\{f, g, h\}) = g$. Choice in this example exhibits the asymmetric dominance effect — the addition of h, which is dominated by g, causes the DM to switch from f to g. IINIA rules out such forms of menu dependence. ▲

Axiom 2 imposes that choice obeys state-wise dominance relationships. Axiom 3 is a technical condition, and Axiom 6 reflects the assumption that the set of acts is non-trivial, whereas Axiom 7 reflects the assumption that the set of models the DM entertains precludes models that induce complete certainty.

Axiom 4 is a weakening of the standard Independence axiom, which states that choice from a menu is invariant to mixing all acts in the menu with another act. It jointly captures two properties of the theory. The first property concerns choice when the DM is restricted to a single model. In this case, the theory collapses to a case of subjective expected utility (SEU), where the DM evaluates prospects according to a fixed set of subjective beliefs (a fixed model) – in SEU, Independence must hold. The second property is a restriction on how model selection can depend on the DM's decision problem. In particular, since the decisiveness of a model depends only on statewise *differences* between the utility levels of alternatives in the menu, shifting the utility levels that all acts deliver in a state by the same constant cannot induce a change in which model the DM selects. Since in the theory, choice for a fixed model satisfies Independence, and model selection also is invariant to such mixing operations performed on DM's decision problem, choice itself must satisfy Independence.

Example (Violation of Mixture Independence). Suppose again that $Z = \{z_h, z_l\}$ with $\mathcal{C}(\{z_h, z_l\}) = z_h$, and again identify each act by the probability it places on prize z_h in each state. Suppose we have the acts f, g, f', g'

	θ_1	θ_2
f	0.6	0.4
g	0.8	0.2

	θ_1	θ_2
f'	0.3	0.7
g'	0.4	0.6

and suppose $\mathcal{C}(\{f, g\}) = g$ and $\mathcal{C}(\{f', g'\}) = f'$. Choice in this example can be rationalized by a model of wishful thinking — the DM places higher beliefs in the state under which she can obtain higher payoffs — θ_1 when the menu is $\{f, g\}$ and θ_2 when the menu is $\{f', g'\}$.

Note however, that $f' = \frac{1}{2}f + \frac{1}{2}h$, $g' = \frac{1}{2}g + \frac{1}{2}h$, where h satisfies $h(\theta_1) = z_l, h(\theta_2) = z_h$, and so this choice pattern violates Mixture Independence. Note that although that the above mixing operation changes the payoff levels that acts deliver in either state, the payoff differences within states are preserved; it is the latter consideration, rather than the former, that determines the decisiveness of a given model. ▲.

Axiom 5 imposes that choice satisfies an extremeness property: mixing between two acts cannot result in an act that "improves" on those acts. Axiom 5 directly corresponds to the key extremeness property of the decisiveness criterion formalized in Proposition 1 — the DM's tendency to select extreme models directly translates into a tendency to choose extreme actions – or alternatively a tendency against choosing "diversified" actions or hedging. Note the relationship between Axiom 5 and the Uncertainty Aversion axiom in Gilboa & Schmeidler's (1986) characterization of the Min-max Expected Utility (MEU) model, which embodies the opposite preference towards diversification: if the DM is indifferent between two acts, she must prefer a mixture of those acts to either act. As one might expect, documented choice patterns reflecting ambiguity aversion, which MEU was formalized to rationalize, are inconsistent with Axiom 5, and therefore cannot be rationalized by my theory.

Example (Violation of Mixture Aversion). As before, let $Z = \{z_h, z_l\}$ with $\mathcal{C}(\{z_h, z_l\}) = z_h$, and identify each act by the probability it places on prize z_h in each state. Suppose we have the acts f, g, h

	θ_1	θ_2
f	1	0
g	0	1
h	0.5	0.5

with $\mathcal{C}(\{f, g\}) = \{f, g\}$ but $\mathcal{C}(\{f, g, h\}) = h$. This choice pattern corresponds to a experimental findings for a variant of the classic Ellsberg paradox (Becker and Brownson 1964) in which h corresponds to betting on a black ball being drawn from an urn known to have 50 black and 50 red balls, and f and g correspond to betting on a black and red ball, respectively, being drawn from an urn with an unknown mix of black and red balls. Note that such choice patterns are ruled out by Axiom 5. ▲

Although Axiom 5 is at odds with an established body of evidence for ambiguity-averse preferences in choice settings similar in structure to the example above, there is also a body of evidence for mixture-averse preferences consistent with Axiom 5. For example, Heath and Tversky (1991) demonstrate that aversion to ambiguity reverses when subjects bet in domains in which they have high perceived expertise. In particular, when subjects self-identify as having expertise regarding events for which objective probabilities are unavailable (e.g. outcomes of elections or football matches), they prefer to bet on those events over events with known probabilities, even when the former outcomes are judged by subjects to be equiprobable — behavior which violates Uncertainty Aversion but is consistent with Axiom 5. Viewed through the lens of my theory, subjects who have greater expertise in a given domain may be able to entertain a greater range of arguments (models) for or against any given outcome, which would rationalize the documented mixture-averse behavior. Outside of stylized lab experiments, evidence for behavior consistent with mixture aversion abounds: it is well known that investors tend to hold underdiversified portfolios consisting of too few securities to eliminate idiosyncratic risk (e.g. Mitton and Vorkink 2007) and also tend to exhibit home bias, concentrating their ownership in domestic stocks (French and Poterba 1991). As Section 5 demonstrates, in this model, such behavior can be rationalized as a consequence of uncertainty over the payoff distributions of securities.

The following result states that choice in the model is characterized by Axioms 1–5. Recall that Proposition 1 states that for any decision problem, only extreme models can strictly minimize decisiveness. This implies that in general, only identification of the set of extreme model-induced posteriors from choice data will be possible. The result shows that the set of extreme model-induced posteriors are indeed identified, if we additionally assume Axioms 6 and 7 – that is, if Assumptions 5 and 6 are satisfied.

Theorem 1. $\mathcal{C}$ satisfies Axioms 1–5 if and only if it has a decisiveness-maximizing representation (u, P). $\mathcal{C}$ additionally satisfies Axioms 6 and 7 if and only if u and P satisfy Assumptions 5 and 6, respectively, and for any (u', P') representing $\mathcal{C}$, there exists constants $\alpha > 0$, β such that $u' = \alpha u + \beta$, and $ext(P') = ext(P)$.

The proof of the characterization result in Theorem 1 directly builds on characterization results for the min-max model established in Stoye (2011), which shows that Axioms 1-4, as well as an Uncertainty Aversion axiom, characterize the min-max regret model. Note, however, that while the extreme set of model-implied posteriors is identified in min-max regret model if Axiom 6 (Non-Triviality) is assumed, the same is not the case in my model. Intuitively, if the DM entertains a model-induced posterior $p \in P$ that places certainty in a state θ – that is, if Assumption 6 does not hold – the DM will never select (outside of cases of indifference) a model that does not place certainty in any state, even if such a model is extreme. This precludes the identification of the full set of extreme models in the case where Assumption 6 does not hold, and so the No Certainty axiom, which guarantees that Assumption 6 holds – is required for identification. Below I provide a partial identification result in a setting where Assumption 6 is relaxed.

For a set of model-induced posteriors P, let $ct(P) = \{p \in P : p = \delta_\theta \text{ for some } \theta \in \Theta\}$ denote the set of certainty-inducing posteriors. The following result states that when $ct(P)$ is non-empty it can be identified from choice data, and moreover that choice data identifies whether or not $ct(P)$ is non-empty.

Theorem 2. $\mathcal{C}$ satisfies Axioms 1–6 if and only if it has a decisiveness maximizing representation (u, P), where u satisfies Assumption 5. Also, for any (u', P') representing $\mathcal{C}$, there exists constants $\alpha > 0$, β such that $u = \alpha u' + \beta$, and $ct(P) = ct(P')$.

Note that in the case where the data identifies that $ct(P) = \emptyset$, Assumption 6 holds and so we can use the identification result in Theorem 1 to identify the full set of extreme models from choice data.

A.6.4 Relative Diversification Aversion

Here, I state a comparative static that corresponds to Proposition 6 in the main text, which states that if the DM entertains a larger set of models, the DM will be more averse to diversification. Consider a family of choice functions $\{\mathcal{C}^k\}_{k\in K}$ that have a decisiveness-maximizing representation, and that share the same utility over prizes. It can be shown that if $\mathcal{C}$ has a decisiveness-maximizing representation, choice over menus of constant acts can be described by a preference relation (see Lemma 4); let $\succeq^*$ denote the preference relation that describes choice over constant acts for members of $\{\mathcal{C}^k\}_{k\in K}$. For a menu A, let f^A_{max} denote any act that satisfies $f^A_{max}(\theta) \in \{f(\theta) : f \in A, f(\theta) \succeq^* g(\theta) \, \forall\, g \in A\}$ — that is, f^A_{max} delivers the most preferred outcome in each state among acts in A.

Definition (C-Diversified Act). Say that h is a *C-diversified act* with respect to A if there exists lotteries over prizes $\overline{q}, \underline{q} \in \Delta(Z)$ satisfying $\overline{q} \succeq^* \underline{q}$ such that $f^A_{max}(\theta) \oplus \underline{q} \sim^* h(\theta) \oplus \overline{q}$ for all θ. Let H_A collect acts that are C-diversified with respect to A.

As in the definition of a C-defined action in the main text, a C-diversified act has the following interpretation: to construct a C-diversified act, one takes the maximal payoff achievable in each state and reduces that payoff by a constant.

Definition 3 (Relative Diversification Aversion). Say that $\mathcal{C}'$ is *more diversification-averse* than $\mathcal{C}$ if for any $A \in \mathcal{A}$, $h \in H_A$, $g \notin \mathcal{C}(A \cup \{h\}) \implies g \notin \mathcal{C}'(A \cup \{h\})$.

That is, $\mathcal{C}'$ is more diversification-averse than $\mathcal{C}$ if $\mathcal{C}'$ never chooses a diversified act from a menu whenever $\mathcal{C}$ does not. I now stat the analog of Proposition 6 in the main text.

Theorem 3. Suppose $\mathcal{C}, \mathcal{C}'$ are represented by (u, P), (u, P'), respectively. If $co(P) \subseteq co(P')$, then $\mathcal{C}'$ is more diversification-averse than $\mathcal{C}$. Furthermore, if $\mathcal{C}'$ is more diversification-averse than $\mathcal{C}$ and P' satisfies Assumption 6, then $co(P) \subseteq co(P')$.

A.7 Proofs: Propositions in Main Text

Proof of Proposition 1. Fix some signal s; in what follows I drop the dependence on s in the notation, where appropriate. Since $\max_a \sum_y u(a,\theta)p(\theta)$ is convex in p and $\sum_y \max_{a'} u(a',\theta)p(\theta)$ is linear in p, $R_{\mathcal{D}}$ is concave in p. Bayes' rule implies that for all $\alpha \in (0,1)$, there exists $\lambda \in (0,1)$ such that $p^{\lambda m+(1-\lambda)m'} = \alpha p^m + (1-\alpha)p^{m'}$, and so $I_{\mathcal{D}}$ is convex. Therefore, $I_{\mathcal{D}}(\lambda m + (1-\lambda)m') \leq \max\{I_{\mathcal{D}}(m), I_{\mathcal{D}}(m')\}$, and so $m, m' \notin C_{\mathcal{D}}(M) \implies \lambda m + (1-\lambda)m' \notin C_{\mathcal{D}}(M)$, thus establishing the first part of the proposition.

Now additionally suppose that $A^m, A^{m'}$ are disjoint. Toward a contradiction, suppose that $\lambda m + (1-\lambda)m' \in C_{\mathcal{D}}(M)$ for $\lambda \in (0,1)$. Bayes' rule implies that there exists $\alpha \in (0,1)$ s.t. $\alpha p^m + (1-\alpha)p^{m'} = p^{\lambda m+(1-\lambda)m'}$. We then have $R_{\mathcal{D}}(p^{\lambda m+(1-\lambda)m'}) \leq \alpha R_{\mathcal{D}}(p^m) + (1-\alpha)R_{\mathcal{D}}(p^{m'})$. Fix $a \in A^m_{\mathcal{D}}$ and $a' \in A^{m'}_{\mathcal{D}}$. The preceding inequality implies that for any $a'' \in A^{\lambda m+(1-\lambda)m'}_{\mathcal{D}}$,

$$\sum_\theta u(a'',\theta)(\alpha p^m(\theta) + (1-\alpha)p^{m'}(\theta)) \geq \alpha \sum_\theta u(a,\theta)p^m(\theta) + (1-\alpha) \sum_\theta u(a,\theta)p^{m'}(\theta)$$

Since $\sum_\theta u(a'',\theta)p^m(\theta) \leq \sum_\theta u(a,\theta)p^m(\theta)$ and $\sum_\theta u(a'',\theta)p^{m'}(\theta) \leq \sum_\theta u(a',\theta)p^{m'}(\theta)$, the above implies that $\sum_\theta u(a'',\theta)p^m(\theta) = \sum_\theta u(a,\theta)p^m(\theta)$ and $\sum_\theta u(a'',\theta)p^{m'}(\theta) = \sum_\theta u(a',\theta)p^{m'}(\theta)$, which in turn implies that $a'' \in A^m_{\mathcal{D}}, A^{m'}_{\mathcal{D}}$, a contradiction.

□

Proof of Proposition 2. Fix some signal s; in what follows I drop the dependence on s in the notation, where appropriate. Proposition 1 shows necessity of Extremeness; necessity of Sen's α, β, Certainty Preference, and Scale Invariance are immediate from the definition of the regret criterion.

To see necessity of Continuity, let $E \subset \mathbb{R}^n$ denote a convex, compact superset of $\Delta(\Theta)$. Consider $\tilde{H} : E \to \mathbb{R}$ defined by $\tilde{H}(p) = \max_{a \in A} \left\{ \sum_\theta u(a,\theta)p(\theta) - \sum_p \max_{a' \in A} u(a',\theta)p(\theta) \right\}$. Since $\tilde{H}$ is the maximum over a family of affine functions on a convex, compact set, $\tilde{H}$ is convex. By construction $\tilde{H}$ is bounded from above by 0, and is bounded from below. So $\tilde{H}$ is a proper convex function and is therefore continuous on the relative interior of E[21], which in turn implies that H, the restriction of $\tilde{H}$ to $\Delta(\Theta)$, is continuous. Therefore, $R_{\mathcal{D}}(p) = -H(p)$ is continuous, and so $I_{\mathcal{D}}(m) = R_{\mathcal{D}}(p^m)$ is continuous since the map $m \to p^m$ is continuous.

To show sufficiency, note that Sen's α, β implies that there is a preference relation $\succeq$ where $C_{\mathcal{D}}(M) = \{m \in M : m \succeq m' \, \forall m' \in M\}$. Since $\mathcal{M}$ is a closed, bounded subset of $\mathbb{R}^n$, $\succeq$ can be represented by a continuous utility function $V : \mathcal{M} \to \mathbb{R}$. Let $\tilde{R} = -V$. Since C satisfies Extremeness, $\tilde{R}$ must be concave. Since C satisfies Certainty Preference, $\tilde{R}$ must attain its minimum value at any m such that $p^m = \delta_\theta$ for any $\theta \in \Theta$; normalize this minimum value to 0. Scale Invariance implies that $\tilde{R}(m) = \tilde{R}(m')$ whenever $p^m = p^{m'}$, and so the function $R : \Delta(\Theta) \to \mathbb{R}$ satisfying $R(p^m) = \tilde{R}(m)$ is well defined.

Bayes' rule implies that for all $\lambda \in (0,1)$, there exists $\alpha \in (0,1)$ such that $\lambda p^m + (1-\lambda)p^{m'} = p^{\alpha m + (1-\alpha)m'}$, and so R inherits concavity from $\tilde{R}$, and for any $\theta \in \Theta$, $R(\delta_\theta) = 0$. By Theorem 2 of Frankel and Kamenica (2019), there exists a set of actions A and utility function $u : A \times \Theta \to \mathbb{R}$ such that $R(p) = \sum_{\theta \in \Theta} p(\theta) \max_a u(a,\theta) - \max_a \sum_{\theta \in \Theta} p(\theta)u(a,\theta)$.

□

Proof of Corollary 1. Follows immediately from Proposition 1.

□

Proof of Corollary 2. Fix some signal s; in what follows I drop the dependence on s in the notation, where appropriate. In the binary state case, the residual uncertainty can be expressed as $R_{\mathcal{D}}(p)$, where $p = p(\theta_h)$. For any model m, let $p^m \equiv p^m(\theta_h)$. By Proposition 2, we know that $R_{\mathcal{D}}(0) = R_{\mathcal{D}}(1) = 0$, $R_{\mathcal{D}}(p) \geq 0$, and that $R_{\mathcal{D}}$ is continuous. This implies that there exists $p^* > 0$ where $R_{\mathcal{D}}(p)$ is increasing for all $p \leq p^*$. By Bayes' rule, if $m(s|\theta_h)/m(s|\theta_l) < \frac{(1-\overline{p})p^*}{\overline{p}(1-p^*)}$, then $p^m \leq p^*$ so long as $\rho(\theta_h) \leq \overline{p}$, which in turn implies that $I_{\mathcal{D}}(m') \geq I_{\mathcal{D}}(m)$ for $\rho(\theta_h) \leq \overline{p}$ and any m' in favor of θ_l since $R_{\mathcal{D}}$ is increasing on $[0, p^*]$.

[21] Theorem 10.4 in Rockafellar, "Convex Analysis"

Finally, note that for $\rho(\theta_h) \leq \overline{p}$ and any m'' with $m''(s|\theta_h)/m''(s|\theta_l) \leq m(s|\theta_h)/m(s|\theta_l)$, $p^{m''} \leq p^m \leq p^*$ and so $I_{\mathcal{D}}(m') \geq I_{\mathcal{D}}(m'')$.

□

Proof of Proposition 3. Fix some signal s; in what follows I drop the dependence on s in the notation, where appropriate. Note that $R_{\mathcal{D}'}(p^m) \leq R_{\mathcal{D}}(p^m) + \sum_\theta p^m \Delta u_\theta$. The inequality results from the fact that the expected regret of m under menu A' must be at most the expected regret associated with continuing to choose the action recommended by m under menu A. Since the actions recommended by m' are in the original menu A, we also have the equality $R_{\mathcal{D}'}(p^{m'}) = R_{\mathcal{D}}(p^{m'}) + \sum_\theta p^{m'} \Delta u_\theta$. Model m' must deliver lower expected regret than m under $\mathcal{D}'$, and so we have $R_{\mathcal{D}'}(p^{m'}) \leq R_{\mathcal{D}'}(p^m)$, which in turn yields $\sum_\theta p^{m'}(\theta)\Delta u_\theta \leq \sum_\theta p^m(\theta)\Delta u_\theta$.

□

Proof of Corollary 3. Fix some signal s; in what follows I drop the dependence on s in the notation, where appropriate. Take the actions in A that maximize posterior expected utility for some posterior. These actions can be ordered[22] $a_1, a_2, ..., a_K$ where $u(a_j, \theta_h) > u(a_k, \theta_h)$ and $u(a_j, \theta_l) < u(a_k, \theta_l)$ for $j > k$; in the binary state case, the utility-maximizing action is increasing with respect to this order in $p(\theta_h)$. By assumption, no $m' \in C_{\mathcal{D}'}(M)$ recommends a' from $\mathcal{D}'$, and so Proposition 3 implies that for any $m \in C_{\mathcal{D}}(M)$ and $m' \in C_{\mathcal{D}'}(M)$, $p^{m'}(\theta_h) \leq p^m(\theta_h)$. This in turn implies that any a recommended by m from $\mathcal{D}$ and any a'' recommended by m' from $\mathcal{D}'$ satisfies $u(a, \theta_h) \geq u(a'', \theta_h)$, and $u(a, \theta_l) \leq u(a'', \theta_l)$.

□

Proof of Proposition 4. Fix some signal s; in what follows I drop the dependence on s in the notation, where appropriate. Note that if m recommends a^* from $\mathcal{D}'$, then m recommends a^* from $\mathcal{D}$, since $u'(a^*, \theta) \leq u(a^*, \theta)$ for all θ, and $u'(a^*, \theta) = u(a^*, \theta)$ for all $\theta, a \neq a^*$. Similarly, if m' recommends $a \neq a^*$ from $\mathcal{D}$, then m' also recommends $a \neq a^*$ from $\mathcal{D}'$. Let $k_\theta = u(a^*, \theta) - u'(a^*, \theta)$, and let Θ_+ collect the outcomes for which a^* delivers the maximal payoff. We have

$$\begin{aligned} R_{\mathcal{D}}(p^m) &\leq R_{\mathcal{D}'}(p^m) - \sum_{\theta \in \Theta \setminus \Theta_+} k_\theta p^m(\theta) \\ &\leq R_{\mathcal{D}'}(m) \end{aligned}$$

Furthermore, since m' recommends the same action from both $\mathcal{D}, \mathcal{D}'$, we have

$$\begin{aligned} R_{\mathcal{D}'}(p^{m'}) &= R_{\mathcal{D}}(p^{m'}) - \sum_{\theta \in \Theta_+} \min \left\{ k_\theta, u(a^*, \theta) - \max_{a \neq a^*} u(a, \theta) \right\} p^m(\theta) \\ &\leq R_{\mathcal{D}}(p^{m'}) \end{aligned}$$

Since $m \notin C_{\mathcal{D}}(M), m' \in C_{\mathcal{D}}(M)$, $R_{\mathcal{D}}(p^{m'}) < R_{\mathcal{D}}(p^m)$, and so the above two inequalities imply $R_{\mathcal{D}'}(p^{m'}) < R_{\mathcal{D}'}(p^m)$, which in turn implies that $m \notin C_{\mathcal{D}}(M)$.

[22] This ignores actions that produce identical payoffs, which is without loss of generality.

□

Proof of Corollary 4. Since $c < \bar{c}$, there exists $\bar{p} < 1$ and $\underline{p} > 0$ such that for both types of DMs, $\bar{a}$ and $\underline{a}$ maximize posterior expected utility for posteriors $p(\theta_h) > \bar{p}$ and $p(\theta_h) < \underline{p}$, respectively. Let $\mathcal{D}_1$ and $\mathcal{D}_2$ denote the decision problem of type-1 and type-2 DMs, respectively. Note that if $p^{\overline{m}}, p^{\underline{m}}$ satisfy

$$\begin{aligned} p^{\overline{m}} &> \bar{p} \\ p^{\underline{m}} &< \underline{p} \\ (\bar{k} - \underline{k})(1 - p^{\overline{m}}) &< (\bar{v} - \underline{v})p^{\underline{m}} \\ (\bar{k} - \underline{k} + c)(1 - p^{\overline{m}}) &> (\bar{v} - \underline{v} - c)p^{\underline{m}} \end{aligned}$$

then $I_{\mathcal{D}_1}(\overline{m}) > I_{\mathcal{D}_1}(\underline{m})$ and $I_{\mathcal{D}_2}(\overline{m}) < I_{\mathcal{D}_2}(\underline{m})$, and so polarization occurs. The above conditions simplify to

$$\begin{aligned} p^{\underline{m}} &\in \left(\frac{(\bar{k} - \underline{k})(1 - p^{\overline{m}})}{\bar{v} - \underline{v}}, \min\left\{ \frac{(\bar{k} - \underline{k} + c)(1 - p^{\overline{m}})}{\bar{v} - \underline{v} - c}, \underline{p} \right\} \right) \\ p^{\overline{m}} &\in \left(\min\left\{ 1 - \frac{(\bar{v} - \underline{v})p^{\underline{m}}}{\bar{k} - \underline{k}}, \bar{p} \right\}, 1 - \frac{(\bar{v} - \underline{v} - c)p^{\underline{m}}}{\bar{k} - \underline{k} + c} \right) \end{aligned}$$

yielding the bounds with the desired properties. Note also that for $p^{\overline{m}}$ sufficiently large and $p^{\underline{m}}$ sufficiently small, the constraints $p^{\overline{m}} > \bar{p}, p^{\underline{m}} < \underline{p}$ are not binding, and so for $c \in (0, \bar{c})$, there exists $\overline{m}, \underline{m}$ satisfying the above conditions.

□

Proof of Proposition 5. Follows directly from Theorem 1.

□

Proof of Proposition 6. Follows directly from Theorem 3.

□

Proof of Corollary 5. Suppose that a is strictly diversified with respect to A. There exists an action $\tilde{a}$ that is C-diversified with respect to A satisfying $u(a, \theta) < u(\tilde{a}, \theta) < \bar{u}^A(\theta)$ for all θ. Let $\overline{M}$ denote any set of interior models for which $\sum_\theta u(\tilde{a}, \theta)p^m(\theta|s) < \max_{a \in A} \sum_\theta u(a, \theta)p^m(\theta|s)$ for all $m \in \overline{M}$. To see that such a set exists, let $a^* \in A$ be any action that delivers the maximal payoff some state θ^*; there exists an $\epsilon > 0$ s.t. for any m^* s.t. for any m^* satisfying $p^{m^*}(\theta^*|s) > 1 - \epsilon$, $\sum_\theta u(\tilde{a}, \theta)p^{m^*}(\theta^*|s) < \max_{a \in A} \sum_\theta u(a, \theta)p^{m^*}(\theta^*|s)$ since $u(\tilde{a}, \theta^*) < \bar{a}^A(\theta^*) = u(a^*, \theta^*)$.

By construction, for $\mathcal{C}'$ represented by $(u, \overline{M})$, we have $\tilde{a} \notin \mathcal{C}'(A \cup \{\tilde{a}\})$. By Proposition 6, for any $M \supseteq \overline{M}$ and $\mathcal{C}$ represented by (u, M), $\tilde{a} \notin \mathcal{C}(A \cup \{\tilde{a}\})$, which implies $a \notin \mathcal{C}(A \cup \{a\})$ as desired.

□

Proof of Lemma 1. For some C-diversified action $a \in H_A$ and $\mathcal{C}$ represented by (u, M)

given signal s, $a \in \mathcal{C}(A \cup \{a\})$ implies that $\min_{m \in M} R_{\mathcal{D}'}(p^m(\cdot|s)) = k_a$ for $\mathcal{D}' = (u, A \cup \{a\})$. For $a_d \in H_a$ such that $a_d \in \mathcal{C}(A \cup \{a_d\})$ and a_d minimizes k_{a_d}, it must be the case that there exists some $a^* \in A$ such that $\{a^*, a_d\} \subseteq \mathcal{C}(A \cup \{a_d\})$, which in turn implies that

$$\begin{aligned}
K_A(M|s) &= k_{a_d} \\
&= \min_{m \in M} R_{\mathcal{D}'}(p^m(\cdot|s)) \\
&= \min_{m \in M} \left\{ \sum_\theta \max_{a' \in A \cup \{a_d\}} u(a', \theta) p^m(\theta|s) - \max_{a \in A \cup \{a_d\}} \sum_\theta u(a, \theta) p^m(\theta|s) \right\} \\
&= \min_{m \in M} \left\{ \sum_\theta \max_{a' \in A \cup \{a_d\}} u(a', \theta) p^m(\theta|s) - \max_{a \in A} \sum_\theta u(a, \theta) p^m(\theta|s) \right\} \\
&= \min_{m \in M} \left\{ \sum_\theta \max_{a' \in A} u(a', \theta) p^m(\theta|s) - \max_{a \in A} \sum_\theta u(a, \theta) p^m(\theta|s) \right\} \\
&= \min_{m \in M} R_{\mathcal{D}}(p^m(\cdot|s))
\end{aligned}$$

where the third line follows form the fact that $\{a^*, a_d\} \subseteq \mathcal{C}(A \cup \{a_d\})$ and so both a^* and a_d maximize expected utility for some $m \in \arg\min R_{\mathcal{D}'}(p^m(\cdot|s))$, and the fourth line follows from the fact that $u(a_d, \theta) \leq \overline{u}^A(\theta)$ for all θ.

□

Proof of Proposition 7. To see that for any $m^* \in M^*$, p^{m^*} must lie in the interior of $(\max\{0, p^{m_T} - \sqrt{v}\}, \min\{1, p^{m_T} + \sqrt{v}\})$, suppose not: any sender who proposed some m^* incurs negative payoffs, and so can profitably deviate by proposing m' such that $p^{m'}$=$\max\{0, p^{m_T} - \sqrt{v}\}$, which guarantees a non-negative payoff.

To see that for any $m^* \in M^*$, p^{m^*} cannot lie strictly in the interior of $(\max\{0, p^{m_T} - \sqrt{v}\}, \min\{1, p^{m_T} + \sqrt{v}\})$, suppose not. It must be the case that all models in M^* lie strictly in the interior of $(\max\{0, p^{m_T} - \sqrt{v}\}, \min\{1, p^{m_T} + \sqrt{v}\})$; if not, then by Proposition 1 and the assumption that $R_{\mathcal{D}}(p)$ is non-constant in a neighborhood around p^{m_T}, the receiver would not be indifferent between the models in M^*, a contradiction. Since senders receive strictly positive profits by proposing a model in M^*, it must be the case in such an equilibrium that all senders propose a model in M^*. Let $\underline{p} = \min\{p^{m_T}, \{p^{m^*} : m^* \in M^*\}\}$, $\overline{p} = \max\{p^{m_T}, \{p^{m^*} : m^* \in M^*\}\}$. Consider m', m'' satisfying $p^{m'} = \underline{p} - \epsilon$, $p^{m''} = \overline{p} + \epsilon$, for $\epsilon > 0$. By Proposition 1 and the assumption that $R_{\mathcal{D}}(p)$ is non-constant in a neighborhood around p^{m_T}, the receiver must find at least one of m', m'' strictly more decisive than any model in M^*. Taking $\epsilon \to 0$, any sender has a profitable deviation to either m' or m''.

□

Proof of Proposition 8. Necessity follows from Proposition 2. To show sufficiency, note that following the proof of Proposition 2, there exists a concave $\tilde{R} : \mathcal{M} \to \mathbb{R}$ such that $\tilde{R}(m) = 0$ whenever $p^m = \delta_\theta$ for $\theta \in \Theta$ such that $V = -\tilde{R}$ represents C. Again following the proof of Proposition 2, the function $R : \Delta(\Theta) \to \mathbb{R}$ satisfying $R(p^m) = \tilde{R}(m)$ is well defined, and is concave with $R(\delta_\theta) = 0$ for any $\theta \in \Theta$.

All that remains is to show that there exists a proper prediction problem $\mathcal{D}$ such that $R_{\mathcal{D}}(p) = R(p)$. By Theorem 1 of Gneiting and Raftery, there exists a proper scoring rule u where for $G \equiv -R$, $u(p,\theta) = G(p) - \sum_\theta G^*(p,\theta)p(\theta) + G^*(p,\theta)$ where $G^*(p,\theta)$ satisfies $G(q) \geq G(p) + \sum_\theta G^*(p,\theta)(q(\theta) - p(\theta))$ for all $p, q \in \Delta(\Theta)$. This implies that $u(\delta_\theta, \theta) = G(\delta_\theta)$ and $\sum_\theta u(p,\theta)p(\theta) = G(p)$. Since u is proper, we have

$$\begin{aligned} R_{\mathcal{D}}(p) &= \sum_\theta u(\delta_\theta, \theta)p(\theta) - \sum_\theta u(p,\theta)p(\theta) \\ &= \sum_\theta G(\delta_\theta)p(\theta) - G(p) \\ &= R(p) \end{aligned}$$

where the last step follows from the fact that $G(\delta_\theta) = 0$ and $G = -R$ by construction.

□

Proof of Proposition 9. Suppose Θ_k is not decision-relevant. Let $u_{-k} : \Theta_{-k} \times A \to \mathbb{R}$ satisfy $u_{-k}(a, \theta_{-k}) = u(a, \theta)$, which is well-defined since $u(a,\theta)$ is constant in θ_k. Let $I^{-k}_{\mathcal{D}}(\cdot|s)$ denote the decisiveness of a model defined on state space Θ_{-k} given a decision problem $\mathcal{D}$ defined on Θ_{-k}. Letting $\mathcal{D}_{-k} = (A, u_{-k})$, note that since Θ_k is not decision-relevant, for any $m, m' \in M$, $I_{\mathcal{D}}(m|s) \geq I_{\mathcal{D}}(m'|s) \iff I^{-k}_{\mathcal{D}_{-k}}(m_{-k}|s) \geq I^{-k}_{\mathcal{D}_{-k}}(m'_{-k}|s)$.

Now toward a contradiction, suppose that for all $m \in C_{\mathcal{D}}(M)$, m does not neglect k. By hypothesis, we know that $I_{\mathcal{D}}(m|s) > I_{\mathcal{D}}(m_\varnothing|s)$, which in turn implies that $I^{-k}_{\mathcal{D}_{-k}}(m_{-k}|s) > I^{-k}_{\mathcal{D}_{-k}}(m_\varnothing|s)$. By hypothesis, there exists some $m^* \in M$ neglecting k such that $\lambda m^*_{-k} + (1 - \lambda)m_\varnothing = m_{-k}$; Corollary 1 implies that $I^{-k}_{\mathcal{D}_{-k}}(m^*_{-k}|s) \geq I^{-k}_{\mathcal{D}_{-k}}(m_{-k}|s)$, which in turn implies that $I_{\mathcal{D}}(m^*|s) \geq I_{\mathcal{D}}(m|s)$, a contradiction.

□

Proof of Proposition 10 Without loss of generality, suppose that $\rho(\theta_h) \leq p^{m_T}(\theta_h|s) < p^{m_T}(\theta_h|s')$. This implies that for some k, k' $s_k = l$ and $s_{k'} = h$, and that for some k'', $s'_{k''} = h$. Consider the following two cases:

Case 1: $s'_k = h$ for all k. Note that in this case, $P_{M|s'} \subseteq P_{M|s}$, and so by Proposition 6, $K_A(M|s) \leq K_A(M|s')$.

Case 2: $s'_k = l$ for some k. Note that in this case, $P_{M|s'} = P_{M|s}$, and so $K_A(M|s) \leq K_A(M|s')$.

This implies that $K_A(M|s) - K_A(M|s') \leq 0 < K_A(\{m_T\}|s) - K_A(\{m_T\}|s')$.

□

Proof of Proposition 11. Note that

$$\begin{aligned} I_{\mathcal{D}}^{E}(m) &= \sum_{s} \max_{a \in A} \sum_{\theta} u(a,\theta) p^{m}(\theta|s) p^{m}(s) - \sum_{\theta} \max_{a' \in A} u(a',\theta)\rho(\theta) \\ &= \max_{a \in A} \sum_{s} \sum_{\theta} u(a,\theta) m(s|\theta)\rho(\theta) - \sum_{\theta} \max_{a' \in A} u(a',\theta)\rho(\theta) \end{aligned}$$

and so $I_{\mathcal{D}}^{E}(m)$ is the maximum over a family of functions that are linear in m. This implies that $I_{\mathcal{D}}^{E}$ is convex. Therefore, $I_{\mathcal{D}}^{E}(\lambda m + (1-\lambda)m') \leq \max\{I_{\mathcal{D}}^{E}(m), I_{\mathcal{D}}^{E}(m')\}$, and so $m, m' \notin C_{\mathcal{D}}^{E}(M) \implies \lambda m + (1-\lambda)m' \notin C_{\mathcal{D}}^{E}(M)$

□

Proof of Proposition 12. Follows immediately from the fact that $I_{\mathcal{D}}^{E}(m)$ respects the Blackwell order over $\mathcal{M}$, and that for $m = \lambda m_T + (1-\lambda)m_{\varnothing}$, m_T is a garbling of m and is therefore dominated by m in the Blackwell order.

□

Proof of Proposition 13. Note that if m recommends a^* from $\mathcal{D}'$ for some signal s, then m recommends a^* from $\mathcal{D}$ for s, since $u'(a^*,\theta) \leq u(a^*,\theta)$ for all θ, and $u'(a^*,\theta) = u(a^*,\theta)$ for all $\theta, a \neq a^*$. Similarly, if m' recommends $a \neq a^*$ from $\mathcal{D}$ for some signal s, then m' also recommends $a \neq a^*$ from $\mathcal{D}'$ for s. Let S^{a^*} denote the signals for which m recommends a^*. The above implies that

$$\begin{aligned} I_{\mathcal{D}'}^{E}(m') - I_{\mathcal{D}'}^{E}(m) =& \sum_{s} \max_{a \in A} \sum_{\theta} u'(a,\theta) p^{m'}(\theta|s) p^{m'}(s) - \sum_{s} \max_{a \in A} \sum_{\theta} u'(a,\theta) p^{m}(\theta|s) p^{m}(s) \\ =& \sum_{s} \max_{a \in A} \sum_{\theta} u(a,\theta) p^{m'}(\theta|s) p^{m'}(s) - \sum_{s} \max_{a \in A} \sum_{\theta} u(a,\theta) p^{m}(\theta|s) p^{m}(s) \\ &+ \sum_{s \in S^{a^*}} [u(a^*,\theta) - u'(a^*,\theta)] p^{m}(\theta|s) p^{m}(s) \\ \geq& \sum_{s} \max_{a \in A} \sum_{\theta} u(a,\theta) p^{m'}(\theta|s) p^{m'}(s) - \sum_{s} \max_{a \in A} \sum_{\theta} u(a,\theta) p^{m}(\theta|s) p^{m}(s) \\ =& I_{\mathcal{D}}^{E}(m') - I_{\mathcal{D}}^{E}(m) \end{aligned}$$

Since $m' \in C_{\mathcal{D}}^{E}(M), m \notin C_{\mathcal{D}}^{E}(M)$, $I_{\mathcal{D}}^{E}(m') - I_{\mathcal{D}}^{E}(m) \geq 0$ which in turn implies that $I_{\mathcal{D}'}^{E}(m') - I_{\mathcal{D}'}^{E}(m) \geq 0$ and $m \notin C_{\mathcal{D}'}^{E}(M)$.

□

Proof of Proposition 14. Suppose Θ_k is not decision-relevant. Let $u_{-k} : \Theta_{-k} \times A \to \mathbb{R}$ satisfy $u_{-k}(a,\theta_{-k}) = u(a,\theta)$, which is well-defined since $u(a,\theta)$ is constant in θ_k. Let $I_{\mathcal{D}}^{E,-k}(\cdot)$ denote the ex-ante decisiveness of a model defined on state space Θ_{-k} given a decision problem $\mathcal{D}$ defined on Θ_{-k}. Letting $\mathcal{D}_{-k} = (A, u_{-k})$, note that since Θ_k is not decision-relevant, for any $m, m' \in M$, $I_{\mathcal{D}}^{E}(m) \geq I_{\mathcal{D}}^{E}(m') \iff I_{\mathcal{D}_{-k}}^{E,-k}(m_{-k}) \geq I_{\mathcal{D}_{-k}}^{E,-k}(m'_{-k})$. Now toward a contradiction, suppose for all $m \in C_{\mathcal{D}}^{E}(M)$, m does not neglect k. By hypothesis, we know that there exists some $m^* \in M$ such that $\lambda m^*_{-k} + (1-\lambda)m_{\varnothing} = m_{-k}$ for $\lambda \in (0,1)$. By Proposition 12, we have $I_{\mathcal{D}_{-k}}^{E,-k}(m^*_{-k}) \geq I_{\mathcal{D}_{-k}}^{E,-k}(m_{-k})$, which in turn implies that $I_{\mathcal{D}}^{E}(m^*) > I_{\mathcal{D}}^{E}(m)$, a contradiction.

□

A.8 Proofs: Characterization and Identification Results

First, we establish a basic observation about our representation.

Lemma 2. $\mathcal{C}$ has a decisiveness-maximizing representation $(u, P) \in \mathcal{U} \times \mathcal{P}$ iff

$$\mathcal{C}(A) = \arg\max_{f \in A} U(f|A)$$

where

$$U(f|A) = \max_{p \in P} \sum_{\theta} \left[u(f(\theta)) - \max_{f' \in A} u(f'(\theta)) \right] p(\theta)$$

Proof. Suppose $f^* \in \mathcal{C}(A)$. By definition, we have $f^* \in \arg\max_{f \in A} \sum_{\theta} u(f(\theta)) p^*(\theta)$, for $p^* \in \arg\max_{p \in P} \max_{f \in A} \sum_{\theta} [u(f(\theta)) - \max_{f' \in A} u(f'(\theta))] \, p(\theta)$. But this implies that

$$f^* \in \arg\max_{f \in A} \sum_{\theta} \left[u(f(\theta)) - \max_{f' \in A} u(f'(\theta)) \right] p^*(\theta)$$

which in turn implies that

$$p^* \in \arg\max_{p \in P} \sum_{\theta} \left[u(f^*(\theta)) - \max_{f' \in A} u(f'(\theta)) \right] p(\theta)$$

So for any $f \in A$, we have

$$\begin{aligned} \sum_{\theta} \left[u(f(\theta)) - \max_{f' \in A} u(f'(\theta)) \right] p^*(\theta) &\leq \sum_{\theta} \left[u(f^*(\theta)) - \max_{f' \in A} u(f'(\theta)) \right] p^*(\theta) \\ &\leq \max_{p \in P} \sum_{\theta} \left[u(f^*(\theta)) - \max_{f' \in A} u(f'(\theta)) \right] p(\theta) \end{aligned}$$

and so $f^* \in \arg\max_{f \in A} U(f|A)$ as desired. Now suppose $f^* \in \arg\max_{f \in A} U(f|A)$; we have

$$f^* \in \arg\max_{f \in A} \max_{p \in P} \sum_{\theta} \left[u(f(\theta)) - \max_{f' \in A} u(f'(\theta)) \right] p(\theta)$$

This implies that $f^* \in \arg\max_{f \in A} \sum_{\theta} u(f(\theta)) p^*(\theta)$ for some

$$p^*(\theta) \in \arg\max_{p \in P} \sum_{\theta} \left[u(f^*(\theta)) - \max_{f' \in A} u(f'(\theta)) \right] p^*(\theta)$$

This in turn implies that

$$f^* \in \arg\max_{f\in A} \sum_\theta \left[u(f(\theta)) - \max_{f'\in A} u(f'(\theta))\right] p^*(\theta)$$

and so

$$\begin{aligned}
&\arg\max_{p\in P} \sum_\theta \left[u(f^*(\theta)) - \max_{f'\in A} u(f'(\theta))\right] p^*(\theta) \\
&= \arg\max_{p\in P} \sum_\theta \max_{f\in A} \left[u(f(\theta)) - \max_{f'\in A} u(f'(\theta))\right] p^*(\theta) \\
&= \mathcal{I}(P|A)
\end{aligned}$$

Therefore, $f^* \in \arg\max_{f\in A} \sum_\theta u(f(\theta))p^*(\theta)$ for some $p^* \in \mathcal{I}(P|A)$, and so $f^* \in \mathcal{C}(A)$. $\square$

Now, we establish basic results on $\mathcal{C}$ under Axiom 1.

Lemma 3. Suppose that f^* does not improve A. Under Axiom 1 (IINIA) we have the following:

1. If $f^* \notin \mathcal{C}(A \cup \{f^*\})$, then $\mathcal{C}(A \cup \{f^*\}) = \mathcal{C}(A)$
2. For $f' \neq f$, if $f \in \mathcal{C}(A \cup \{f^*\})$ and $f' \notin \mathcal{C}(A \cup \{f^*\})$, then $f' \notin \mathcal{C}(A)$
3. If $f \in \mathcal{C}(A)$ but $f \notin \mathcal{C}(A \cup \{f^*\}))$, then $f^* \in \mathcal{C}(A \cup \{f^*\})$

Proof. Assume that f^* does not improve A.

To show (1), take any $f \in \mathcal{C}(A\cup\{f^*\})$. Since $f^* \notin \mathcal{C}(A\cup\{f^*\})$, $f \neq f^*$ and therefore $f \in A$; by Axiom 1, $f \in \mathcal{C}(A)$. We therefore have $\mathcal{C}(A) \subseteq \mathcal{C}(A\cup\{f^*\})$. Now suppose $f \in \mathcal{C}(A)$, and towards a contradiction, suppose that $f \notin \mathcal{C}(A\cup\{f^*\})$. Note that it cannot be the case that $f' \in A$, $f' \in \mathcal{C}(A, \cup\{f^*\})$ since Axiom 1 would then imply that $f \in\in \mathcal{C}(A, \cup\{f^*\})$. Therefore, it must be the case that $f^* \in \mathcal{C}(A \cup \{f^*\})$, a contradiction. Therefore, $\mathcal{C}(A \cup \{f^*\}) \subseteq \mathcal{C}(A)$ and so $\mathcal{C}(A \cup \{f^*\}) = \mathcal{C}(A)$.

To show (2), suppose not; $f' \in \mathcal{C}(A)$. By Axiom 1, $f \in \mathcal{C}(A)$, and again applying Axiom 1, we have $f \notin \mathcal{C}(A \cup \{f^*\})$ since $f' \notin \mathcal{C}(A \cup \{f^*\})$, a contradiction.

To show (3), suppose not; $f^* \notin \mathcal{C}(A \cup \{f^*\})$. Then we must have $f' \in \mathcal{C}(A \cup \{f^*\})$, where $f' \neq f$ and $f' \in A$. By Axiom 1, this implies that $f' \in \mathcal{C}(A)$. But since $f \in \mathcal{C}(A)$, Axiom 1 in turn implies that $f \in \mathcal{C}(A \cup \{f^*\})$, a contradiction. $\square$

Say that $u : \Delta(Z) \to \mathbb{R}$ *represents the restriction of* $\mathcal{C}$ *to menus of constant acts* if $u(f_a(\theta)) \geq u(f_b(\theta))$ iff $f_a \in \mathcal{C}(\{f_a, f_b\})$ for all $f_a, f_b \in X_c$. The following lemma states that under Axioms 1-4, such a u exists and is linear.

Lemma 4. Suppose that Axioms 1, 2, 3, and 4 hold. Then there exists a linear $u : \Delta(Z) \to \mathbb{R}$ that represents the restriction of $\mathcal{C}$ to menus of constant acts.

Proof. Define the binary relation $\succeq^*$ on X_c as follows: for $f_a, f_b \in X_c$, $f_b \succeq f_a$ if $f_b \in \mathcal{C}(\{f_b, f_a\})$. Note that $\succeq^*$ is transitive. To see this, suppose that $f_c \succeq^* f_b$, $f_b \succeq^* f_a$. We have $f_b \in \mathcal{C}(\{f_b, f_a\})$, $f_c \in \mathcal{C}(\{f_c, f_b\})$. If $f_a \in \mathcal{C}(\{f_c, f_b, f_a\})$, then by Axiom 2 (Monotonicity), we have $f_b \in \mathcal{C}(\{f_c, f_b, f_a\})$ and subsequently $f_c \in \mathcal{C}(\{f_c, f_b, f_a\})$; if $f_a \notin \mathcal{C}(\{f_c, f_b, f_a\})$, since f_a does not improve f_b, Lemma 3 implies $f_c \in \mathcal{C}(\{f_c, f_b, f_a\})$. Therefore, we have $f_c \in \mathcal{C}(\{f_c, f_b, f_a\})$; since f_b does not improve f_c Axiom 1 (IINIA) implies that $f_c \in \mathcal{C}(\{f_c, f_a\}) \implies f_c \succeq^* f_a$.

So $\succeq^*$ is a complete and transitive preference relation on X_c that agrees with the restriction of $\mathcal{C}$ to menus of constant acts. Furthermore, $\succeq^*$ inherits continuity and independence properties from Axiom 3 (Mixture Continuity) and Axiom 4 (Mixture Independence): in particular, $\succeq^*$ satisfies

1. For $f_a, f_b, f_c \in X_c$, The sets $\{\alpha \in [0,1] : \alpha f_a + (1-\alpha) f_b \succeq^* f_c\}$ and $\{\alpha \in [0,1] : f_c \succeq^* \alpha f_a + (1-\alpha) f_b\}$ are closed.
2. For $f_a, f_b, f_c \in X_c$, $\alpha \in (0,1)$, $f_a \succeq^* f_b \iff \alpha f_a + (1-\alpha) f_c \succeq^* \alpha f_b + (1-\alpha) f_c$.

Therefore, by the Expected Utility Theorem (von-Neumann and Morgenstern 1994) there exists a linear $u : \Delta(Z) \to \mathbb{R}$ that represents the restriction of $\mathcal{C}$ to menus of constant acts. □

Say that Axiom 7 *fails for state* θ^* if exists f_a, f_b, f_c, with $f_a = \mathcal{C}(f_a, f_b)$, $f_b = \mathcal{C}(f_b, f_c)$, such that for all $\lambda \in (0,1)$, for act g satisfying $g_\lambda(\theta) = q(\lambda a + (1-\lambda) b)$ for $\theta = \theta^*$ and $g_\lambda(\theta) = q(c)$ for $\theta \neq \theta^*$, there exists a menu A that neither improves nor is improved by $g(\theta^*)$ for which $f_b \in A$ but also $g \in \mathcal{C}(A)$.

For any collection of states $\overline{\Theta} \in 2^\Theta$, we say that $\mathcal{C}$ *satisfies the support condition for* $\overline{\Theta}$ if for all $f_a, f_b, f_c \in X_c$ s.t. $f_a = \mathcal{C}(f_a, f_b)$, $f_b = \mathcal{C}(f_b, f_c)$, there exists $\lambda^* \in (0,1)$ such that for all $\lambda \in (0, \lambda^*)$, act g with $g(\theta) = \lambda f_a(\theta) + (1-\lambda) f_b(\theta)$ for $\theta \in \overline{\Theta}$, $g(\theta) = f_c(\theta)$ for $\theta \notin \theta^*$, and any A that neither improves nor is improved by $g(\theta)$ for $\theta \in \overline{\Theta}$, $g \notin \mathcal{C}(A)$ whenever $f_b \in A$.

Lemma 5. Suppose that Axioms 1, 2, 3, 4, and 7 hold. If $\mathcal{C}$ does not satisfy the support condition for $\overline{\Theta} \in 2^\Theta$, then $\{f \in A : \forall\, g \in A,\, g(\theta) \text{ does not improve } f(\theta)\, \forall\, \theta \in \overline{\Theta}\} \subseteq \mathcal{C}(A)$. If Axiom 7 fails for state θ^*, then $\{f \in A : \forall\, g \in A,\, g(\theta^*) \text{ does not improve } f(\theta^*)\} \subseteq \mathcal{C}(A)$.

Proof. By Lemma 4, there exists a linear $u : \Delta(Z) \to \mathbb{R}$ that represents the restriction of $\mathcal{C}$ to menus of constant acts. Additionally, since Axiom 6 holds, there exists $z, z' \in Z$ s.t. $u(z) > u(z')$. Take some $\overline{z} \in \arg\max_{z \in Z} u(z)$ and $\underline{z} \in \arg\min_{z \in Z} u(z)$; by linearity of u, we can without loss of generality take $u(\overline{z}) = 1$, $u(\underline{z}) = -1$.

For any $c \in [-1, 1]$, let $q(c) = \frac{1-c}{2} \circ \underline{z} + \frac{1+c}{2} \circ \overline{z}$; by linearity of u, $u(q(c)) = c$. let f_c denote the constant act with $f_c(\theta) = q(c)$ for all θ. Consider the constant act $f_{-1/2}$.We introduce the following notation for any $A \in \mathcal{A}$, $f \in A$:

- Let f^A_{max} denote the *maximal* act corresponding to A, satisfying $f^A_{max}(\theta) = q(\max_{f \in A} u(f(\theta)))$, and define the *minimal* act f^A_{min} analogously.

- Let $\overline{f}$ be the *normalized* act satisfying $\overline{f}(\theta) = q(\frac{1}{4}[u(f(\theta)) - u(f^A_{max}(\theta))])$; since $\frac{1}{4}[u(f(\theta)) - u(f^A_{max}(\theta))] \in [-1, 1]$ for any $f \in A$, $\overline{f}$ is well-defined. Let $\overline{A} = \{\overline{f} : f \in A\}$ collect the normalized acts in A.

We will first show that if $\mathcal{C}$ does not satisfy the support condition for $\overline{\Theta} \in 2^{\Theta}$, then $\{f \in A : \forall\, g \in A,\, g(\theta) \text{ does not improve } f(\theta)\, \forall\, \theta \in \overline{\Theta}\} \subseteq \mathcal{C}(A)$.

Claim 1. Suppose that $\mathcal{C}$ does not satisfy the support condition for $\overline{\Theta}$. Then for the act g satisfying $g(\theta) = q(0)$ for $\theta \in \overline{\Theta}$ and $g(\theta) = q(-1)$ for $\theta \notin \overline{\Theta}$, $g \in \mathcal{C}(g, f_0)$.

Proof of Claim 1. Since $\mathcal{C}$ does not satisfy the support condition for $\overline{\Theta}$, there there exists f_a, f_b, f_c, with $a > b > a$, such that for all $\lambda \in (0, 1)$, for act g_λ satisfying $g_\lambda(\theta) = q(\lambda a + (1-\lambda)b)$ for $\theta \in \overline{\Theta}$ and $g_\lambda(\theta) = q(c)$ for $\theta \notin \overline{\Theta}$, there exists a menu A for which $u \circ f^A_{max} = \lambda a + (1-\lambda)b$ for which $f_b \in A$ but also $g_\lambda \in \mathcal{C}(A)$. By Axiom 4 (Mixture Independence), we can without loss of generality take $b = 0$ and $a - c < 1$. Now, note that for $\phi = -c \in (0, 1)$, by Axiom 4 (Mixture Independence) we have $\phi g_{\lambda_k} + (1-\phi) f_0 \in \mathcal{C}(\phi A + (1-\phi) f_0)$.

Noting that $u \circ f^{\phi A + (1-\phi) f_0}_{min} > c$, the above implies that for all $\lambda < \phi$, there exists an act $\tilde{g}_\lambda$ satisfying $\tilde{g}_\lambda(\theta) = q(\lambda a)$ for $\theta \in \overline{\Theta}$ and $\tilde{g}_\lambda(\theta) = q(\phi c)$ for $\theta \notin \overline{\Theta}$ and a menu A satisfying $u \circ f^A_{max} = \lambda a$ and $u \circ f^A_{min} > c$, for which $f_0 \in A$ but also $\tilde{g}_\lambda \in \mathcal{C}(A)$.

Now take a sequence $\lambda_k \to 0$, where $\lambda_k \in (0, \phi)$ for all k. The above implies the existence of acts $\tilde{g}_{\lambda_k}$ satisfying

$$\tilde{g}_{\lambda_k}(\theta) = \begin{cases} q(\lambda_k a) & \theta \in \overline{\Theta} \\ q(\phi c) & \text{otherwise} \end{cases}$$

and a sequence of menus A_k satisfying $u \circ f^{A_k}_{max} = \lambda_k a$ and $f^{A_k}_{min} > c$, where $f_0 \in A_k$ but $\tilde{g}_{\lambda_k} \in \mathcal{C}(A_k)$. Let $B_k = A_k \setminus \tilde{g}_{\lambda_k}$. Let $\hat{g}_{\lambda_k}$ denote the act satisfying

$$\hat{g}_{\lambda_k}(\theta) = \begin{cases} q(\lambda_k a) & \theta \in \overline{\Theta} \\ q(\phi c + \lambda_k a) & \text{otherwise} \end{cases}$$

and for all $\theta \neq \theta^*$, let $h_{\theta,k}$ denote the acts satisfying

$$h_{\theta,k}(\theta') = \begin{cases} q(\lambda_k a) & \theta' = \theta \\ q(-1 + \lambda_k a) & \text{otherwise} \end{cases}$$

Let $H_k = \{h_{\theta,k}\}_{\theta \notin \overline{\Theta}}$. Note that since $u \circ \hat{g}_{\lambda_k} \geq u \circ \tilde{g}_{\lambda_k}$, by Axiom 2 (Monotonicity), $\hat{g}_{\lambda_k} \in \mathcal{C}(B_k \cup \{\tilde{g}_{\lambda_k}\})$. Furthermore, since $a - c < 1$, by construction $u(h_{\theta,k}(\theta')) \leq c$ for $\theta' \neq \theta$, and so since $u \circ f^A_{min} \geq c$ and therefore $u \circ f^{B_k \cup \{\hat{g}_{\lambda_k}\}}_{min} \geq c$ each $h_{\theta,k}$ is dominated by an act from $B_k \cup \{\hat{g}_{\lambda_k}\}$. Axiom 2 (Monotonicity) then implies that $\hat{g}_{\lambda_k} \in \mathcal{C}(B_k \cup H_k \cup \{\hat{g}_{\lambda_k}\})$; Axiom 1 (IINIA) subsequently implies $\hat{g}_{\lambda_k} \in \mathcal{C}(H_k \cup \{\hat{g}_{\lambda_k}, f_0\})$, noting that B_k does not improve H_k by construction.

Now, note that for all k, $\frac{1}{2}\hat{g}_{\lambda_k} + \frac{1}{2}f_{-\lambda_k a} = \dot{g}$, and $\frac{1}{2}h_{\theta,k} + \frac{1}{2}f_{-\lambda k} = h_\theta$, where

$$\dot{g}(\theta) = \begin{cases} q(0) & \theta \in \overline{\Theta} \\ q(1/2\phi c) & \text{otherwise} \end{cases} \qquad h_\theta(\theta') = \begin{cases} q(0) & \theta' = \theta \\ q(-1/2) & \text{otherwise} \end{cases}$$

Let $H = \{h_\theta\}_{\theta \notin \overline{\Theta}}$ By Axiom 4 (Mixture Independence), we therefore have $\dot{g} \in \mathcal{C}(H \cup \{\dot{g}, \frac{1}{2}f_0 + \frac{1}{2}f_{-\lambda_k a}\})$, which can be rewritten as $\dot{g} \in \mathcal{C}(H \cup \{\dot{g}, \lambda_k f_{-1/2a} + (1-\lambda_k)f_0\})$. Since this holds for $\lambda_k \to 0$, Axiom 3 (Mixture Continuity) implies that $\dot{g} \in \mathcal{C}(H \cup \{\dot{g}, f_0\})$, and Axiom 1 (IINIA) in turn implies $\dot{g} \in \mathcal{C}(\{\dot{g}, f_0\})$, noting that H does not improve f_0 by construction.

Now, note that for act g satisfying $g(\theta) = 0$ for $\theta = \theta^*$, $g(\theta) = -1$ for $\theta \neq \theta^*$, $\dot{g} = (\frac{1}{2}\phi c)g + (1 - \frac{1}{2}\phi c)f_0$ and therefore by Axiom 4 (Mixture Independence) we have $\dot{g} \in \mathcal{C}(\{\dot{g}, f_0\}) \implies g \in \mathcal{C}(\{g, f_0\})$ as desired.

△

Claim 2. For $A \in \mathcal{A}$, $g \in A$, suppose that for all $f \in A \setminus \{f\}$, $g(\theta)$ improves $f(\theta)$ for all $\theta \in \overline{\Theta}$. Then $g \in \mathcal{C}(A)$.

Proof of Claim 2. Take $g \in A$ where $g(\theta)$ improves $f(\theta)$ for all $f \in A \setminus \{g\}$ and $\theta \in \overline{\Theta}$. Consider the normalized menu $\overline{A}$; we have $u(\overline{f}(\theta)) < u(\overline{g}(\theta)) = 0$ for all $\overline{f} \in \overline{A}$, $\theta \in \overline{\Theta}$. Let $v = \max_{\theta \in \overline{\Theta}} \max_{\overline{f} \in \overline{A} \setminus \{\overline{g}\}} u(\overline{f}(\theta)) < 0$, and let $\gamma = \frac{1}{1-v} \in (0,1)$. Consider act h satisfying

$$h(\theta) = \begin{cases} q(0) & \theta \in \overline{\Theta} \\ q(-1) & \text{otherwise} \end{cases}$$

For any act f, let $f^v = \frac{1}{2}(\gamma f + (1-\gamma)h) + \frac{1}{2}f_{-\frac{v}{v-1}}$, and for any menu A, let $A^v = \{f^v : f \in A\}$. By construction, we have $u \circ \overline{f}^v \leq 0$ for all $f \in A \setminus \{g\}$, whereas $u(\overline{g}^v(\theta)) = \frac{v}{v-1} > 0$ for all $\theta \in \overline{\Theta}$ and $u(\overline{g}^v(\theta)) \leq 0$ for all $\theta \notin \overline{\Theta}$.

Now by Claim 1, for the act $\tilde{g}$ satisfying $\tilde{g}(\theta) = q(0)$ for $\theta \in \overline{\Theta}$ and $\tilde{g}(\theta) = q(-1)$ for $\theta \notin \overline{\Theta}$, we have $\tilde{g} \in \mathcal{C}(\{\tilde{g}, f_0\})$. Noting that $\overline{A}^v \setminus \{\overline{g}^v\}$ does not improve f_0, and that $\overline{g}^v$ dominates $\tilde{g}$, we have

$$\begin{aligned} \tilde{g} \in \mathcal{C}(\{\tilde{g}, f_0\}) &\implies \tilde{g} \in \mathcal{C}(\{\tilde{g}, f_0\} \cup \overline{A}^v \setminus \{\overline{g}^v\}) \qquad \text{by Axioms 1, 2 (IINIA, Monotonicity)} \\ &\implies \overline{g}^v \in \mathcal{C}(\{f_0\} \cup \overline{A}^v) \qquad \text{by Axiom 2 (Monotonicity)} \\ &\implies \overline{g}^v \in \mathcal{C}(\overline{A}^v) \qquad \text{by Axiom 1 (IINIA)} \end{aligned}$$

where for the the last step we note that by construction f_0 does not improve $\overline{A}^v$. By Axiom 4 (Mixture Independence) and Axiom 2 (Monotonicity), $\overline{g}^v \in \mathcal{C}(\overline{A}^v) \implies g \in \mathcal{C}(A)$ as desired.

△

Now, Take any $A \in \mathcal{A}$, $f \in A$, and suppose for all $g \in A$, $g(\theta)$ does not improve $f(\theta)$ for all $\theta \in \overline{\Theta}$. By Axiom 4 (Mixture Independence), we can without loss of generality assume $u(f(\theta)) < 1$ for all θ. For any $\alpha \in (0,1]$, $\alpha f_1 + (1-\alpha)f$ improves $A \setminus \{f\}$, and so

Claim 2 implies that

$$\alpha f_1 + (1-\alpha) f \in \mathcal{C}(\{\alpha f_1 + (1-\alpha) f\} \cup A \setminus \{f\})$$

Since the above holds for all $\alpha \in (0,1]$, Axiom 3 (Mixture Continuity) implies that $f \in \mathcal{C}(A)$ as desired; we have $\{f \in A : \forall\, g \in A,\, g(\theta) \text{ does not improve } f(\theta)\, \forall\, \theta \in \overline{\Theta}\} \subseteq \mathcal{C}(A)$.

Now, suppose that Axiom 7 fails for state θ^*. Note that by definition, this holds only if $\mathcal{C}$ does not satisfy the support condition for $\{\theta^*\}$; the above then implies that $\{f \in A : \forall\, g \in A,\, g(\theta^*) \text{ does not improve } f(\theta^*)\} \subseteq \mathcal{C}(A)$. □

Lemma 6. Suppose that Axioms 1, 2, 3, 4, and 7 hold, and suppose that $\mathcal{C}$ does not satisfy the support condition for the collection of states $\mathcal{F} \subseteq 2^\Theta$ and satisfies the support condition for $2^\Theta \setminus \mathcal{F}$, where $\mathcal{F}$ contains $\{\theta^*\}$ for some $\theta^* \in \Theta$. Then

$$\mathcal{C}(A) = \bigcup_{\overline{\Theta} \in \mathcal{F}} \{f \in A : \forall\, g \in A,\, g(\theta) \text{ does not improve } f(\theta)\, \forall\, \theta \in \overline{\Theta}\}.$$

Proof. By Lemma 4, there exists a linear $u : \Delta(Z) \to \mathbb{R}$ that represents the restriction of $\mathcal{C}$ to menus of constant acts. Additionally, since Axiom 6 holds, there exists $z, z' \in Z$ s.t. $u(z) > u(z')$. Take some $\overline{z} \in \arg\max_{z \in Z} u(z)$ and $\underline{z} \in \arg\min_{z \in Z} u(z)$; by linearity of u, we can without loss of generality take $u(\overline{z}) = 1$, $u(\underline{z}) = -1$.

For any $c \in [-1, 1]$, let $q(c) = \frac{1-c}{2} \circ \underline{z} + \frac{1+c}{2} \circ \overline{z}$; by linearity of u, $u(q(c)) = c$. let f_c denote the constant act with $f_c(\theta) = q(c)$ for all θ. Consider the constant act $f_{-1/2}$.We introduce the following notation for any $A \in \mathcal{A}$, $f \in A$:

- Let f^A_{max} denote the *maximal* act corresponding to A, satisfying $f^A_{max}(\theta) = q(\max_{f \in A} u(f(\theta)))$, and define the *minimal* act f^A_{min} analogously.
- Let $\overline{f}$ be the *normalized* act satisfying $\overline{f}(\theta) = q(\frac{1}{4}[u(f(\theta)) - u(f^A_{max}(\theta))])$; since $\frac{1}{4}[u(f(\theta)) - u(f^A_{max}(\theta))] \in [-1, 1]$ for any $f \in A$, $\overline{f}$ is well-defined. Let $\overline{A} = \{\overline{f} : f \in A\}$ collect the normalized acts in A.

Let g_θ denote the act satisfying $g_\theta(\theta') = q(0)$ for $\theta' = \theta$ and $g_\theta(\theta') = q(-1)$ for $\theta' \neq \theta$.

Claim 1. Suppose that $\mathcal{C}$ does not satisfy the support condition for the collection of states $\mathcal{F} \subseteq 2^\Theta$ and satisfies the support condition for $2^\Theta \setminus \mathcal{F}$, where $\mathcal{F}$ contains $\{\theta^*\}$ for some $\theta^* \in \Theta$. Then for all $\Theta^* \subseteq \Theta$ s.t. $\theta^* \in \Theta^*$ and $\Theta \setminus \Theta^* \notin \mathcal{F}$, for any $c < 0$, $f_c \notin \mathcal{C}(\{f_c\} \cup \{g_\theta\}_{\theta \in \Theta^*})$.

Proof of Claim 1. Towards a contradiction, suppose not: there exists $c < 0$ such that $f_c \in \mathcal{C}(\{f_c\} \cup \{g_\theta\}_{\theta \in \Theta^*})$.

First, consider the case where $\tilde{\Theta} \equiv \Theta \setminus \Theta^* = \varnothing$. By Axiom 2 (Monotonicity), for any $c' \in (c, 0)$, $f_{c'} \in \mathcal{C}(\{f_{c'}, f_c\} \cup \{g_\theta\}_{\theta \in \Theta^*})$. Also, by Lemma 5, we have $g_{\theta^*} \in \mathcal{C}(\{f_c\} \cup \{g_\theta\}_{\theta \in \Theta^*})$ and $g_{\theta^*} \in \mathcal{C}(\{f_{c'}, f_c\} \cup \{g_\theta\}_{\theta \in \Theta^*})$. Since by construction $f_{c'}$ does not improve $\{f_c\} \cup \{g_\theta\}_{\theta \in \Theta^*}$, Axiom 1 (IINIA) implies that $f_c \in \mathcal{C}(\{f_{c'}, f_c\} \cup \{g_\theta\}_{\theta \in \Theta^*})$, which implies a contradiction by

Axiom 2 (Monotonocity).

Now, consider the case where $\tilde{\Theta}$ is non-empty; by hypothesis, $\mathcal{C}$ satisfies the support condition for $\tilde{\Theta}$. This implies that there exists $a \in (0, -1/2c)$ such that for act $\tilde{h}$ satisfying $\tilde{h}(\theta) = q(a)$ for $\theta \in \tilde{\Theta}$, $\tilde{h}(\theta) = q(1/2c + a)$ for $\theta \notin \tilde{\Theta}$, we have $\tilde{h} \notin \mathcal{C}(\{\tilde{h}, f_0\} \cup \{\tilde{g}_\theta\}_{\theta \in \Theta^*})$, where $\tilde{g}_\theta(\theta') = q(a)$ for $\theta' = \theta$ and $\tilde{g}_\theta(\theta') = q(-1+a)$ for $\theta' \neq \theta$. By Axiom 4 (Mixture Independence) we have $\frac{1}{2}\tilde{h} + \frac{1}{2}f_{-a} \notin \mathcal{C}(\{\frac{1}{2}\tilde{h} + \frac{1}{2}f_{-a}, f_{-1/2a}\} \cup \{\frac{1}{2}\tilde{g}_\theta + \frac{1}{2}f_{-a}\}_{\theta \in \Theta^*})$. By Axiom 4 (Mixture Independence) this in turn implies that for act h satisfying

$$h(\theta) = \begin{cases} q(0) & \theta \in \tilde{\Theta} \\ q(1/2c) & \text{otherwise} \end{cases}$$

we have $h \notin \mathcal{C}(\{h, f_{-a}\} \cup \{\tilde{g}_\theta\}_{\theta \in \Theta^*})$, which in turn implies, by Axiom 1 (IINIA), that $h \notin \mathcal{C}(\{h, f_{-a}, f_c\} \cup \{\tilde{g}_\theta\}_{\theta \in \Theta^*})$. Note, however, since $h(\theta)$ dominates f_c, by Axiom 2 (Monotonicity), $f_c \in \mathcal{C}(\{f_c\} \cup \{\tilde{g}_\theta\}_{\theta \in \Theta^*}) \implies h \in \mathcal{C}(\{h, f_c\} \cup \{\tilde{g}_\theta\}_{\theta \in \Theta^*})$. Now, since f_{-a} does not improve $\{h, f_c\} \cup \{\tilde{g}_\theta\}_{\theta \in \Theta^*}$, by Lemma 3, we have $f_{-a} \in \mathcal{C}(\{h, f_{-a}, f_c\} \cup \{\tilde{g}_\theta\}_{\theta \in \Theta^*})$

Now, by Lemma 5, we have $g_{\theta^*} \in \mathcal{C}(\{h, f_c\} \cup \{g_\theta\}_{\theta \in \Theta^*})$ and also $g_{\theta^*} \in \mathcal{C}(\{h, f_{-a}, f_c\} \cup \{g_\theta\}_{\theta \in \Theta^*})$, and so noting again that f_{-a} does not improve $\{h, f_c\} \cup \{g_\theta\}_{\theta \in \Theta^*}$, Axiom 1 (IINIA) implies that $f_c \in \mathcal{C}(\{h, f_{-a}, f_c\} \cup \{g_\theta\}_{\theta \in \Theta^*})$. However, since $-a > 1/2c > c$, and we also have $f_a \in \mathcal{C}(\{h, f_{-a}, f_c\} \cup \{g_\theta\}_{\theta \in \Theta^*})$, by Axiom 2 (Monotonicity) we have a contradiction.

$\triangle$

Now suppose that $f \notin \bigcup_{\overline{\Theta} \in \mathcal{F}} \{f \in A : \forall\, g \in A,\ g(\theta) \text{ does not improve } f(\theta)\, \forall\, \theta \in \overline{\Theta}\}$. We will show that $f \notin \mathcal{C}(A)$, which in conjunction with Lemma 5 completes the proof. Let $\Theta^* \equiv \{\theta \in \Theta \text{ s.t. } f^A_{max}(\theta) \text{ improves } f(\theta)\}$; by construction, we have $\theta^* \in \Theta^*$, and also that $\mathcal{C}$ satisfies the full support condition for $\tilde{\Theta} \equiv \Theta \setminus \Theta^*$.

Consider the normalized menu $\overline{A}$; we have $u(\overline{f}(\theta)) < 0$ for all $\theta \in \Theta^*$. Let $v = \max_{\theta \in \Theta^*} u(\overline{f}(\theta)) < 0$, and let $\gamma = \frac{1}{1-v} \in (0, 1)$. Consider act h satisfying

$$h(\theta) = \begin{cases} q(0) & \theta \in \Theta^* \\ q(-1) & \text{otherwise} \end{cases}$$

For any act f, let $f^v = \gamma f + (1-\gamma)h$, and for any menu A, let $A^v = \{f^v : f \in A\}$. By construction, we have $u \circ \overline{f}^v \leq -\frac{v}{v-1} < 0$ for all $\theta \in \Theta^*$, and $u(f^{\overline{A}^v}_{max}(\theta)) = 0$ for all $\theta \in \Theta^*$. We have

$$\begin{aligned} & f_{-v/(v-1)} \notin \mathcal{C}(\{f_{-v/(v-1)}\} \cup \{g_\theta\}_{\theta \in \Theta^*}) && \text{by Claim 1} \\ \implies & \overline{f}^v \notin \mathcal{C}(\{\overline{f}^v\} \cup \{g_\theta\}_{\theta \in \Theta^*}) && \text{by Axiom 2 (Monotonicity)} \\ \implies & \overline{f}^v \notin \mathcal{C}(\overline{A}^v \cup \{g_\theta\}_{\theta \in \Theta^*}) && \text{by Axiom 1 (IINIA)} \\ \implies & \overline{f}^v \notin \mathcal{C}(\overline{A}^v) \end{aligned}$$

where the last line follows from Axiom 1 (IINIA) and Axiom 2 (Monotonicity), noting that For each $\theta \in \Theta^*$ there exists an $\overline{g}^v \in \overline{A}^v$ that dominates g_θ. Axiom 4 (Mixture Independence) then implies that $f \notin \mathcal{C}(A)$ as desired. $\square$

A key object in our proof will be a real map on the set of negative utility acts $\varphi : \mathbb{R}^\Theta_- \to \mathbb{R}$. For $c \in \mathbb{R}$, let ξ_c be the constant utility act satisfying $\xi_c(\theta) = c$ for all θ. We say that φ is *C-additive* if for all $\xi \in \mathbb{R}^\Theta_-$ and constant utility acts $\xi_c \in \mathbb{R}^\Theta$ with $c < 0$, $\varphi(\xi + \xi_c) = \varphi(\xi) + \varphi(\xi_c)$. We say that φ is *sublinear* if it is convex and positively homogenous.

Lemma 7. If $\varphi : \mathbb{R}^\Theta_- \to \mathbb{R}$ is increasing, sublinear, and C-additive, then there exists a non-empty convex subset $\mathcal{L}$ of positive linear functionals on $\mathbb{R}^\Theta$ such that $\varphi(\xi) = \max_{L\in\mathcal{L}} L(\xi)$ for all $\xi \in \mathbb{R}^\Theta_-$, and $L(\xi_{-1}) = \varphi(\xi_{-1})$ for all $L \in \mathcal{L}$.

Proof. By sublinearity of φ, proposition G.3 in Ok (2007) implies that there exists a nonempty convex subset $\widetilde{\mathcal{L}}$ of linear functionals on $\mathbb{R}^\Theta$ such that $\varphi(\xi) = \max_{L\in\widetilde{\mathcal{L}}} L(\xi)$ for all $\xi \in \mathbb{R}^\Theta_-$. Take any $\xi \in \mathbb{R}^\theta_-\_l$; we can find some $c < 0$ and $\xi' \in \mathbb{R}^\Theta_-$ s.t. $\xi' + \xi_c = \xi$. There exists a linear functional $L_{\xi'+\xi_c}$ on $\mathbb{R}^\Theta$ s.t. $\varphi \geq L_{\xi'+\xi_c}$ on $\mathbb{R}^\Theta_-$ and $\varphi(\xi' + \xi_c) = L_{\xi'+\xi_c}(\xi' + \xi_c)$. By C-additivity of φ and linearity of $\mathcal{L}$, we have

$$\varphi(\xi') + \varphi(\xi_c) = \mathcal{L}_{\xi'+\xi_c}(\xi') + \mathcal{L}_{\xi'+\xi_c}(\xi_c)$$

Since $L_{\xi'+\xi_c}(\xi') \leq \varphi(\xi')$ and $L_{\xi'+\xi_c}(\xi_c) \leq \varphi(\xi_c)$, the above implies $L_{\xi'+\xi_c}(\xi_c) = \varphi(\xi_c) \implies L_\xi(\xi_c) = \varphi(\xi_c)$, which in turn implies $L_\xi(\xi_{-1}) = \varphi(\xi_{-1})$ by linearity of L and homogeneity of φ

Now, take $\mathcal{L} = co\{L_\xi : \xi \in \mathbb{R}^\Theta_{--}\}$. By construction, $\mathcal{L}$ is a non-empty convex subset of linear functionals that satisfies $L(\xi_{-1}) = \varphi(\xi_{-1})$ for all $L \in \mathcal{L}$. To see that each $L \in \mathcal{L}$ is positive, take any $\xi \leq 0$; since φ is increasing we have $L(\xi) \leq \varphi(\xi) \leq \varphi(0) = 0$, where the last equality follows from the fact that φ is sublinear.

Finally to see that $\varphi(\xi) = \max_{L\in\mathcal{L}} L(\xi)$ for all $\xi \in \mathbb{R}^\Theta_-$, take any $\xi \in \mathbb{R}^\Theta_-$. By construction, $L(\xi) \leq \varphi(\xi)$ for all $L \in \mathcal{L}$. Also, for $c < 0$, we have $L_{\xi+\xi_c} \in \mathcal{L}$, with

$$\begin{aligned}
\mathcal{L}_{\xi+\xi_c}(\xi + \xi_c) &= \varphi(\xi + \xi_c) \\
\mathcal{L}_{\xi+\xi_c}(\xi) + \mathcal{L}_{\xi+\xi_c}(\xi_c) &= \varphi(\xi) + \varphi(\xi_c) \qquad \text{by C-additivity} \\
\mathcal{L}_{\xi+\xi_c}(\xi) &= \varphi(\xi) \qquad \text{since } \mathcal{L}_{\xi+\xi_c}(\xi_{-1}) = \varphi(\xi_{-1})
\end{aligned}$$

and so $\varphi(\xi) = \max_{L\in\mathcal{L}} L(\xi)$. □

Lemma 8. If $\varphi : \mathbb{R}^\Theta_- \to \mathbb{R}$ is increasing, sublinear, and C-additive, with $\varphi(\xi_{-1}) < 0$, then there exists a non-empty, closed $P \subset \Delta(\Theta)$ s.t.

$$\varphi(\xi) = \max_{p\in P} \sum_\theta \xi(\theta)p(\theta)$$

for all $\xi \in \mathbb{R}^\Theta_-$.

Proof. By Lemma 7, non-empty convex subset $\mathcal{L}$ of positive linear functionals on $\mathbb{R}^\Theta$ such that $\varphi(\xi) = \max_{L\in\mathcal{L}} L(\xi)$ for all $\xi \in \mathbb{R}^\Theta_-$, and $L(\xi_{-1}) = \varphi(\xi_{-1})$ for all $L \in \mathcal{L}$. We can associate each $L \in \mathcal{L}$ with $q_L \in \mathbb{R}^\Theta$ such that $L(\xi) = \max q_L : L \in \mathcal{L} \sum_\theta q_L(\theta)\xi(\theta)$ for all $\xi \in \mathbb{R}^\Theta$. Since L is positive, we have $q_L \geq 0$; furthermore, $L(\xi_{-1}) = \varphi(\xi_{-1}) < 0 \implies \sum_\theta q_L(\theta) = -\varphi(\xi_{-1}) > 0$.

For each L, define $p_L = -q_L/\varphi(\xi_{-1})$; by the above, $p_L \geq 0$ with $\sum_\theta p_L(\theta) = 1$; we have $\varphi(\xi) = \max_{p_L \in L \in \mathcal{L}} \sum_\theta \xi(\theta)p_L(\theta)$. Let $P = cl(\{p_L : L \in \mathcal{L}\})$; we have $p \geq 0$, $\sum_\theta p(\theta) = 1$ and so therefore $P \subseteq \Delta(\Theta)$, and furthermore, since $\max_{p\in P} \sum_\theta \xi(\theta)p(\theta) = \max_{p_L\in L\in\mathcal{L}} \sum_\theta \xi(\theta)p_L(\theta)$, we have $\varphi(\xi) = \max_{p\in P} \sum_\theta \xi(\theta)p(\theta)$. □

We are now ready to prove Theorem 1.

Theorem 1. $\mathcal{C}$ satisfies Axioms 1–5 if and only if it has a decisiveness-maximizing representation (u, P). $\mathcal{C}$ additionally satisfies Axioms 6 and 7 if and only if u and P satisfy Assumptions 5 and 6, respectively, and for any (u', P') representing $\mathcal{C}$, there exists constants $\alpha > 0$, β such that $u' = \alpha u + \beta$, and $ext(P') = ext(P)$.

Proof. We first prove the characterization result. Necessity of the axioms follow directly from the definition of the representation. The proof of sufficiency proceeds in four steps.

Step 1: Construction of $u : \Delta(Z) \to \mathbb{R}$.

By Lemma 4, there exists a linear $u : \Delta(Z) \to \mathbb{R}$ that represents the restriction of $\mathcal{C}$ to menus of constant acts; that is if $u(f_a(\theta)) \geq u(f_b(\theta))$ iff $f_a \in \mathcal{C}(\{f_a, f_b\})$ for all $f_a, f_b \in X_c$. In particular, since u is linear, we have $u \in \mathcal{U}$.

Step 2: Characterization without Axioms 6 and 7.

We will now show that in the case where Axiom 6 (Non-Triviality) does not hold, $\mathcal{C}$ has a decisiveness-maximizing representation. Suppose that Axiom 6 (Non-Triviality) does not hold. This implies that for all $z, z' \in Z$, $u(z) = u(z')$. By linearity of u, we then have that for $q, q' \in \Delta(Z)$, $u(q) = u(q')$, and so for any $f, f' \in X$, $f(\theta) \in \mathcal{C}(\{f(\theta), f'(\theta)\})$. Axiom 2 (Monotonicity) then implies that $\mathcal{C}(A) = A$ for all $A \in \mathcal{A}$. Note that by taking any $P \in \mathcal{P}$ we are done; $\mathcal{C}$ has a decisiveness-maximizing representation (u, P).

The remainder of the proof will deal with the case where Axiom 6 (Non-Triviality) holds. There exists $z, z' \in Z$ s.t. $u(z) > u(z')$; that is, u satisfies Assumption 5. Take some $\overline{z} \in \arg\max_{z\in Z} u(z)$ and $\underline{z} \in \arg\min_{z\in Z} u(z)$; by linearity of u, we can without loss of generality take $u(\overline{z}) = 1$, $u(\underline{z}) = -1$.

Now, we will show that in the case where Axiom 7 (No Certainty) does not hold, C has a decisiveness-maximizing representation.

For any collection of states $\overline{\Theta} \in 2^\Theta$, we say that $\mathcal{C}$ *satisfies the support condition for* $\overline{\Theta}$ if for all $f_a, f_b, f_c \in X_c$ s.t. $f_a = \mathcal{C}(f_a, f_b)$, $f_b = \mathcal{C}(f_b, f_c)$, there exists $\lambda^* \in (0, 1)$ such that for all $\lambda \in (0, \lambda^*)$, act g with $g(\theta) = \lambda f_a(\theta) + (1-\lambda) f_b(\theta)$ for $\theta \in \overline{\Theta}$, $g(\theta) = f_c(\theta)$ for $\theta \notin \theta^*$, and any A that neither improves nor is improved by $g(\theta)$ for $\theta \in \overline{\Theta}$, $g \notin \mathcal{C}(A)$ whenever $f_b \in A$.

Suppose that Axiom 7 (No Certainty) does not hold. This implies that $\mathcal{C}$ does not satisfy the support condition for some $\{\theta^*\}$, $\theta^* \in \Theta$, and so there exists a collection of states

$\mathcal{F} \subseteq 2^\Theta$ containing $\{\theta^*\}$ for which $\mathcal{C}$ does not satisfy the support condition and for which $\mathcal{C}$ satisfies the support condition for $2^\Theta \setminus \mathcal{F}$. Lemma 6 then implies that

$$\mathcal{C}(A) = \bigcup_{\overline{\Theta} \in \mathcal{F}} \{f \in A : \forall\, g \in A,\, g(\theta) \text{ does not improve } f(\theta)\, \forall\, \theta \in \overline{\Theta}\}.$$

For each $\overline{\Theta} \in \mathcal{F}$, let $p_{\overline{\Theta}} \in \Delta(\Theta)$ be a model-induced posterior that has support only on Θ. Letting $P = \{p_{\overline{\Theta}}\}_{\overline{\Theta} \in \mathcal{F}} \in \mathcal{P}$, note that (u, P) represents $\mathcal{C}(A)$, since $p_{\overline{\Theta}} \in \mathcal{I}(P|A)$ iff there exists $f \in A$ for which $u(f(\theta)) \geq u(g(\theta))$ for all $g \in A$, $\theta \in \overline{\Theta}$, in which case $f \in \mathcal{C}(A)$.

The remainder of the proof will deal with the case where Axiom 7 (No Certainty) holds.

Step 3: Construction of $U : X \times \mathcal{A} \to \mathbb{R}$.

Here, our goal is construct a utility map $U : X \times \mathcal{A} \to \mathbb{R}$ from choice data that represents $\mathcal{C}$. We begin by introducing notation that will be useful in this construction.

For any $c \in [-1, 1]$, let $q(c) = \frac{1-c}{2} \circ \underline{z} + \frac{1+c}{2} \circ \overline{z}$; by linearity of u, $u(q(c)) = c$. let f_c denote the constant act with $f_c(\theta) = q(c)$ for all θ. Consider the constant act $f_{-1/2}$. By Axiom 7 (No Certainty), there exists a constant $k > -1/2$ such that for all θ, where given the act g_θ satisfying

$$g_\theta(\theta') = \begin{cases} = q(k) & \theta' = \theta \\ = q(-1) & \text{otherwise} \end{cases}$$

we have $g_\theta \notin \mathcal{C}(A)$ if $f_{-1/2} \in A$ for any A that neither improves nor is improved by f_k. Let $E = \{g_\theta\}_{\theta \in \Theta}$.

Fix k as defined above. We introduce the following notation for any $A \in \mathcal{A}$, $f \in A$:

- Let f^A_{max} denote the *maximal* act corresponding to A, satisfying $f^A_{max}(\theta) = q(\max_{f \in A} u(f(\theta)))$, and define the *minimal* act f^A_{min} analogously.
- Let $\overline{f}$ be the *normalized* act satisfying $\overline{f}(\theta) = q(\frac{1}{4}[u(f(\theta)) - u(f^A_{max}(\theta))])$; since $\frac{1}{4}[u(f(\theta)) - u(f^A_{max}(\theta))] \in [-1, 1]$ for any $f \in A$, $\overline{f}$ is well-defined. Let $\overline{A} = \{\overline{f} : f \in A\}$ collect the normalized acts in A.
- Let $\dot{f}$ be the *k-normalized* act given by $\dot{f} = \lambda f_{-1/2} + (1 - \lambda) f$, for $\lambda = -2k \in [0, 1)$. Let $\dot{A} = \{\dot{f} : f \in A\}$ collect the k-normalized acts in A.
- Let $f^k = \frac{1}{2} f^k + \frac{1}{2} f_{-k}$, with $A^k = \{f^k : f \in A\}$.

We are now ready to construct our utility map.

Claim 1. There exists a map $U : X \times \mathcal{A} \to \mathbb{R}$ satisfying, for all $A \in \mathcal{A}$, $f \in A$,

$$U(f|A) = c$$
$$\text{for } c \text{ satisfying } \{\dot{\overline{f}}^k, \dot{f}^k_c\} = \mathcal{C}(\{\dot{\overline{f}}^k, \dot{f}^k_c\} \cup E^k)$$

that also represents $\mathcal{C}$: that is, for any $f \in \mathcal{C}(A)$ iff $U(f|A) > U(f'|A)$ for any $f' \in A$.

Proof of Claim 1. We first establish several elementary results.

Observation 1. Fixing any $f \in A$, for any $B \in \mathcal{A}$ neither improves nor is improved by f_0, $g_\theta^k \notin \mathcal{C}(B \cup \{\dot{\overline{f}}^k\})$. To see this, fix any $f \in A$, $B \in \mathcal{A}$, where B neither improves nor is improved on by f_0. Note also that for all θ, $u(f_{max}^{\dot{\overline{A}}^k}(\theta)) = u(q(\frac{1}{2}(1+2k)u(f_{max}^{\overline{A}}(\theta)))) = 0$ since $u(f_{max}^{\overline{A}}(\theta)) = 0$, and so $\dot{\overline{f}}^k$ does not improve B. By construction and linearity of u we have $u(f_{min}^{\overline{A}}(\theta)) \geq -1/2$ for all θ, and so $f_{-1/2}$ does not improve $\overline{f}$; by linearity of u this in turn implies that $\dot{f}_{-1/2}^k$ does not improve $\dot{\overline{f}}^k$. Now, since $B \cup \{\dot{\overline{f}}^k, \dot{f}_{-1/2}^k\}$ neither improves nor is improved by f_k, we have $g_\theta \notin \mathcal{C}(B \cup \{\dot{\overline{f}}^k, \dot{f}_{-1/2}^k\})$, which in turn implies that $g_\theta \notin \mathcal{C}(B \cup \{\dot{\overline{f}}^k\})$ by IINIA, as desired. $\circ$

Observation 2. For any $A \in \mathcal{A}$, E^k does not improve $\dot{\overline{A}}^k$, and vice versa. To see this, note that by linearity of u $u(f_{max}^{E^k}(\theta)) = u(\frac{1}{2}q(k) + \frac{1}{2}q(-k)) = 0$ for all θ and also $u(f_{max}^{\dot{\overline{A}}^k}(\theta)) = u(q(\frac{1}{2}(1+2k)u(f_{max}^{\overline{A}}(\theta)))) = 0$ since $u(f_{max}^{\overline{A}}(\theta)) = 0$ for all θ. $\circ$

Observation 3. $f \in \mathcal{C}(A) \iff \dot{\overline{f}}^k \in \mathcal{C}(\dot{\overline{A}}^k)$. To see this, take any f, A, and define f_{-max}^A to be the act satisfying $f_{-max}^A(\theta) = q(-u(f_{max}^A))$ Note that by linearity of u, $u(\overline{f}(\theta)) = u((\frac{1}{4}f + \frac{1}{4}f_{-max}^A + \frac{3}{4}f_0)(\theta))$. We have, by Axiom 4 (Mixture Independence)

$$\begin{aligned} f \in \mathcal{C}(A) &\iff \left(\frac{1}{4}f + \frac{1}{4}f_{-max}^A + \frac{3}{4}f_0\right) \in \mathcal{C}\left(\frac{1}{4}A + \frac{1}{4}f_{-max}^A + \frac{3}{4}f_0\right) \\ &\iff \overline{f} \in \mathcal{C}(\overline{A}) \qquad \text{by Axiom 2 (Monotonicity)} \\ &\iff \dot{\overline{f}}^k \in \mathcal{C}(\dot{\overline{A}}^k) \end{aligned}$$

where the last step follows from Axiom 4 (Mixture Independence) and the fact that $\dot{\overline{f}}^k = \frac{1}{2}(1-\lambda)\overline{f} + \frac{1}{2}\lambda f_{-1/2} + \frac{1}{2}f_{-k}$.

Now, we show that U is well defined.

Observation 4. For each f, A, there exists a unique $c \in [-1,1]$ such that $\{\dot{\overline{f}}^k, \dot{f}_c^k\} = \mathcal{C}(\{\dot{\overline{f}}^k, \dot{f}_c^k\} \cup E^k)$, and furthermore $c \leq 0$.

To see this, Fix some $A \in \mathcal{A}$, $f \in A$. By Axiom 3 (Mixture Continuity), the sets

$$\begin{aligned} \overline{W} &= \{\alpha \in [0,1] : \alpha \dot{f}_{-1}^k + (1-\alpha)\dot{f}_1^k \in \mathcal{C}(\{\alpha \dot{f}_{-1}^k + (1-\alpha)\dot{f}_1^k\} \cup \{\dot{\overline{f}}^k\} \cup E^k)\} \\ \underline{W} &= \{\alpha \in [0,1] : \dot{\overline{f}}^k \in \mathcal{C}(\{\alpha \dot{f}_{-1}^k + (1-\alpha)\dot{f}_1^k\} \cup \{\dot{\overline{f}}^k\} \cup E^k)\} \end{aligned}$$

are closed. Since $u(\dot{f}_1^k) = \frac{1}{2} + k > 0$ and the payoffs of $\dot{\overline{f}}^k$ and acts in E^k are bounded from above by 0, Axiom 2 (Monotoncity) guarantees that $\dot{f}_1^k \in \mathcal{C}(\{\dot{f}_1^k\} \cup \{\dot{\overline{f}}^k\} \cup E^k)$ and so $\overline{W}$ is non-empty. Also, noting that $E^k \cap \mathcal{C}(\{\dot{f}_{-1}^k\} \cup \{\dot{\overline{f}}^k\} \cup E^k) = \varnothing$ by Observation

1, Axiom 2 (Monotonicity) guarantees that $\dot{\overline{f}}^k \in \mathcal{C}(\{\dot{f}^k_{-1}\} \cup \{\dot{\overline{f}}^k\} \cup E^k)$, and so $\underline{W}$ is nonempty. Note also $E^k \cap \mathcal{C}(\{\alpha \dot{f}^k_{-1} + (1-\alpha)\dot{f}^k_1\} \cup \{\dot{\overline{f}}^k\} \cup E^k) = \varnothing$ for all $\alpha \in [0,1]$; to see this, note that for $\alpha \in [0, 1/2)$, $\alpha \dot{f}^k_{-1} + (1-\alpha)\dot{f}^k_1$ does not improve E^k, and by Observation 1 $E^k \cap \mathcal{C}(\{\alpha \dot{f}^k_{-1} + (1-\alpha)\dot{f}^k_1\} \cup \{\dot{\overline{f}}^k\} \cup E^k) = \varnothing$, whereas for $\alpha \in (1/2, 1]$ we have $E^k \cap \mathcal{C}(\{\alpha \dot{f}^k_{-1} + (1-\alpha)\dot{f}^k_1\} \cup \{\dot{\overline{f}}^k\} \cup E^k) = \varnothing$ by Axiom 2 (Monotonicity). This implies that $\mathcal{C}(\{\dot{\overline{f}}^k, \dot{f}^k_c\} \cup E^k) = \{\dot{\overline{f}}^k, \dot{f}^k_c\}$ iff $c \in \overline{W} \cap \underline{W}$.

Axiom 2 (Monotonicity) guarantees that $\overline{W}$ and $\underline{W}$ are convex; this implies that $\overline{W}$ and $\underline{W}$ nonempty intervals, the union of which is equal to $[0,1]$. This in turn implies that $\inf \overline{W} = \sup \underline{W}$, and that $\alpha^k \equiv \inf \overline{W} = \sup \underline{W}$ is the unique member of $\overline{W} \cap \underline{W}$, and so there exists a unique c satisfying $\mathcal{C}(\{\dot{\overline{f}}^k, \dot{f}^k_c\} \cup E^k) = \{\dot{\overline{f}}^k, \dot{f}^k_c\}$.

Finally, to see that $c \leq 0$, suppose not: $c > 0$. Then $\dot{f}^k_c$ strictly dominates $\dot{\overline{f}}^k$, and so $\mathcal{C}(\{\dot{\overline{f}}^k, \dot{f}^k_c\} \cup E^k) = \{\dot{\overline{f}}^k, \dot{f}^k_c\}$ would violate Axiom 2 (Monotonicity). $\circ$

Now we show that this $U(f|A)$ represents $\mathcal{C}$, in particular that $f \in \mathcal{C}(A) \iff U(f|A) \geq U(f'|A)$ for all $f \in A$.

($\Longrightarrow$). Suppose $f \in \mathcal{C}(A)$. Towards a contradiction, suppose that, $U(f|A) < U(f'|A)$ for some $f' \in A$. This implies that there exists $c < c'$ with $U(f|A) = c, U(f'|A) = c'$ such that $\dot{\overline{f'}}^k, \dot{f}^k_{c'} \in \mathcal{C}(\{\dot{\overline{f'}}^k, \dot{f}^k_{c'}\} \cup E^k)$ and $\dot{\overline{f}}^k, \dot{f}^k_c \in \mathcal{C}(\{\dot{\overline{f}}^k, \dot{f}^k_c\} \cup E^k)$. By Observation 4, we know that $\dot{f}^k_c, \dot{f}^k_{c'}$ do not improve E^k or $\dot{\overline{A}}^k$. By Axiom 2 (Monotonicity), we have

$$
\begin{aligned}
& f^k_c \notin \mathcal{C}(\{\dot{\overline{f}}^k, \dot{f}^k_c, \dot{f}^k_{c'}\} \cup E^k) \\
\Longrightarrow\; & \dot{\overline{f}}^k \notin \mathcal{C}(\{\dot{\overline{f}}^k, \dot{f}^k_c, \dot{f}^k_{c'}\} \cup E^k) \qquad \text{by Axiom 1 (IINIA)} \\
\Longrightarrow\; & \dot{\overline{f}}^k \notin \mathcal{C}(\{\dot{\overline{f}}^k, \dot{f}^k_{c'}\} \cup E^k) \qquad \text{by Lemma 3, since } \dot{f}^k_c \notin \mathcal{C}(\{\dot{\overline{f}}^k, \dot{f}^k_c, \dot{f}^k_{c'}\} \cup E^k) \\
\Longrightarrow\; & \dot{\overline{f}}^k \notin \mathcal{C}(\{\dot{f}^k_{c'}\} \cup \dot{\overline{A}}^k \cup E^k) \qquad \text{By Axiom 1 (IINIA), since } \dot{\overline{A}}^k \text{ does not improve } E^k \\
\Longrightarrow\; & \dot{\overline{f}}^k \notin \mathcal{C}(\{\dot{f}^k_{c'}\} \cup \dot{\overline{A}}^k) \qquad \text{By Lemma 3, since } E^k \cap \mathcal{C}(\{\dot{\overline{f}}^k, \dot{f}^k_c, \dot{f}^k_{c'}\} \cup E^k) = \varnothing
\end{aligned}
$$

By Observation 3, $f \in \mathcal{C}(A) \implies \dot{\overline{f}}^k \in \mathcal{C}(\dot{\overline{A}}^k)$. Since $\dot{\overline{f}}^k \notin \mathcal{C}(\{\dot{f}_{c'}\} \cup \dot{\overline{A}}^k)$, and $\dot{f}^k_{c'}$ does not improve $\dot{\overline{A}}^k$, Lemma 3 implies that $\dot{f}^k_{c'} \in \mathcal{C}(\{\dot{f}^k_{c'}\} \cup \dot{\overline{A}}^k)$. Since $E^k \cap \mathcal{C}(\{\dot{f}^k_{c'}\} \cup \dot{\overline{A}}^k \cup E^k) = \varnothing$ by Observation 1, by Lemma 3 we have $\dot{f}^k_{c'} \in \mathcal{C}(\{\dot{f}^k_{c'}\} \cup \dot{\overline{A}}^k \cup E^k)$; applying Axiom 1 (IINIA), this in turn implies that $\dot{\overline{f'}}^k \in \mathcal{C}(\{\dot{f}^k_{c'}\} \cup \dot{\overline{A}}^k \cup E^k)$. Again applying Axiom 1 (IINIA) we have $\dot{\overline{f'}}^k \in \mathcal{C}(\{\dot{f}^k_{c'}\} \cup \dot{\overline{A}}^k)$ and also $\dot{\overline{f'}}^k \in \mathcal{C}(\dot{\overline{A}}^k)$.

Now, since $\dot{\overline{f}}^k, \dot{\overline{f'}}^k \in \mathcal{C}(\dot{\overline{A}}^k)$, Axiom 1 (IINIA) implies that $\dot{\overline{f}}^k \in \mathcal{C}(\{\dot{f}^k_{c'}\} \cup \dot{\overline{A}}^k)$, a contradiction.

($\Longleftarrow$). We proceed by showing the contrapositive: if $f \notin \mathcal{C}(A)$ and $f' \in \mathcal{C}(A)$, then $U(f|A) < U(f'|A)$. Suppose that $f \notin \mathcal{C}(A)$ and $f' \in \mathcal{C}(A)$. Towards a contradiction, suppose that $U(f|A) \geq U(f'|A)$. This implies that there exists $c \geq c'$ with $U(f|A) = c, U(f'|A) = c'$

such that $\dot{\overline{f'}}^k, \dot{f}^k_{c'} \in \mathcal{C}(\{\dot{\overline{f'}}^k, \dot{f}^k_{c'}\} \cup E^k)$ and $\dot{\overline{f}}^k, \dot{f}^k_c \in \mathcal{C}(\{\dot{\overline{f}}^k, \dot{f}^k_c\} \cup E^k)$.

First, note that $\dot{f}^k_c \in \mathcal{C}(\{\dot{\overline{f'}}^k, \dot{f}^k_{c'}, \dot{\overline{f}}^k, \dot{f}^k_c\} \cup E^k)$. To see this, suppose not; if $\dot{f}^{kk}_c \notin \mathcal{C}(\{\dot{\overline{f'}}^k, \dot{f}^k_{c'}, \dot{\overline{f}}^k, \dot{f}^k_c\} \cup E^k)$, then by Axiom 1 (IINIA) we $\dot{\overline{f}}^k \notin \mathcal{C}(\{\dot{\overline{f'}}^k, \dot{f}^k_{c'}, \dot{\overline{f}}^k, \dot{f}^k_c\} \cup E^k)$. This implies that at least one of $\dot{\overline{f'}}^k, \dot{f}^k_{c'}$ must belong to $\mathcal{C}(\{\dot{\overline{f'}}^k, \dot{f}^k_{c'}, \dot{\overline{f}}^k, \dot{f}^k_c\} \cup E^k)$, since $E^k \cap \mathcal{C}(\{\dot{\overline{f'}}^k, \dot{f}^k_{c'}, \dot{\overline{f}}^k, \dot{f}^k_c\} \cup E^k) = \varnothing$, and so by Axiom 1 (IINIA) we have $\dot{\overline{f'}}^k, \dot{f}^k_{c'} \in \mathcal{C}(\{\dot{\overline{f'}}^k, \dot{f}^k_{c'}, \dot{\overline{f}}^k, \dot{f}^k_c\} \cup E^k)$. In turn, by Axiom 2 (Monotonicity), $\dot{f}^k_c \in \mathcal{C}(\{\dot{\overline{f'}}^k, \dot{f}^k_{c'}, \dot{\overline{f}}^k, \dot{f}^k_c\} \cup E^k)$, a contradiction; it must be the case that $\dot{f}^k_c \in \mathcal{C}(\{\dot{\overline{f'}}^k, \dot{f}^k_{c'}, \dot{\overline{f}}^k, \dot{f}^k_c\} \cup E^k)$, and therefore $\dot{\overline{f}}^k \in \mathcal{C}(\{\dot{\overline{f'}}^k, \dot{f}_{c'}, \dot{\overline{f}}^k, \dot{f}^k_c\} \cup E^k)$ by Axiom 1 (IINIA). Again applying Axiom 1 (IINIA), we have $\dot{\overline{f}}^k \in \mathcal{C}(\{\dot{\overline{f'}}^k, \dot{\overline{f}}^k\} \cup E^k)$.

Now, $f' \in \mathcal{C}(A)$ implies that $\dot{\overline{f'}}^k \in \mathcal{C}(\dot{\overline{A}}^k)$ by Observation 3. By Lemma 3, since $E^k \cap \mathcal{C}(\dot{\overline{A}}^k \cup E^k) = \varnothing$, $\dot{\overline{f'}}^k \in \mathcal{C}(\dot{\overline{A}}^k \cup E^k)$, and by Axiom 1 (IINIA), $\dot{\overline{f'}}^k \in \mathcal{C}(\{\dot{\overline{f'}}^k, \dot{\overline{f}}^k\} \cup E^k)$. Now, since $\dot{\overline{f'}}^k \in \mathcal{C}(\dot{\overline{A}}^k \cup E^k)$ and $\dot{\overline{f'}}^k, \dot{\overline{f}}^k \in \mathcal{C}(\{\dot{\overline{f'}}^k, \dot{\overline{f}}^k\} \cup E^k)$, by Axiom 1 (IINIA) $\dot{\overline{f}}^k \in \mathcal{C}(\dot{\overline{A}}^k \cup E^k)$ which in turn implies $\dot{\overline{f}}^k \in \mathcal{C}(\dot{\overline{A}}^k)$. By Observation 3, this implies that $f \in \mathcal{C}(A)$, a contradiction.

△

Step 4: Construction of $\varphi : \mathbb{R}^\Theta_- \to \mathbb{R}$. Now, we show that there exists an increasing, superlinear, and C-additive map $\varphi : \mathbb{R}^\Theta_- \to \mathbb{R}$ such that for all $A \in \mathcal{A}$, $f \in A$, and f^A_{max} the maximal act constructed from A,

$$\varphi(u \circ f - u \circ f^A_{max}) = U(f|A)$$

which also satisfies $\varphi(\xi_{-1}) = -1$.

Note that $\{(u \circ f - u \circ f^A_{max}) : A \in \mathcal{A}, f \in A\} = [-2, 0]^\Theta$. Note also that for any f, A, f', A' s.t. $(u \circ f - u \circ f^A_{max}) = (u \circ f' - u \circ f^{A'}_{max})$, $\overline{f} = \overline{f'}$ and $\overline{A} = \overline{A'}$, which in turn implies $U(f|A) = U(f'|A')$ by construction of U. Therefore, we can define the map $\psi : [-2, 0]^\Theta \to \mathbb{R}$ by taking

$$\psi(u \circ f - u \circ f^A_{max}) = U(f|A) \qquad \text{for all } A \in \mathcal{A}, f \in A$$

We first establish that ψ is positively homogenous on its domain.

Claim 2: For all $\xi \in [-2, 0]^\Theta, \alpha > 0$ s.t. $\alpha\xi \in [-2, 0]^\Theta$.

Proof of Claim 2. Take $\xi \in [-2, 0]^\Theta, \alpha \in (0, 1)$. There exists $A \in \mathcal{A}, f \in A$ such that $u \circ f - u \circ f^A_{max} = \xi$. Take $f' = \alpha f + (1 - \alpha) f_0$, $A' = \alpha A + (1 - \alpha) f_0$; by linearity of u, we have $u \circ f' - u \circ f^{A'}_{max} = \alpha\xi$. By definition of ψ, we have $\psi(\xi) = U(f|A), \psi(\alpha\xi) = U(f'|A')$, and by definition of U, we have

$$\{\dot{\overline{f}}^k, \dot{f}^k_c\} = \mathcal{C}(\{\dot{\overline{f}}^k, \dot{f}^k_c\} \cup E^k)$$
$$\{\dot{\overline{f'}}^k, \dot{f}^k_{c'}\} = \mathcal{C}(\{\dot{\overline{f}}^k, \dot{f}^k_{c'}\} \cup E^k) \qquad \text{for } c = U(f|A), c' = U(f'|A')$$

By linearity of u, we have that for all θ, $\overline{f}^k(\theta) = q(\frac{1}{2}(1+2k)u(f(\theta)))$. Again applying linearity of u, this implies that $\dot{\overline{f'}}^k(\theta) = \alpha\overline{f}^k(\theta) + (1-\alpha)f_0$. Note also that $\dot{f}_c^k(\theta) = q(\frac{1}{2}(1+2k)c)$ for all θ by linearity of u, and so again applying linearity of u, $\alpha\dot{f}_c^k + (1-\alpha)f_0 = \dot{f}_{\alpha c}^k$. We have

$$
\begin{aligned}
\{\dot{\overline{f}}^k, \dot{f}_c^k\} = \mathcal{C}(\{\dot{\overline{f}}^k, \dot{f}_c^k\} \cup E^k) &\implies \{\dot{\overline{f'}}^k, \dot{f}_{\alpha c}^k\} = \mathcal{C}(\{\dot{\overline{f'}}^k, \dot{f}_{\alpha c}^k\} \cup (\alpha E^k + (1-\alpha)f_0)) \\
&\implies \{\dot{\overline{f'}}^k, \dot{f}_{\alpha c}^k\} = \mathcal{C}(\{\dot{\overline{f'}}^k, \dot{f}_{\alpha c}^k\} \cup (\alpha E^k + (1-\alpha)f_0) \cup E^k) \\
&\implies \{\dot{\overline{f'}}^k, \dot{f}_{\alpha c}^k\} = \mathcal{C}(\{\dot{\overline{f'}}^k, \dot{f}_{\alpha c}^k\} \cup E^k)
\end{aligned}
$$

The first step follows from Axiom 4 (Mixture Independence); the second step follows from Lemma 3, the fact that E^k does not improve $\alpha E^k + (1-\alpha)f_0$, and the fact that by Observation 1, $E^k \cap \mathcal{C}(\{\dot{\overline{f'}}^k, \dot{f}_{\alpha c}^k\} \cup (\alpha E^k + (1-\alpha)f_0) \cup E^k) = \varnothing$; the third step follows from Axiom 1 (IINIA). So we have $U(f'|A') = c' = \alpha c = \alpha U(f|A)$, and so $\psi(\alpha\xi) = U(f'|A') = \alpha U(f|A) = \psi(\xi)$.

Now, take any $\xi \in [-2, 0]^\Theta$, $\alpha > 0$ s.t. $\alpha\xi \in [-2, 0]^\Theta$. Note that

$$
\begin{aligned}
\psi(\xi) &= \psi(\frac{1}{\alpha}\alpha\xi) \\
\psi(\xi) &= \frac{1}{\alpha}\psi(\alpha\xi) \qquad \text{by the previously established result} \\
\alpha\psi(\xi) &= \psi(\alpha\xi)
\end{aligned}
$$

△

Now, define $\varphi : \mathbb{R}_-^\Theta \to \mathbb{R}$ as follows:

$$\varphi(\xi) = \frac{1}{\alpha}\psi(\alpha\xi)$$

for any $\alpha > 0$ s.t. $\alpha\xi \in [-2, 0]^{-\Theta}$. To see that φ is well defined, take any $\beta > \alpha > 0$ s.t. $\alpha\xi, \beta\xi \in [-2, 0]^\Theta$. We have, by Claim 2 (positive homogeneity of ψ), $\frac{1}{\alpha}\psi(\alpha\epsilon) = \frac{1}{\alpha}\psi(\frac{\alpha}{\beta}\beta\psi) = \frac{1}{\beta}\psi(\xi)$.

Claim 3: φ is positively homogenous. Also, for any $a < 0$, $\varphi(\xi_a) = \frac{1}{4}a$.

Proof of Claim 3. To see that φ is positively homogenous, take $\xi \in \mathbb{R}_-^\Theta, \alpha > 0$, by picking some $\beta > 0$ s.t. $\beta\alpha\xi, \beta\xi \in [-2, 0]^\Theta$. We have, by definition of φ and by Claim 2 (positive homogeneity of ψ), $\varphi(\alpha\xi) = \frac{1}{\beta}\psi(\beta\alpha\xi) = \frac{1}{\beta}\alpha(\beta\xi) = \alpha\varphi(\xi)$.

To show the second statement, start by taking any $a \in [-2, 0]$. There exists some $A \in \mathcal{A}$, $f \in A$ such that $u \circ f - u \circ f_{max}^A = \xi_c$. Since $\overline{f} = q \circ (\frac{1}{4}(u \circ f - u \circ f_{max}^A))$, linearity of u implies that $\overline{f} = f_{1/4a}$. Therefore, we have $\dot{\overline{f}}^k = \dot{f}_{1/4a}^k$, and in particular

$$\{\dot{\overline{f}}^k, \dot{f}_{1/4a}^k\} = \mathcal{C}(\{\dot{\overline{f}}^k, \dot{f}_{1/4a}^k\} \cup E^k)$$

which implies that $\varphi(\xi_a) = U(f|A) = \frac{1}{4}a$.

To see that this holds for any $a < 0$, fix some $a < 0$, and fix some $\beta > 0$ s.t. $\beta\xi_a \in [-2,0]^\Theta$. We have, by positive homogeneity of φ, and the previously established result, $\varphi(\xi_a) = \varphi(\frac{1}{\beta}\beta\xi_a) = \frac{1}{\beta}\varphi(\beta\xi_a) = \frac{1}{\beta}\beta\frac{1}{4}a = \frac{1}{4}a$.

△

Claim 4: φ is C-additive.

Proof of Claim 4. To see that φ is C-additive, we first show that C-additivity holds for any $\xi \in [-2,0]^\Theta$, $a \in [-1,0]$ s.t. $\xi + \xi_a \in [-2,0]^\Theta$: that is, for such ξ, a, we have $\varphi(\xi + \xi_a) = \varphi(\xi) + \varphi(\xi_a)$.

Fix such ξ, a. There exists $f, g \in X_c$, $A = \{f, g\}$ s.t. $u \circ f - u \circ f^A_{max} = \xi$. Now consider $f' = \frac{1}{2}f + \frac{1}{2}f_a$ $A' = \{f', \frac{1}{2}g + \frac{1}{2}f_0\}$; since $\xi + \xi_a \leq 0$, we have $u \circ f + a - u \circ f^A_{max} \leq 0$ which in turn implies, by linearity of u, that $u \circ f' \leq u \circ (\frac{1}{2}g + \frac{1}{2}f_0)$. Therefore, $u \circ f' - u \circ f^{A'}_{max} = \frac{1}{2}\xi + \frac{1}{2}a$. We therefore have

$$\{\dot{\overline{f}}^k, \dot{f}^k_c\} = \mathcal{C}(\{\dot{\overline{f}}^k, \dot{f}^k_c\} \cup E^k)$$
$$\{\dot{\overline{f'}}^k, \dot{f}^k_{c'}\} = \mathcal{C}(\{\dot{\overline{f'}}^k, \dot{f}^k_{c'}\} \cup E^k)$$

for $c = U(f|A) = \varphi(\xi), c' = U(f'|A') = \varphi(\frac{1}{2}\xi + \frac{1}{2}\xi_a)$.

Let $\tilde{a} = \frac{1}{4}a$. Since $u \circ f' - u \circ f^{A'}_{max} = \frac{1}{2}\xi + \frac{1}{2}a$, and $\overline{f'} = q \circ (\frac{1}{4}(u \circ f' - u \circ f^{A'}_{max})) = q \circ (\frac{1}{4}(\frac{1}{2}\xi + \frac{1}{2}a))$ and $\overline{f} = q \circ (\frac{1}{4}\xi)$, by linearity of u we have $\overline{f'} = \frac{1}{2}\overline{f} + \frac{1}{2}f_{\tilde{a}}$, and since $\dot{\overline{f}}^k(\theta) = q(\frac{1}{2}(1+2k)\overline{f}(\theta))$ for all f, again applying linearity of u we have $\dot{\overline{f'}}^k(\theta) = \frac{1}{2}\dot{\overline{f}}^k(\theta) + \frac{1}{2}\dot{f}^k_{\tilde{a}}$. Linearity of u also implies, for any constant act f_c, that $\frac{1}{2}\dot{f}^k_c + \frac{1}{2}\dot{f}^k_{\tilde{a}} = \dot{f}^k_{1/2c+1/2\tilde{a}}$. We have, by Axiom 4 (Mixture Independence)

$$\{\dot{\overline{f}}^k, \dot{f}^k_c\} = \mathcal{C}(\{\dot{\overline{f}}^k, \dot{f}^k_c\} \cup E^k)$$
$$\implies \{\dot{\overline{f'}}^k, \dot{f}^k_{1/2c+1/2\tilde{a}}\} = \mathcal{C}(\{\dot{\overline{f'}}^k, \dot{f}^k_{1/2c+1/2\tilde{a}}\} \cup \frac{1}{2}E^k + \frac{1}{2}\dot{f}^k_{\tilde{a}})$$

Now, note that $c \leq \max_\theta u(\overline{f}(\theta))$; if not, then f_c would dominate $\overline{f}$ and by linearity of u, $\dot{f}^k_c$ would dominate $\dot{\overline{f}}^k$, which contradicts $\{\dot{\overline{f}}^k, \dot{f}^k_c\} = \mathcal{C}(\{\dot{\overline{f}}^k, \dot{f}^k_c\} \cup E^k)$ by Axiom 2 (Monotonicity). This in turn implies that $c \leq \max_\theta \frac{1}{4}\xi(\theta)$, and since $\xi + \xi_a \leq 0$, we have $c - \tilde{a} \leq 0$. Therefore $\dot{f}^k_{1/2c+1/2\tilde{a}}$ does not improve E^k, and so by Observation 1, we have $E^k \cap \mathcal{C}(\{\dot{\overline{f'}}^k, \dot{f}^k_{1/2c+1/2\tilde{a}}\} \cup E^k) = \varnothing$. Also, by Observation 1 and Axiom 4 (Mixture Independence), we have $(\frac{1}{2}E^k + \frac{1}{2}\dot{f}^k_{\tilde{a}}) \cap \mathcal{C}(\{\dot{\overline{f'}}^k, \dot{f}^k_{1/2c+1/2\tilde{a}}\} \cup \frac{1}{2}E^k + \frac{1}{2}\dot{f}^k_{\tilde{a}}) = \varnothing$. Therefore, by Axiom 1 (IISIA), we have

$$\{\dot{\overline{f'}}^k, \dot{f}^k_{1/2c+1/2\tilde{a}}\} = \mathcal{C}(\{\dot{\overline{f'}}^k, \dot{f}^k_{1/2c+1/2\tilde{a}}\} \cup \frac{1}{2}E^k + \frac{1}{2}\dot{f}^k_{\tilde{a}})$$
$$\implies \{\dot{\overline{f'}}^k, \dot{f}^k_{1/2c+1/2\tilde{a}}\} = \mathcal{C}(\{\dot{\overline{f'}}^k, \dot{f}^k_{1/2c+1/2\tilde{a}}\} \cup E^k)$$

Therefore, $U(f'|A') = c' = \frac{1}{2}c + \frac{1}{2}\tilde{a} = \frac{1}{2}U(f|A) + \frac{1}{2}\tilde{a}$. This in turn implies, by Claim 3 (homogeneity of φ), that $\varphi(\xi + \xi_a) = \varphi(\xi) + \tilde{a}$. Note also that by Claim 3, we have for any

$a < 0$, $\varphi(\xi_a) = \frac{1}{4}a$. This implies that $\varphi(\xi + \xi_a) = \varphi(\xi) + \varphi(\xi_a)$ as desired.

Now, take any $\xi \in \mathbb{R}^\Theta_-$, $a \leq 0$ s.t. $\xi + \xi_a \leq 0$. There exists some $\beta > 0$ s.t. $\beta(\xi + \xi_a) \in [-2, 0]^\Theta$ and $\beta a \in [-1, 0]$. We have

$$\begin{aligned}
\varphi(\xi + \xi_a) &= \frac{1}{\beta}\varphi(\beta(\xi + \xi_a)) \\
&= \frac{1}{\beta}\varphi(\beta\xi) + \frac{1}{\beta}\varphi(\beta\xi_a) \\
&= \varphi(\xi) + (\xi_a)
\end{aligned}$$

where the first and third equalities follow from Claim 3 (homogeneity of φ), and the second equality follows from the preceding result.

△

Claim 5: φ is sublinear.

Proof of Claim 5. By Claim 3, we know that φ is positively homeogenous. All that remains is to show that φ is convex: that is for any $\xi, \xi' \in \mathbb{R}^\Theta_-, \alpha \in (0, 1)$ $\varphi(\alpha\xi + (1 - \alpha)\xi') \leq \alpha\varphi(\xi) + (1 - \alpha)\varphi(\xi')$.

Start by showing that the result holds for any $\xi, \xi' \in [-2, 0]^\Theta$ such that $\varphi(\xi) = \varphi(\xi')$. Fix such ξ, ξ' and $\lambda \in (0, 1)$. There exists acts f, g, f', g', satisfying $u \circ f \leq u \circ g$, $u \circ f' \leq u \circ g'$, and with $\xi = u \circ f - u \circ f^A_{max}$, $\xi' = u \circ f' - u \circ f^{A'}_{max}$ for $A = \{f, g\}$, $A' = \{f', g'\}$. Let $f'' = \lambda f + (1-\lambda) f'$, $g'' = \lambda g + (1-\lambda) g'$, $A'' = \{f'', g''\}$. By linearity of u, we have $u \circ f'' + u \circ f^{A''}_{max} = \lambda\xi + (1-\lambda)\xi$. We therefore have $\varphi(\xi) = U(f|A)$, $\varphi(\xi') = U(f'|A')$, and $\varphi(\lambda\xi + (1-\lambda)\xi') = U(f''|A'')$. Since $u \circ f'' - u \circ f^{A''}_{max} = \lambda(u \circ f'' - u \circ f^{A''}_{max}) + (1-\lambda)(u \circ f'' - u \circ f^{A''}_{max})$, we have $\overline{f''} = \lambda \overline{f} + (1-\lambda)\overline{f'}$, which in turn implies, by linearity of u, that $\dot{\overline{f''}}^k = \lambda \dot{\overline{f}}^k + (1 - \lambda)\dot{\overline{f'}}^k$.

Now towards a contradiction, suppose that $\varphi(\lambda\xi + (1-\lambda)\xi') > \varphi(\lambda\xi) + \varphi((1-\lambda)\xi') = \varphi(\xi)$. This implies that for $c = U(f|A) = U(f'|A') < c'' = U(f''|A'')$, we have

$$\begin{aligned}
\{\dot{\overline{f}}^k, \dot{f}^k_c\} &= \mathcal{C}(\{\dot{\overline{f}}^k, \dot{f}^k_c\} \cup E^k) \\
\{\dot{\overline{f'}}^k, \dot{f}^k_c\} &= \mathcal{C}(\{\dot{\overline{f'}}^k, \dot{f}^k_c\} \cup E^k) \\
\{\dot{\overline{f''}}^k, \dot{f}^k_{c''}\} &= \mathcal{C}(\{\dot{\overline{f''}}^k, \dot{f}^k_{c''}\} \cup E^k)
\end{aligned}$$

First, note that $\{\dot{\overline{f}}^k, \dot{\overline{f'}}^k\} \in \mathcal{C}(\{\dot{\overline{f}}^k, \dot{\overline{f'}}^k, \dot{f}^k_c\} \cup E^k)$; if not, then $\dot{f}^k_c \in \mathcal{C}(\{\dot{\overline{f}}^k, \dot{\overline{f'}}^k, \dot{f}^k_c\} \cup E^k)$ and so nothing that $\dot{\overline{f'}}^k$ does not improve E^k, Axiom 1 (IINIA) would imply that $\dot{\overline{f}}^k \in \mathcal{C}(\{\dot{\overline{f}}^k, \dot{\overline{f'}}^k, \dot{f}^k_c\} \cup E^k)$, a contradiction. Axiom 5 (Mixture Aversion) then implies that

$$\begin{aligned}
&\dot{\overline{f}}^k \in \mathcal{C}(\{\dot{\overline{f}}^k, \dot{\overline{f'}}^k, \dot{\overline{f''}}^k, \dot{f}^k_c\} \cup E^k) \\
\implies &\dot{f}^k_c \in \mathcal{C}(\{\dot{\overline{f}}^k, \dot{\overline{f'}}^k, \dot{\overline{f''}}^k, \dot{f}^k_c\} \cup E^k) \\
\implies &\dot{f}^k_c \in \mathcal{C}(\{\dot{\overline{f''}}^k, \dot{f}^k_c\} \cup E^k)
\end{aligned}$$

where, noting that noting that $\{\dot{\overline{f}}^k, \dot{\overline{f'}}^k, \dot{\overline{f''}}^k\}$ do not improve E^k, the second line follows from Axiom 1 (IINIA) and the fact that $\{\dot{\overline{f}}^k, \dot{f}_c^k\} \in \mathcal{C}(\{\dot{\overline{f}}^k, \dot{f}_c^k\} \cup E^k)$, and the third line follows from Axiom 1 (IINIA).

Now, note that $\dot{\overline{f''}}^k \in \mathcal{C}(\{\dot{f}_c^k, \dot{f}_{c''}^k, \dot{\overline{f''}}^k\} \cup E^k)$; if not, then $\dot{f}_{c''}^k \notin \mathcal{C}(\{\dot{f}_c^k, \dot{f}_{c''}^k, \dot{\overline{f''}}^k\} \cup E^k)$ by Axiom 1 (IINIA), which implies that $\dot{f}_c^k \in \mathcal{C}(\{\dot{f}_c^k, \dot{f}_{c''}^k, \dot{\overline{f''}}^k\} \cup E^k)$ since $E^k \cap \mathcal{C}(\{\dot{f}_c^k, \dot{f}_{c''}^k, \dot{\overline{f''}}^k\} \cup E^k) = \varnothing$; this implies a contradiction by Axiom 2 (Monotonicity). Applying Axiom 1 (IINIA), $\dot{\overline{f''}}^k \in \mathcal{C}(\{\dot{f}_c^k, \dot{f}_{c''}^k, \dot{\overline{f''}}^k\} \cup E^k) \implies \dot{\overline{f''}}^k \in \mathcal{C}(\{\dot{f}_c^k, \dot{\overline{f''}}^k\} \cup E^k)$. So we have

$$\{\dot{\overline{f''}}^k, \dot{f}_c^k\} = \mathcal{C}(\{\dot{\overline{f''}}^k, \dot{f}_c^k\} \cup E^k)$$

a contradiction since $c'' > c$ and $\{\dot{\overline{f''}}^k, \dot{f}_{c''}^k\} = \mathcal{C}(\{\dot{\overline{f''}}^k, \dot{f}_{c''}^k\} \cup E^k)$; we must therefore have $\varphi(\lambda\xi + (1-\lambda)\xi') \leq \varphi(\lambda\xi) + \varphi((1-\lambda)\xi')$.

Now extend this result to any $\xi, \xi' \in \mathbb{R}_-^\Theta$ such that $\varphi(\xi) = \varphi(\xi')$. Take any such ξ, ξ', and $\lambda \in (0,1)$. There exists $\alpha > 0$ s.t. $\alpha\xi, \alpha\xi' \in [-2,0]^\Theta$; We have

$$\begin{aligned}
\varphi(\lambda\xi + (1-\lambda)\xi) &= \frac{1}{\alpha}\varphi(\alpha(\lambda\xi + (1-\lambda)\xi')) \\
&\leq \frac{1}{\alpha}\lambda\varphi(\alpha\xi) + \frac{1}{\alpha}(1-\lambda)\varphi(\alpha\xi') \\
&= \lambda\varphi(\xi) + (1-\lambda)\varphi(\xi')
\end{aligned}$$

where the first and third equalities follow from Claim 3 (homogeneity of φ) and the inequality follows from the preceding result.

Finally, consider $\xi, \xi' \in \mathbb{R}_-^\Theta$ with $\varphi(\xi) > \varphi(\xi')$, and $\lambda \in (0,1)$. Take $a = \varphi(\xi') - \varphi(\xi) < 0$, and define $\tilde{\xi} = \xi + \xi_{4a}$; by Claim 4, we have $\varphi(\tilde{\xi}) = \varphi(\xi) + a = \varphi(\xi')$. Now by the preceding result, we have

$$\begin{aligned}
\varphi(\lambda\xi + (1-\lambda)\xi' + \lambda\xi_{4a}) &= \varphi(\lambda\tilde{\xi} + (1-\lambda)\xi') \\
&\leq \lambda\varphi(\tilde{\xi}) + (1-\lambda)\varphi(\xi') \\
&= \lambda(\varphi(\xi) + a) + (1-\lambda)\varphi(\xi')
\end{aligned}$$

and so by C-additivity of φ, we have $\varphi(\lambda\xi + (1-\lambda)\xi') \leq \lambda\varphi(\xi) + (1-\lambda)\varphi(\xi')$.

△

Claim 6: φ is increasing.

Proof of Claim 6. First, show that φ is increasing on $[-2,0]^\Theta$. Take $\xi, \xi' \in [-2,0]^\Theta$ with $\xi \geq \xi'$. There exists $A, A' \in \mathcal{A}$, $f \in A$, $f \in A'$, such that $\xi = u \circ f - u \circ f_{max}^A$, $\xi' = u \circ f' - u \circ f_{max}^{A'}$ we have $\varphi(\xi) = U(f|A) = c$, $\varphi(\xi') = U(f'|A') = c'$, where

$$\{\dot{\overline{f}}^k, \dot{f}_c^k\} = \mathcal{C}(\{\dot{\overline{f}}^k, \dot{f}_c^k\} \cup E^k)$$
$$\{\dot{\overline{f'}}^k, \dot{f}_c^k\} = \mathcal{C}(\{\dot{\overline{f'}}^k, \dot{f}_c^k\} \cup E^k)$$

Since $\overline{f} = q \circ (\frac{1}{4}(u \circ f - u \circ f^A_{max}))$, $\overline{f'} = q \circ (\frac{1}{4}(u \circ f' - u \circ f^{A'}_{max}))$ and $\xi \geq \xi'$, we have $u \circ \overline{f} \geq u \circ \overline{f'}$, and therefore $u \circ \dot{\overline{f}}^k \geq u \circ \dot{\overline{f'}}^k$.

Now, note that $\dot{\overline{f}}^k \in \mathcal{C}(\{\dot{\overline{f}}^k, \dot{f}^k_c, \dot{\overline{f'}}^k, \dot{f}^k_{c'}\} \cup E^k)$. If not, then by Axiom 1 (IINIA), $\dot{f}^k_c \notin \mathcal{C}(\{\dot{\overline{f}}^k, \dot{f}^k_c, \dot{\overline{f'}}^k, \dot{f}^k_{c'}\} \cup E^k)$; since $E^k \cap \mathcal{C}(\{\dot{\overline{f}}^k, \dot{f}^k_c, \dot{\overline{f'}}^k, \dot{f}^k_{c'}\} \cup E^k) = \varnothing$, we must have that one of $\dot{\overline{f'}}^k, \dot{f}^k_{c'}$ belongs to $\mathcal{C}(\{\dot{\overline{f}}^k, \dot{f}^k_c, \dot{\overline{f'}}^k, \dot{f}^k_{c'}\} \cup E^k)$, which by Axiom 1 (IINIA), implies that $\dot{\overline{f'}}^k \in \mathcal{C}(\{\dot{\overline{f}}^k, \dot{f}^k_c, \dot{\overline{f'}}^k, \dot{f}^k_{c'}\} \cup E^k)$, a contradiction by Axiom 2 (Monotinicity) since $u \circ \dot{\overline{f}}^k \geq u \circ \dot{\overline{f'}}^k$.

By Axiom 1 (IINIA), $\dot{\overline{f}}^k \in \mathcal{C}(\{\dot{\overline{f}}^k, \dot{f}^k_c, \dot{\overline{f'}}^k, \dot{f}^k_{c'}\} \cup E^k) \implies \dot{f}^k_c \in \mathcal{C}(\{\dot{\overline{f}}^k, \dot{f}^k_c, \dot{\overline{f'}}^k, \dot{f}^k_{c'}\} \cup E^k)$, and so by Axiom 2 (Monotinicity), it must be the case that $c \geq c'$, and so we have $\varphi(\xi) \geq \varphi(\xi')$.

Now to extend this result to all of $\mathbb{R}^\Theta_{-1}$, take any $\xi, \xi' \in \mathbb{R}^\Theta_{-1}$ with $\xi \geq \xi'$. Take $\alpha > 0$ s.t. $\alpha\xi, \alpha\xi' \in [-2, 0]^\Theta$; we have

$$\begin{aligned}\varphi(\xi) &= \frac{1}{\alpha}\varphi(\alpha\xi) \\ &\leq \frac{1}{\alpha}\varphi(\alpha\xi') \\ &= \varphi(\xi')\end{aligned}$$

by Claim 3 (homogeneity of φ) and the preceding result.

△

By Claims 3-6, we are possession of a increasing, sublinear, C-additive, $\varphi : \mathbb{R}^\Theta_- \to \mathbb{R}$ with $\varphi(\xi_{-1}) < 0$; we also have $U(f|A) = \varphi(u \circ f - u \circ f^A_{max})$ for all $A \in \mathcal{A}, f \in A$. By Lemma 8, there exists $P \subseteq \mathcal{P}$ s.t. $\varphi(\xi) = \max_{p \in P} \sum_\theta \xi(\theta)p(\theta)$, and therefore

$$\begin{aligned}U(f|A) &= \max_{p \in P} \sum_\theta (u \circ f - u \circ f^A_{max})(\theta)p(\theta) \\ &= \max_{p \in P} \sum_\theta \left[u(f(\theta)) - \max_{f' \in A} u(f'(\theta))\right] p(\theta)\end{aligned}$$

for all $A \in \mathcal{A}$, $f \in A$. Furthermore, by Claim 1, U represents $\mathcal{C}$, and so for any $A \in \mathcal{A}$,

$$\begin{aligned}\mathcal{C}(A) &= \arg\max_{f \in A} U(f|A), \\ \text{for } U(f|A) &= \max_{p \in P} \sum_\theta \left[u(f(\theta)) - \max_{f' \in A} u(f'(\theta))\right] p(\theta)\end{aligned}$$

To see that P satisfies Assumption 6, suppose not; there exists $\delta_{\theta^*} \in P$. By the above, we have that for any $f \in A$ s.t. $f(\theta^*) > g(\theta^*)$ for all $g \in A$, $f \in \mathcal{C}(A)$, which contradicts Axiom 7 (No Certainty).

Lemma 2 then implies that $\mathcal{C}$ has a decisiveness-maximizing representation (u, P), where u satisfies Assumption 5 and P satisfies Assumption 6.

Now we prove the identification result. Fix some (u, P) and (u, P') that represent $\mathcal{C}$, where $\mathcal{C}$ additionally satisfies Axiom 7 (No Certainty) Axiom 6 (Non-Triviality). Since $\mathcal{C}$ satisfies Axiom 6 (Non-Triviality), u, u' must satisfy Assumption 5, and since $\mathcal{C}$ satisfies Axiom 7 (No Certainty), P, P' satisfy Assumption 6. The proof of the identification result proceeds in the following four steps.

Step 5: Identification of u.

First we show that u is identified up to an affine transformation using choice data on menus of constant acts. For any constant act $f \in X_c$, let $q_f \in \Delta(Z)$ denote the objective lottery induced by f. Note that for any finite menu of constant acts, $A \subset X_c$,

$$\begin{aligned}\mathcal{C}(A) &= \arg\max_{f \in A} u(q_f) \\ &= \arg\max_{f \in A} \sum_{z \in Z} u(z) q_f(z)\end{aligned}$$

Therefore, for any u, the restriction of $\mathcal{C}$ to menus of constant acts has a von-Neumann-Morgenstern (v-NM) representation, and so by the Expected Utility Theorem (von-Neumann and Morgenstern 1994), u is identified up to an affine transformation. As result, we can assume without loss for the remainder of the proof that $u = u'$.

Step 6: Utility acts and preliminaries.

Under Assumption 5, there exists $\underline{z}, \overline{z} \in Z$ s.t. $u(\underline{z}) = \min_{z \in Z} u(z)$, $u(\overline{z}) = \max_{z \in Z} u(z)$, with $u(\underline{z}) < u(\overline{z})$. Using the identification result from Step 1, we can without loss take an affine transformation of u so that $u(\underline{z}) = 0$, $u(\overline{z}) = 1$. Note that we can associate each act $f \in X$, with a utility act $\xi^f \in [0,1]^\Theta$ satisfying $\xi^f(\theta) = u(f(\theta))$ for all θ. Note that $\{\xi^f : f \in X\} = [0,1]^\Theta$. For the remainder of this proof we will work with choice data over menus of utility acts, dropping the f superscript.

For any constant utility act ξ_c delivering an interior utility value $c \in (0,1)$, define

$$\begin{aligned}B(\xi_c|P) &= \bigcap_{p \in P} \left\{\xi \in [0,1]^\Theta \ : \ \xi \cdot p < c\right\} \\ \overline{B}(\xi_c|P) &= \bigcap_{p \in P} \left\{\xi \in [0,1]^\Theta \ : \ \xi \cdot p \leq c\right\}\end{aligned}$$

Note that $\overline{B}(\xi_c|P) = cl(B(\xi_c|P))$. To see this, take any $\xi \in [0,1]^\Theta$ such that for all $p \in P$, $\xi \cdot p \leq c$; we will show that ξ is a limit point of elements in $B(\xi_c|P)$. Consider the sequence (ξ_k) satisfying $\xi_k = (1 - 1/k)\xi$; we have $\xi_k \in B(\xi_c|P)$ and $\xi_k \to \xi$. Since $B(\xi_c|P) \subset \overline{B}(\xi_c|P)$, $\overline{B}(\xi_c|P) = cl(B(\xi_c|P))$.

Step 7: Identification of $\overline{B}$.

Now, we will show that B, and subsequently $\overline{B}$, is identified from choice data. Fix any

ξ_c and ξ. If $\xi, \xi_c \in A$ and $\xi \in \mathcal{C}(A)$, then there must exist a model-induced posterior $p \in P$ s.t. $\xi \cdot p \geq c$, and so $\xi \notin B(\xi_c|P)$. For any $\lambda \in (0,1)$, let $\xi_\lambda = (1-\lambda)\xi + \lambda\xi_c$. The preceding implies that if $\xi_\lambda, \xi_c \in A$ and $\xi_\lambda \in \mathcal{C}(A)$, $\xi \notin B(\xi_c|P)$. We will show that if $\xi \notin B(\xi_c|P)$, there will always exist a constant $\lambda \in (0,1)$ and menu A containing ξ_c, ξ_λ such that $\xi_\lambda \in \mathcal{C}(A)$, which delivers identification of $B(\xi_c|P)^C$ and therefore $B(\xi_c|P)$.

Suppose that there exists some $p^* \in P$ s.t. $\xi \cdot p \geq c$. Take any sequence $\epsilon_k \to 0$, where $\epsilon_k \in (0, 1-c)$ for all k. For all k, there exists $\lambda_k \in (0,1)$ satisfying $\max_{\theta\in\Theta} \xi_{\lambda_k}(\theta) < c + \epsilon_k$; fix such a sequence of λ_k. For each k, define a set of acts $E_k = (\overline{\xi}_{k,\theta})_{\theta\in\theta}$ where for each θ, k,

$$\overline{\xi}_{k,\theta}(\theta') = \begin{cases} 0 & \theta \neq \theta' \\ c + \epsilon_k & \theta = \theta' \end{cases}$$

Define a sequence of menus $A_k = E_k \cup \{\xi_{\lambda_k}, \xi_c\}$. Note that for all A_k,

$$\begin{aligned} \mathcal{I}(P|A_k) &= \arg\max_{p\in P} \left\{ \max_{\xi\in A_k} \sum_{\theta\in\Theta} \left[\xi(\theta) - \max_{\xi'\in A_k} \xi'(\theta) \right] p(\theta) \right\} \\ &= \arg\max_{p\in P} \left\{ \max_{\xi\in A_k} \sum_{\theta\in\Theta} \left[\xi(\theta) - (c+\epsilon_k) \right] p(\theta) \right\} \\ &= \arg\max_{p\in P} \left\{ \max_{\xi\in A_k} \xi \cdot p \right\} \end{aligned}$$

Since P is closed and no $p \in P$ places probability 1 in any state by Assumption 6, there exists some K such that for all $k > K$, $\max_{p\in P} \left\{ \overline{\xi}_{k,\theta} \cdot p \right\} \leq c$ for all θ. In particular, this implies that for $k > K$,

$$\max_{p\in P} \left\{ \overline{\xi}_{k,\theta} \cdot p \right\} \leq c \leq \xi_{\lambda_k} \cdot p^*$$

which in turn implies that for all $k > K$, any $p \in \mathcal{I}(P|A_k)$ must choose ξ_{λ_k} from A_k. Using choice data on such sequences of menus, for any ξ_c, ξ, we can determine whether $\xi \in B(\xi_c|P)$, therefore identifying $B(\xi_c|P)$ and subsequently $\overline{B}(\xi_c|P)$, the closure of $B(\xi_c|P)$.

Formally, since (u, P) and (u', P') both represent $\mathcal{C}$, it must be the case that $B(\xi_c|P) = B(\xi_c|P')$ since B is identified from $\mathcal{C}$; this implies that $cl(B(\xi_c|P)) = cl(B(\xi_c|P')) \implies \overline{B}(\xi_c|P) = \overline{B}(\xi_c|P')$.

Step 8. Identification of Extreme Models.

Let $\mathcal{P}_{ext} = \{ext(P) : P \in \mathcal{P}\}$ denote the collection of extreme points formed from elements in $\mathcal{P}$, and let $\mathcal{P}_{conv} = \{P \in \mathcal{P} : P \text{ is convex}\}$ denote the set of convex elements in $\mathcal{P}$. First, we will show that $co : \mathcal{P}_{ext} \to \mathcal{P}_{conv}$ is one-to-one. To see this, note that for $P \in \mathcal{P}_{ext}$, $ext(co(P)) = P$: $p \in P \implies p \in ext(co(P))$ since all points in P are extreme by construction. Also, for $P \in \mathcal{P}_{conv}$, since P is a convex, closed subset of $\Delta(\Theta)$ and is therefore compact, by the Minkowski–Caratheodory Theorem $co(ext(P)) = P$. This implies that $co : \mathcal{P}_{ext} \to \mathcal{P}_{conv}$ is one-to-one as desired.

Now, towards a contradiction, suppose that $ext(P) \neq ext(P')$. Define $P_{conv} = co(ext(P))$ and $P'_{conv} = co(ext(P'))$; by the above result, we have $P_{conv} \neq P'_{conv}$. Since P is closed and therefore compact, $co(P)$ is also compact. Further, since $ext(P) = ext(co(P))$, we have

$$\begin{aligned} P_{conv} &= co(ext(P)) \\ &= co(ext(co(P))) \\ &= co(P) \end{aligned}$$

where the last step follows from the Minkowski–Caratheodory Theorem. Therefore P_{conv} is compact, and by the same argument, so is P'_{conv}. Since $P_{conv} \neq P'_{conv}$, we can without loss of generality take there to be $p \in P_{conv} \setminus P'_{conv}$. By a separating hyperplane theorem (Dunford and Schwartz, 1957, Theorem V.2.10), there exists a nonzero $\xi \in \mathbb{R}^\Theta$ and a $c \in \mathbb{R}$ such that

$$\max_{p' \in P'_{conv}} \xi \cdot p' < c < \xi \cdot p$$

Without loss of generality, we can take an affine transformation of ξ and c so that $\xi \in [0,1]^\Theta$ and $c \in (0,1)$. This then implies that $B(\xi_c|P_{conv}) \neq B(\xi_c|P'_{conv})$. By Proposition 1, the fact that (u, P) and (u, P') represent $\mathcal{C}$ implies that (u, P_{conv}) and (u, P'_{conv}) also represent $\mathcal{C}$. Since B is identified from $\mathcal{C}$, it must be the case that $B(\xi_c|P_{conv}) = B(\xi_c|P'_{conv})$; we have a contradiction and it must be the case that $ext(P) = ext(P')$. □

Now, we give a proof of Theorems 2 and 3.

Theorem 2. C satisfies Axioms 1–6 if and only if it has a decisiveness maximizing representation (u, P), where u satisfies Assumption 5. Also, for any (u', P') representing $\mathcal{C}$, there exists constants $\alpha > 0$, β such that $u = \alpha u' + \beta$, and $ct(P) = ct(P')$.

Proof. The proof for the characterization result is contained in Steps 1–4 from the proof for Theorem 1.

To show the identification property, fix some $(u, P), (u', P')$ that represent $\mathcal{C}$. Following Step 5 from the proof of Theorem 1, we know that u is identified up to an affine transformation, and as a result we can assume without loss for the remainder of the proof that $u = u'$. Following Step 6 from the proof of Theorem 1, construct the set of utility acts $[0,1]^\Theta$; for the remainder of this proof we will work with choice data over menus of utility acts.

Fix any $\theta \in \Theta$. Let ξ_c denote an constant utility act satisfying $\xi_c(\theta') = c \in (0,1)$ for all $\theta' \in \Theta$. Take any sequence $\epsilon_k \to 0$ satisfying $\epsilon_k \in (0, 1-c)$ for all k. Define a sequence of utility acts $\overline{\xi}_{k,\theta}$ where for each k,

$$\overline{\xi}_{k,\theta}(\theta') = \begin{cases} 0 & \theta \neq \theta' \\ c + \epsilon_k & \theta = \theta' \end{cases}$$

Define a sequence of menus $A_k = \{\xi_c, \overline{\xi}_{k,\theta}\}$. We will show that $\delta_\theta \notin P$ iff there exists K s.t. for all $k > K$, $\overline{\xi}_{k,\theta} \notin \mathcal{C}(A_k)$. First, suppose that $\delta_\theta \notin P$. Since P is closed by Assumption 2, there exists some $\overline{p} < 1$ such that $\max_{p \in P} p(\theta) \leq \overline{p} < 1$ and so for all $p \in P$, $\overline{\xi}_{k,\theta} \cdot p \leq \overline{p}(c+\epsilon_k)$. Since $\overline{p} < 1$ and $\epsilon_k \to 0$, there exists K s.t. for all $k > K$, $\overline{p}(c+\epsilon_k) < c$, which implies that $\overline{\xi}_{k,\theta} \notin \mathcal{C}(A_k)$. Conversely, suppose that $\delta_\theta \in P$. This implies $\delta_\theta \in \mathcal{I}(P|A)$ for any menu A, and so $\overline{\xi}_{k,\theta} \in \mathcal{C}(A_k)$ for all k.

Now towards a contradiction, suppose that $ct(P) \neq ct(P')$. Without loss of generality, we can take there to be some $\delta_\theta \in P \setminus P'$. By the above result this means that (u, P) and (u, P') cannot represent the same choice correspondence $\mathcal{C}$, a contradiction; $ct(P) = ct(P')$ as desired. □

Theorem 3. Suppose $\mathcal{C}, \mathcal{C}'$ are represented by (u, P), (u, P'), respectively. If $co(P) \subseteq co(P')$, then $\mathcal{C}'$ is more diversification-averse than $\mathcal{C}$. Furthermore, if $\mathcal{C}'$ is more diversification-averse than $\mathcal{C}$ and P' satisfies Assumption 6, then $co(P) \subseteq co(P')$.

Proof. Begin by proving the first statement. Suppose that $co(P) \subseteq co(P')$, and suppose that for some $h \in H_A$, $h \notin \mathcal{C}(A \cup \{h\})$. By Lemma 4, $\succeq^*$ is represented by a linear u, and note that $h \notin \mathcal{C}(A \cup \{h\})$ implies that u satisfies Assumption 5. Take some $\overline{z} \in \arg\max_{z \in Z} u(z)$ and $\underline{z} \in \arg\min_{z \in Z} u(z)$; since u satisfies Assumption 5 and by linearity of u, we can without loss of generality take $u(\overline{z}) = 1$, $u(\underline{z}) = -1$. For any $c \in [-1, 1]$, let $q(c) = \frac{1-c}{2} \circ \underline{z} + \frac{1+c}{2} \circ \overline{z}$; by linearity of u, $u(q(c)) = c$. let f_c denote the constant act with $f_c(\theta) = q(c)$ for all θ. Let $\overline{f}$ be the *normalized* act satisfying $\overline{f}(\theta) = q(\frac{1}{4}[u(f(\theta)) - u(f^A_{max}(\theta))])$. Let $\overline{A} = \{\overline{f} : f \in A\}$ collect the normalized acts in A.

Since $h \in H_A$, there exists $\underline{q}, \overline{q} \in \Delta(Z)$ with $u(\underline{q}) - u(\overline{q}) \equiv c \leq 0$ such that $\frac{1}{2}u(f^A_{max}(\theta)) + \frac{1}{2}u(\underline{q}) = \frac{1}{2}u(g(\theta)) + \frac{1}{2}u(\overline{q})$ for all θ, which in turn implies that $u(g(\theta)) = u(f^A_{max}(\theta)) + c$, and subsequently $\overline{g} = f_{\tilde{c}}$, for $\tilde{c} = \frac{1}{4}c$. Note that

$$\begin{aligned}\max_{p \in P} \sum_\theta \left[u(\overline{h}(\theta)) - \max_{f' \in \overline{A} \cup \{\overline{h}\}} u(f'(\theta))\right] p(\theta) &= \max_{p \in P} \sum_\theta \left[u(\overline{h}(\theta))\right] p(\theta) \\ &= \max_{p \in P'} \sum_\theta \left[u(\overline{h}(\theta))\right] p(\theta) \\ &= \max_{p \in P'} \sum_\theta \left[u(\overline{h}(\theta)) - \max_{f' \in \overline{A} \cup \{\overline{h}\}} u(f'(\theta))\right] p(\theta)\end{aligned}$$

whereas for any $\overline{f} \in \overline{A}$,

$$\begin{aligned}\max_{p \in P} \sum_\theta \left[u(\overline{f}(\theta)) - \max_{f \in \overline{A} \cup \{\overline{f}\}} u(f(\theta))\right] p(\theta) &= \max_{p \in P} \sum_\theta \left[u(\overline{h}(\theta))\right] p(\theta) \\ &\leq \max_{p \in P'} \sum_\theta \left[u(\overline{f}(\theta))\right] p(\theta) \\ &= \max_{p \in P'} \sum_\theta \left[u(\overline{h}(\theta)) - \max_{f \in \overline{A} \cup \{\overline{f}\}} u(f(\theta))\right] p(\theta)\end{aligned}$$

since $co(P) \subseteq co(P')$. By Lemma 2, we have $\overline{h} \notin \mathcal{C}(\overline{A} \cup \{\overline{h}\}) \implies \overline{h} \notin \mathcal{C}'(\overline{A} \cup \{\overline{h}\})$. Since $\mathcal{C}, \mathcal{C}'$ satisfy Axiom 4 (Mixture Independence), we have $\overline{h} \notin \mathcal{C}(\overline{A} \cup \{\overline{h}\}) \iff h \notin \mathcal{C}(A \cup h)$, and likewise $\overline{h} \notin \mathcal{C}'(\overline{A} \cup \{\overline{h}\}) \iff h \notin \mathcal{C}'(A \cup h)$, and so $h \notin \mathcal{C}(A \cup \{h\}) \implies h \notin \mathcal{C}'(A \cup \{h\})$ as desired.

Now prove the second statement. Suppose that $\mathcal{C}'$ is more diversification averse than $\mathcal{C}$, and P' satisfies Assumption 6. Towards a contradiction, suppose that $co(P) \nsubseteq co(P')$; this guarantees the existence of some $p \in co(P) \setminus co(P')$, where $p \in P$. By a separating hyperplane theorem (Dunford and Schwartz, 1957, Theorem V.2.10), there exists a nonzero $\xi \in \mathbb{R}^{\Theta}$ and $c \in \mathbb{R}$ s.t.

$$\max_{p' \in P'} \xi \cdot p' < c < \xi \cdot p$$

Without loss of generality we can take $c \in (-1, 1)$. Since P' satisfies Assumption 6, there exists $k > c$ s.t. for acts $(g_\theta)_{\theta \in \Theta}$ satisfying

$$g_\theta(\theta') = \begin{cases} q(k) & \theta' = \theta \\ q(-1) & \text{otherwise} \end{cases}$$

such that $(u \circ g_\theta) \cdot p' < c$ for all $p' \in P'$ and all θ. By rescaling, we can without loss of generality take $\xi \in [-1, 1]^{\Theta}$ with $\max_\theta \xi(\theta) < k$ satisfying the above inequality. Let f denote an act satisfying $u \circ f = \xi$, and let $A = \{f\} \cup \{g_\theta\}_{\theta \in \Theta}$; we that $h \equiv f_c$ is C-diversified wrt. A.

Since $c < \xi \cdot p$, $h \notin \mathcal{C}(A \cup \{h\})$. Note, however, since $\max_{p' \in P'} \xi \cdot p' < c$ and also $\max_{p \in P'} (u \circ g_\theta) \cdot p' < c$, $h = \mathcal{C}'(A \cup \{h\})$. This implies that $\mathcal{C}'$ is not more diversification averse than $\mathcal{C}$, a contradiction; we have $co(P) \subseteq co(P')$ as desired. $\square$