Essays in Economic Theory

by

Giacomo Lanzani

Submitted to the Department of Economics

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Economics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2023

Authored by: Giacomo Lanzani

Department of Economics

May 12, 2023

Certified by: Drew Fudenberg

Paul A. Samuelson Professor of Economics

Thesis Supervisor

Certified by: Stephen Morris

Peter A. Diamond Professor of Economics

Thesis Supervisor

Accepted by: Abhijit Banerjee

Ford International Professor of Economics

‘Chairman, Department Committee on Graduate Theses

Essays in Economic Theory

by

Giacomo Lanzani

Submitted to the Department of Economics
on May 12, 2023, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Economics

Abstract

The first chapter of this thesis considers an agent who posits a set of probabilistic models for the payoff-relevant outcomes. The agent has a prior over this set but fears the actual model is omitted and hedges against this possibility. The concern for misspecification is endogenous: If a model explains the previous observations well, the concern attenuates. We show that different static preferences under uncertainty (subjective expected utility, maxmin, robust control) arise in the long run, depending on how quickly the agent becomes unsatisfied with unexplained evidence and whether they are misspecified. The misspecification concern's endogeneity naturally induces behavior cycles, and we characterize the limit action frequency. This model is consistent with the empirical evidence on monetary policy cycles and choices in the face of complex tax schedules. Finally, we axiomatize in terms of observable choices this decision criterion and how quickly the agent adjusts their misspecification concern.

The second chapter offers an axiomatization of risk models where the choices of the decision maker are correlation sensitive. By extending the techniques of conjoint measurement to the nondeterministic case, we show that transitivity is the vN-M axiom that has to be relaxed to allow for these richer patterns of behavior. To illustrate the advantages of our modeling choice, we provide a simple axiomatization for the salience theory model within our general framework. This approach leads to a clear comparison to popular preexisting models, such as regret and reference dependence, and lets us single out the ordering property as the feature that brings salience theory outside the prospect theory realm. This chapter is published in the Quarterly Journal of Economics, vol 137.

The third chapter proposes a model of non-Bayesian social learning in networks that accounts for heuristics and biases in opinion aggregation. The updating rules are represented by nonlinear opinion aggregators from which we extract two extreme networks capturing strong and weak links. We provide graph-theoretic conditions on these networks that characterize opinions' convergence, consensus formation, and efficient or biased information aggregation. Under these updating rules, agents may ignore some of their neighbors' opinions, reducing the number of effective connections and inducing long-run disagreement for finite populations. For the wisdom of

the crowd in large populations, we highlight a trade-off between how connected the society is and the nonlinearity of the opinion aggregator. Our framework bridges several models and phenomena in the non-Bayesian social learning literature, thereby providing a unifying approach to the field. This chapter is the result of joint work with Simone Cerreia-Vioglio and Roberto Corrao.

JEL codes: D9

Thesis Supervisor: Drew Fudenberg
Title: Paul A. Samuelson Professor of Economics

Thesis Supervisor: Stephen Morris
Title: Peter A. Diamond Professor of Economics

Acknowledgments

I am indebted to many people who supported me throughout my Ph.D.

I am immensely lucky to have Drew Fudenberg and Stephen Morris as my advisors. They have been invaluable mentors and role models. The innumerable meetings in their offices pushed me to think beyond my intellectual comfort zone and develop the ideas contained in this thesis. Their devotion to their advisees inspires me, and I hope to be able to bring this legacy forward. Drew is also a passionate and extremely engaging collaborator. I hope this relationship will continue for many years.

I also learned so much from collaborating with my external advisors, Simone Cerreia-Vioglio and Philipp Strack. I hope I will be able to adhere to their scientific rigor and perseverance. Finally, I cannot exaggerate the importance of my friend, roommate, and coauthor, Roberto Corrao. Throughout five years together at MIT, every research project I developed originated or has been improved from our endless conversations.

Beyond the academic sphere, I am extremely indebted to my parents, Arturo and Patrizia, and my brother, Tommaso. They made it possible to accomplish this degree by making my youth nurturing, stimulating, and above all, happy.

Finally, this thesis is dedicated to my wife, Francesca. She transformed my life for the better. She has been my constant companion in every new venture, new country, new job, and now, new family. Her love and courage are my inspiration and source of strength.

Contents

Chapter 1

Dynamic Concern for Misspecification

1.1 Introduction

Bayesian rationality requires that an agent uncertain about the data-generating process postulates multiple probabilistic descriptions of the environment and uses Bayes rule to adjust their relative weights. However, even rational agents may fear that they are misspecified and that none of these descriptions is correct. This concern is remarkably natural in complex and high-dimensional settings, where uncertainty needs to be simplified to obtain well-behaved optimization and learning procedures.

For example, none of the model economies considered by a central bank perfectly describes the underlying data-generating process for output and inflation. Similarly, the consumer response models that a firm uses to set prices and qualities are unlikely to include one that considers all relevant decision factors. Moreover, the diffusion of complex and not explicitly described machine learning algorithms naturally creates new reasons for misspecification. Indeed, consumers increasingly rely on automated recommendations. Although they may have some conjecture on how the alternative's features translate into a score or a "match quality" with their profile, they certainly do not consider the specific algorithm used by these recommendation systems. Misspecification is even more relevant when dealing with entirely novel issues, such as those faced by a regulatory body that tries to mitigate the effect of climate change using theoretical models that take into account human impacts never experienced in

history.

Misspecification has been analyzed from two distinct perspectives. On the one hand, several papers have studied the long-run implications of subjective expected utility (SEU) maximizers learning with misspecified beliefs (see, e.g., Esponda and Pouzo, 2016, Fudenberg, Lanzani, and Strack, 2021, Frick, Iijima, and Ishii, 2023, and the references therein). These works assume that the agents have no concern about being misspecified. Here we show that the absence of such concern is normatively unappealing, as it can induce long-run average payoffs lower than a safe guarantee. It also seems descriptively unrealistic, as the widely documented ambiguity-averse behavior may be seen as a way to hedge against the incorrect specification of the model. On the other hand, the robust control literature in macroeconomics pioneered by Hansen and Sargent (2001) considers agents who fear model misspecification. In particular, the first axioms-based decision criterion that accounts for model misspecification was proposed in Cerreia-Vioglio, Hansen, Maccheroni, and Marinacci (2022).[1]

This work reconciles these approaches and shows how popular decision criteria such as maxmin expected utility, robust control preferences, and subjective expected utility arise as the limit behavior of an agent concerned about misspecification and learning about the actual data-generating process (DGP). We consider an agent that repeatedly chooses among actions whose payoffs have an unknown distribution. This choice is taken using an average of robust control assessments, where each assessment takes a different structured model as the benchmark. We introduce endogeneity in the misspecification concern: the better the structured models explain the past, the less concerned the agent is.

There are two critical determinants for the long-run dynamics: whether the agent is correctly specified and how demanding they are in evaluating their models' performance. First, we consider the case of a correctly specified agent. In that case, the behavior converges to a self-confirming equilibrium, regardless of how demanding the agent is in evaluating their model. A self-confirming equilibrium means that they

[1]It has as a particular case the robust control model of Hansen and Sargent (2001) axiomatized by Strzalecki (2011). Since in Strzalecki (2011) the reference probability is subjective, it can also be interpreted as an axiomatization of robust prior analysis, see Hansen and Sargent (2022).

play an SEU best reply to a belief supported over the data-generating processes that are observationally equivalent to the true one given the chosen action.

Instead, to characterize the limit behavior under misspecification, a taxonomy of how demanding the agent is turns out to be crucial. In particular, a "statistically sophisticated" agent performs a likelihood ratio evaluation of their model that keeps the concern form misspecification informative about the model's fitness. To support the identification of these statistically sophisticated types with rationality, we show that the achievement of two desirable properties uniquely characterizes them: safety under misspecification (i.e., guaranteeing at least the minmax payoff) and consistency under almost correct specification (i.e., no regret with small misspecification).[2]

We allow departures from this normative benchmark to obtain descriptive predictions on the effect of an endogenous misspecification concern. We consider agents that are too demanding in evaluating the models' performance (this case includes believers in the Law of Small Numbers, LSN, Tversky and Kahneman, 1971, that treat failures in explaining early realizations as a statistician treats long-run failures). Similarly, we allow the opposite case in which the agent is too lenient in evaluating their model and attributes too much unexplained evidence to sampling variability.

We then characterize the long-run behavior of these different types of misspecified agents. The actions of the lenient type converge to a Berk-Nash equilibrium, i.e., to an SEU best reply to beliefs supported on the models closest in relative entropy to the actual data-generating process. Instead, overemphasis on the model's failures in explaining the data by the demanding type induces convergence to a maxmin best reply to the models that are absolutely continuous with respect to the true one.

In contrast, a statistically sophisticated type maintains a non-trivial concern for misspecification. If their behavior converges, it converges to a robust control best reply to the models closest in relative entropy to the actual data-generating process. Moreover, the misspecification concern is endogenously determined by how well the best models fit the evidence generated by the limit action.

[2]Moreover, we observe that SEU maximization and the original robust control of Hansen and Sargent (2001) fail to jointly satisfy these requirements.

Therefore, our learning results provide several novel predictions about the relation between uncertainty attitudes and other individual traits.[3] First, the extent of long-run uncertainty aversion positively correlates with the agent being initially misspecified and their belief in the LSN. Second, these correlations are causal: repeated failures to explain the data (misspecification) and demanding evaluation of these failures induce the agent to shift to cautious behavior. Third, even keeping constant the misspecification and understanding of probability rules, the limit uncertainty attitudes are stochastic. Initial realizations leading to a limit action with consequences poorly explained by the agent's models induce a long-run uncertainty aversion higher than realizations leading to a limit action whose consequences are well explained.

We thus use the equilibrium behavior predicted by an endogenous concern for misspecification to rationalize the labor supply in the face of complex tax schedules documented in Rees-Jones and Taubinsky (2020). In particular, they show that around 40% of the agents have beliefs corresponding to a heuristic that simplifies the tax schedule to a linear one but that 20% fewer agents act accordingly to this heuristic. This is predicted by an endogenous concern for misspecification, as agents with an incorrect model are less prone to base their decisions on the conclusions they reach within the model.

In general, the behavior of a statistically sophisticated type is not guaranteed to converge. Indeed, it is possible that their behavior cycles between phases of different misspecification concerns. Still, we characterize the limit action frequency and concern for misspecification. We apply this result to revisit the cyclical behavior of monetary policies documented in Sargent (1999) and Sargent (2008). Intuitively, the cycles have the following structure. The agent plays an action whose consequences are well explained by one of their structured models (a conservative monetary policy in the application). Playing this action lowers the concern for misspecification and eventually leads to a more misspecification-vulnerable action (a more aggressive monetary policy). Failures to explain the distribution of outcomes observed under this

[3]The empirical study of the correlation between behavioral biases is an active area of recent development. See, e.g., Dean and Ortoleva (2019) and the references therein.

action lead to a return to the more misspecification robust action.

We also obtain two results that provide a testable foundation to the model employed in the learning part of the chapter: An axiomatization of the static average robust control criterion and testable axioms for when the agent is of the lenient, statistically sophisticated, or demanding type. Two primary axioms pin down the static decision criterion. The first is a weaker form of the Sure-Thing Principle imposed only on bets on the data-generating process (e.g., bets on the urn composition) and bets conditional on the data-generating process (e.g., bets on the ball color conditional on having been told the urn composition). The second requires that conditional on being told the best-fitting model, the agent is equally concerned about misspecification regardless of which one it is.

For the dynamic representation, a dynamic consistency axiom on the acts that bet on the data-generating process is shown to guarantee Bayesian updating over models. More interestingly, the preference adjustment of a statistically sophisticated type is pinned down by a novel Asymptotic Frequentism axiom, requiring arbitrarily similar preferences conditional to sufficiently long histories with the same outcome frequency.

The rest of the chapter is structured as follows. Section 1.2 introduces the average robust control decision criterion and how preferences are adjusted. Section 1.3 studies what attitudes toward model failures induce good payoff performance and provides a learning foundation for the different uncertainty attitudes. Section 1.4 characterizes the limit frequency of time spent using the different actions when the behavior does not converge and applies the result to a central banking problem. Section 1.5 provides the axiomatization to the static decision criterion and how preferences are updated. Section 1.6 discusses the related literature and possible extensions. Section 1.7 concludes. All proofs are collected in the Appendix.

1.2 Decision Criterion

1.2.1 Static Decision Criterion

We describe the criterion used in the repeated decision problem and defer its axiomatization to Section 1.5. We consider an agent who evaluates a finite number of actions $a \in A$ and let Y be a compact metric space representing the set of possible outcomes. The agent has a continuous utility index $u : A \times Y \to \mathbb{R}$ over the action-outcome pairs that captures their preference when the subjective uncertainty is resolved. However, the realized outcome is stochastic and endogenous as each action $a \in A$ induces an objective probability measure $p_a^* \in \Delta(Y)$ over outcomes.[4]

Subjective Beliefs The agent correctly believes that the map from actions to probability distributions over outcomes is fixed and depends only on their current action. Still, they do not know $p^* = (p_a^*)_{a \in A}$ and deal with this uncertainty in a quasi-Bayesian way. The agent postulates a set $Q \subseteq \Delta(Y)^A$ of *structured models,* i.e., action-dependent probability measures over outcomes $q = (q_a)_{a \in A}$. They have a prior belief $\mu \in \Delta(Q)$ with support Q that describes the relative likelihood assigned to these models. For example, the agent may be a central bank that considers a Keynesian Samuelson-Solow model where the monetary policy affects the unemployment rate or a new classical Lucas-Sargent model with no systematic effect of inflation on unemployment.[5]

We must impose a few regularity conditions.

Assumption 1. Q is compact and for every $a \in A$: (i) For all $q \in Q$, $p_a^* \sim q_a$ and the density of q_a with respect to p_a^*, denoted as $\tilde{q}_a$, is continuous and p_a^*-a.s. bounded

[4] For every subset C of a metric space, we denote as $\Delta(C)$ the Borel probability measures on C, endowed with the topology of weak convergence of measures.

[5] This formulation follows the recent literature on misspecified learning in assuming that both the true data-generating process and the subjective models the agent considers are i.i.d. conditionally on the agent's behavior. This makes the true extent of misspecification time-invariant and "learnable". It is important to notice that this time invariance is often relaxed in the literature on dynamic decisions with robust control preferences that follows Hansen and Sargent (2001).

away from 0, uniformly in Q,[6] (ii) For p_a^*-almost every $y \in Y$ the map $q \mapsto \tilde{q}_a(y)$ is continuous.

Condition (i) allows us to compute the relevant expectations while allowing for both discrete and continuous outcome spaces and guarantees that no subjective model of the agent is ruled out in finite time.[7] Continuity of the map from models to outcome distributions is a standard requirement for parametric models.

A Bayesian agent with complete trust in their models evaluates action a according to its subjective expected utility (see, e.g., Cerreia-Vioglio, Maccheroni, Marinacci, and Montrucchio, 2013b):

$$\int_Q \mathbb{E}_{q_a}[u(a,y)] \, \mathrm{d}\mu(q).$$

That is, they compute a two-stage expectation of the utility function: they evaluate the utility of the action given the candidate model q, $\mathbb{E}_{q_a}[u(a,y)]$, and then they average over the models with weights given by their subjective belief μ.

However, we are interested in agents concerned with the possibility that none of these models is the exact description of the data-generating process but only a valid approximation, i.e., that are concerned that there is no $q \in Q$ with $q = p^*$. Therefore, in the spirit of the robustness criterion advocated by Hansen and Sargent (2001), they penalize actions that perform poorly under alternative distributions that are close in relative entropy $R(\cdot||\cdot)$ to some of the structured models.[8]

With this, an agent evaluates each action $a \in A$ accordingly to the *average robust control* criterion:

$$\int_Q \min_{p_a \in \Delta(Y)} \left(\mathbb{E}_{p_a}[u(a,y)] + \frac{1}{\lambda} R(p_a||q_a) \right) \mathrm{d}\mu(q) \tag{1.1}$$

[6]For every $p, q \in \Delta(Y)$, $p \gg q$ means that q is absolutely continuous with respect to p, and $p \sim q$ means that they are mutually absolutely continuous.

[7]Part (i) also plays a technical role in guaranteeing the existence of the equilibrium concepts we consider. It is known it can be relaxed, see Anderson, Duanmu, Ghosh, and Khan (2022), but this relaxation comes at the cost of requiring nonstandard analysis techniques (where nonstandard means using infinitesimal numbers), something beyond this chapter's scope.

[8]Recall that for every $p, p' \in \Delta(Y)$, $R(p||p') = \int_Y \log\left(\frac{\mathrm{d}p}{\mathrm{d}p'}\right) \mathrm{d}p$ if $p' \gg p$ and $R(p||p') = \infty$ otherwise. Appendix .1.3 explains under what other distances between probability distributions our results continue to hold.

where $\lambda > 0$ is a parameter that trade-offs between decision robustness and performance under the structured models.[9]

The original robust control model introduced by Hansen and Sargent (2001) is the case in which μ is a Dirac measure (that in macroeconomics applications is often assumed to satisfy rational expectations, i.e., to be degenerate on the actual data-generating process). As described in Hansen, Sargent, Turmuhambetova, and Williams (2006), this case corresponds to when "[...] a maximizing player ('the decision maker') chooses a best response to a malevolent player ('nature') who can alter the stochastic process within prescribed limits. The minimizing player's malevolence is the maximizing player's tool for analyzing the fragility of alternative decision rules." Equation (1.1) follows Hansen and Sargent (2007) and Cerreia-Vioglio, Hansen, Maccheroni, and Marinacci (2022) in extending this interpretation to a situation in which the agent is still uncertain about the best-approximating model (i.e., μ is nondegenerate), allowing the malevolent nature to alter each of the candidate structured models.

The representation adopts the distinction between two levels of uncertainty. At the first level, given a probabilistic model q, the uncertainty about the exact specification of the model is captured by minimizing the expected utility for probabilities that are not too far away from q. At a higher level, the agent is also uncertain about the identity of the best structured model and posits a prior probability μ over them. While the higher level of uncertainty is already present under subjective expected utility, the lower level captures the agent's concern for misspecification.

1.2.2 Preference Evolution

The average robust control criterion of equation (1.1) describes how the agent chooses for a *given* belief and level of misspecification concern. However, the behavior responds to the received information. Formally, time is discrete, and a history is a finite vector of past actions and outcomes. In particular, the set of histories of finite length $t \in \mathbb{N}$ is $\mathcal{H}_t = (A \times Y)^t$, and the set of all finite histories is $\mathcal{H} = \bigcup_{t=0}^{\infty} \mathcal{H}_t$.

[9]Lemma 1 justifies the use of a min rather than an inf in equation (1.1) and throughout the chapter.

We will denote with $\mathbf{a}_t, \mathbf{y}_t$, and $\mathbf{h}_t$ the random variables corresponding to the action, outcome, and history at time t, and we use the non-bold version for their realizations.

On the one hand, we stick to the classical dynamic treatment of tastes over certain alternatives and beliefs about the possible data-generating processes. We let the utility index u be constant over time, and the belief be updated through standard Bayesian updating. That is, for every measurable subset C of Q, we denote by

$$\mu(C \mid (a^t, y^t)) = \frac{\int_{q\in C} \prod_{\tau=1}^{t} \tilde{q}_{a_\tau}(y_\tau) \mathrm{d}\mu_0(q)}{\int_{q\in Q} \prod_{\tau=1}^{t} \tilde{q}_{a_\tau}(y_\tau) \mathrm{d}\mu_0(q)} \qquad \text{(Bayes Rule)}$$

the subjective belief the agent obtains using Bayes rule after history $(a^t, y^t) \in \mathcal{H}_t$.[10]

On the other hand, we introduce an endogenous and time-evolving concern for misspecification, i.e., λ depends on the realized history. In particular, we want the concern for misspecification to be a function of how well the structured models explain the current history, i.e., to be determined by a function $\Lambda : \mathcal{H} \to \mathbb{R}_+$.

Likelihood Ratio Test In statistics, the most standard measure of fit of a set of distributions Q against a set of unstructured alternatives $N(Q) \subseteq \Delta(Y)^A$ is the log-likelihood ratio:[11]

$$LLR((a^t, y^t), Q) = -\log\left(\frac{\max_{q\in Q} \prod_{\tau=1}^{t} \tilde{q}_{a_\tau}(y_\tau)}{\max_{p\in N(Q)} \prod_{\tau=1}^{t} \tilde{p}_{a_\tau}(y_\tau)}\right) \qquad \forall t \in \mathbb{N}, \forall (a^t, y^t) \in \mathcal{H}_t.$$

Here we want to take a conservative approach and not impose structure over the set of alternative unstructured distributions $N(Q)$ used to evaluate the model's fit. If Y is finite (or, under some regularity conditions, countable) and all outcomes have positive probability, there is a natural way to do so, i.e., to consider as the set of

[10] By Assumption 1 (ii), the posterior is well-defined after every positive probability history. We allow for arbitrary belief revisions after events with zero ex-ante subjective probability.

[11] The Neyman-Pearson Lemma establishes the performance of the log-likelihood ratio test under correct specification. At the same time, Foutz and Srivastava (1977) and Vuong (1989) contain the classical results about the informativeness of the LRT under misspecification. Schwartzstein and Sunderam (2021) is a recent paper that models agents in a persuasion problem who perform model selection using this statistic. Lemma 1 justifies the use of max in the definition of the LRT under Assumption 2.

alternatives unstructured distributions the entire (action-indexed) simplex $\Delta(Y)^A$. However, considering a completely unrestricted set of distributions with a continuum of outcomes leads to an utterly uninformative test of the model, as the (discrete) empirical distribution is an infinitely better fit to itself than any continuous distribution, i.e., the log-likelihood ratio always returns $+\infty$. To maintain informativeness, $N(Q)$ must then include only distributions that are mutually absolutely continuous with respect to the ones in Q. In particular, all our results are invariant to the $N(Q)$ choice as long as the following assumption is satisfied.

Assumption 2. (i) $N(Q) \supseteq Q$ is closed and $p^* \in N(Q)$. (ii) For every $a \in A$, the family of densities $\{\tilde{p}_a : p \in N(Q)\}$ is equicontinuous.

We require that the unstructured set is a relaxation of the parametric structure sufficiently large to include the actual distribution and a continuity condition that rules out a Q that only contains continuous distributions and an $N(Q)$ that includes discrete distributions. With this, an important role will be played by the rule

$$\Lambda(h_t) = \frac{\mathrm{LRT}(h_t,Q)}{ct} \qquad \forall t \in \mathbb{N}, \forall h_t \in \mathcal{H}_t \tag{1.2}$$

where $c \in \mathbb{R}$.

Beyond the log-likelihood ratio, a key role is played by averaging over periods. Indeed, if the agent is misspecified, the expected one-period increase in the LRT is strictly positive, regardless of the distance between the actual DGP and the models in Q. For this reason, the average log likelihood ratio is used to measure the extent of model misspecification.[12] Motivated by these results, we often informally refer to an agent who uses such rule as a "statistically sophisticated type". Of course, the objective of a statistician can be very different from that of an agent involved in

[12]This use of the LLR complements its classical role in deciding whether to reject or accept a model. In particular, Wilks' Theorem (see, e.g., Theorem 10.3.3 in Casella and Berger, 2021) shows that *under correct specification,* the likelihood ratio test statistic converges to a χ^2 distribution. However, it says nothing about the distribution of the LLR if the model is misspecified. See Hausman (1978) and the subsequent literature for a complementary approach to the measurement of model misspecification when the statistician can compute a consistent quasi-maximum-likelihood estimator.

a decision problem under uncertainty. Proposition 1 confirms that this rule is also a rationality benchmark in repeated decision problems, as it uniquely identifies the behavior that induces no regret when the agent is correctly specified and is always maxmin safe.

1.3 Long-run Payoffs and Actions

In this section, we study the long-run consequences of using the decision criterion above. Our primary interest is in what attitudes towards unexplained evidence, i.e., what Λ, induce good payoff performance across environments and what are the limit actions and preferences under uncertainty attitudes that arise given a specific attitude.

Let $BR^{\lambda}(\nu)$ denote the set of average robust control best replies to belief ν when the concern for misspecification is λ, i.e.,[13]

$$BR^{\lambda}(\nu) = \underset{a\in A}{\operatorname{argmax}} \int_{Q} \min_{p_a\in\Delta(Y)} \left(\mathbb{E}_{p_a}\left[u(a,y)\right] + \frac{R(p_a||q_a)}{\lambda} \right) \mathrm{d}\nu(q) \qquad \forall \lambda \geq 0, \forall \nu \in \Delta(Q).$$

Also let

$$BR^{Seu}(\nu) = \underset{a\in A}{\operatorname{argmax}} \int_{Q} \mathbb{E}_{q_a}\left[u(a,y)\right] \mathrm{d}\nu(q) \qquad \forall \nu \in \Delta(Q)$$

denote the actions that maximize the (classical) subjective expected utility of an agent with belief ν and

$$BR^{Meu}(C) = \underset{a\in A}{\operatorname{argmax}} \inf_{p\in C} \mathbb{E}_{p_a}\left[u(a,y)\right]$$

denote the actions preferred by a maxmin agent a la Gilboa and Schmeidler (1989) with models $C \subseteq \Delta(Y)^A$.

A (pure) *policy* is a measurable $\Pi : \mathcal{H} \to A$ that specifies an action for every history. The objective action-contingent probability distribution and a policy Π induce a probability measure $\mathbb{P}_{\Pi}$ on $(A \times Y)^{\mathbb{N}}$.[14] Our interest is in policies derived from maximizing the value in equation (1.1) for some rule Λ determining how the concern

[13] Throughout the chapter, we use the convention $0 \cdot \infty = 0$.
[14] We spell out the $\mathbb{P}_{\Pi}$ derivation in Appendix .1.1.

for misspecification is adjusted.

Definition 1. Policy Π is Λ-*optimal* if for all $h_t \in \mathcal{H}$, $\Pi(h_t) \in BR^{\Lambda(h_t)}(\mu(\cdot|h_t))$.

1.3.1 Safety and Consistency

We have mentioned that using the rule in equation (1.2) has the good statistical property of keeping Λ asymptotically informative about the fit of the model. Now, we provide a normative justification for considering it the relevant benchmark of rationality, showing that it satisfies the desirable properties of safety and consistency (cf. Fudenberg and Levine, 1995) across all possible decision problems the agent can face.

Definition 2. Let $\varepsilon > 0$. Λ is ε-*safe* for the decision problem (u, A, Y) if for every Λ-optimal policy Π and DGP $p^* \in \Delta(Y)^A$

$$\liminf_{t\to\infty} \frac{\sum_{i=1}^{t} u(\mathbf{a}_i, \mathbf{y}_i)}{t} \geq \max_{a\in A} \min_{y\in Y} u(a, y) - \varepsilon \qquad \mathbb{P}_\Pi\text{-a.s.} \tag{1.3}$$

This is a very mild condition that only requires the agent to obtain an average payoff at least ε close to what they can guarantee against *every* possible outcome. However, when paired with misspecification, ε-safety has a significant bite: a Bayesian SEU agent fails it in many decision problems. Indeed, such failures have been the basis of many critiques of learning under misspecification with Bayesian SEU agents.

Example 1 (Unsafe SEU)**.** *Suppose* $A = \{$*Bet Heads, Bet Tails, Out*$\}$ *and* $Y = \{$*Heads, Tails*$\}$*. Utility is* 0 *if Out,* 1 *if action matches the outcome,* -1 *if mismatch. Each agent's model is an action-independent probability of Heads. So identify* $Q = \{0.9, 0.4\}$*, and let* $p_a^*(\textit{Heads}) = 0.6$*, and* $\mu(0.9) = \frac{1}{2} = \mu(0.4)$*. The actions of a Bayesian SEU maximizer converge to Bet Tails with average performance* -0.2 *versus a safe payoff of* 0 *under action Out. This is the simplest possible example, but safety also fail in the more economically motivated case of overconfidence, the key aspect being that good statistical fit does not necessarily induces good decisions.*

Definition 3. Let $\varepsilon > 0$. Λ is *ε-consistent under almost correct specification* for the decision problem (u, A, Y) if there exists $\delta > 0$ such that for every Λ-optimal policy Π and DGP $p^* \in \Delta(Y)$

$$\min_{q \in Q} \max_{a \in A} R\left(p_a^* || q_a\right) < \delta \implies \liminf_{t \to \infty} \frac{\sum_{i=1}^{t} u\left(\mathbf{a}_i, \mathbf{y}_i\right)}{t} \geq \max_{a \in A} \mathbb{E}_{p^*}\left[u\left(a, y\right)\right] - \varepsilon \qquad \mathbb{P}_{\Pi}\text{-a.s.}$$

ε-consistency under almost correct specification requires that sufficiently low levels of misspecification (i.e., the existence of a model q with distance less than δ from the true data generating process) cannot induce considerable ex-post regret (i.e., a limit average payoff more than ε lower than the expected payoff of the objectively optimal action). Intuitively, we want that if the misspecification is minor, in the long run, the agent approximately identifies the actual model and starts best replying to it.[15]

Proposition 1. *1. For every decision problem with $\{q^*\} = argmin_{q \in Q} Q(a)$ for all $a \in A$ and $\varepsilon > 0$ there exists $c > 0$ such that if*

$$\Lambda\left(h_t\right) = \frac{LLR\left(h_t, Q\right)}{ct} \qquad \forall t \in \mathbb{N}, \forall h_t \in \mathcal{H}_t,$$

Λ is both ε-safe and ε-consistent under correct specification.

2. There exists a decision problem with $\{q^*\} = argmin_{q \in Q} Q(a)$ and $\varepsilon > 0$ for which there is no ε-safe and ε-consistent under almost correct specification Λ with either

$$\Lambda\left(h_t\right) = o\left(\frac{LLR\left(h_t, Q\right)}{t}\right) \qquad \forall \left(h_t\right)_{t \in \mathbb{N}} \in \times_{t \in \mathbb{N}} \mathcal{H}_t.$$

or

$$o\left(\Lambda\left(h_t\right)\right) = \frac{LLR\left(h_t, Q\right)}{t} \qquad \forall \left(h_t\right)_{t \in \mathbb{N}} \in \times_{t \in \mathbb{N}} \mathcal{H}_t.$$

The safety and consistency conditions we require are weak but are enough to

[15] Recall that given two sequences $(x_n)_{n \in \mathbb{N}}, (x'_n)_{n \in \mathbb{N}}$ of real numbers, $x_n = o(x'_n)$ means that $\lim_{n \to \infty} \frac{x_n}{x'_n} = 0$. Here we use the convention that $\frac{0}{0} = 0$, making the definition of o more permissive. Since o will always appear as a requirement for a sequence in the hypothesis of our statements, such convention makes our results stronger and able to cover a larger range of cases.

single out the statistically sophisticated type. When the concern for misspecification is adjusted accordingly to equation (1.2), the combination of Bayesian updating over parameters and dynamically adjusted concern for misspecification is consistent with Savage's distinction between small and large worlds (see pages 82-91 in Savage, 1954). Indeed, Savage advocates reducing the large-world uncertainty to small worlds (for us, the structured Q) where Bayesian updating has appealing properties, but being aware that this description is incomplete and that the agent should evaluate the fit of that simplification (for us, using a test that can measure the failures of this description). In terms of performance, this result in a behavior that is safe and consistent under correct specification.

At the same time, some less normatively appealing but descriptively relevant phenomena are captured by other rules. On the one hand, a rule such that $\lim_{t\to\infty} \frac{\Lambda(h_t)}{\mathrm{LRT}(h_t,Q)/t} = \infty$, e.g., $\Lambda(h_t) = \frac{\mathrm{LRT}(h_t,Q)}{\sqrt{t}}$, overly penalizes minor imperfections of the model, expecting that the frequency quickly converges to its theoretical value, as in the fallacy called the Law of Small Numbers. On the other hand, an agent for which $\lim_{t\to\infty} \frac{\Lambda(h_t)}{\mathrm{LRT}(h_t,Q)/t} = 0$ applies an excessively lenient adjustment to the likelihood ratio statistic and attributes too much of the unexplained evidence to sampling variability. In this regard, Proposition 1 tells us that there are decision problems where any way to adjust the concern for misspecification that is globally more demanding or lenient than the average LRT violates either ε-safety or ε-consistency under almost correct specification. In contrast, standard SEU maximization is not safe, while always using a maxmin best reply to Q induces a behavior that is not consistent under almost correct specification.[16]

1.3.2 Long-run Behavior

We are interested in the actions that can arise as the long-run behavior of agents with an evolving concern for misspecification. The main results of this section show

[16]This observation about the inconsistency of a misspecified single agent complements the results of Fudenberg and Kreps (1993) and Fudenberg and Levine (1995) about the inconsistency of a correctly specified SEU player in games.

that we can describe this limit behavior through fixed point conditions involving the agent's action, belief, and concern for misspecification. To this end, let $Q(a) = \text{argmin}_{q \in Q} R(p_a^* || q_a)$ be the structured models that best fit the actual data-generating process when action a is played.

Definition 4. Action a^* is a:

1. *Self-confirming equilibrium* (SCE) if there exists $\nu \in \Delta(Q)$ with

$$\text{supp}\nu \subseteq \{q \in Q : q_{a^*} = p_{a^*}^*\} \text{ and } a^* \in BR^{Seu}(\nu).$$

2. *Berk-Nash equilibrium* (B-NE) if there exists $\nu \in \Delta(Q)$ with

$$\text{supp}\nu \subseteq Q(a^*) \text{ and } a^* \in BR^{Seu}(\nu).$$

3. *Maxmin equilibrium* if

$$a^* \in BR^{Meu}\left(\left\{p \in \Delta(Y)^A : \exists q \in Q, \forall a \in A, q_a \gg p_a\right\}\right).$$

4. *c-robust equilibrium* if there exists $\nu \in \Delta(Q)$ with

$$\text{supp}\nu \subseteq Q(a^*), \ a^* \in BR^{\lambda}(\nu), \text{ and } \lambda = \min_{q \in Q} R(p_{a^*}^* || q_{a^*}^q)/c.$$

Self-confirming equilibrium (Battigalli, 1987 and Fudenberg and Levine, 1993) describes a stable situation where the agent's action is a best reply to a belief that is on-path confirmed, in the sense of being concentrated over models that perfectly match the distribution over outcomes induced by the equilibrium action.

Berk-Nash equilibrium (Esponda and Pouzo, 2016) relaxes the confirmed beliefs condition of SCE by only requiring that the supporting beliefs are concentrated on the models that provide the best fit to the outcome distribution induced by the equilibrium action. Importantly, this fit is not required to be perfect.

In a maxmin equilibrium, the agent evaluates each action under the worst-case scenario that is minimally consistent with their structured descriptions of the environment (i.e., those scenarios that do not assign positive probability to events impossible for the structured models).

c-robust equilibrium is similar to Berk-Nash in requiring best reply to the best-fitting models. However, the best reply is the average robust control, with misspecification concern that decreases in how well the models fit the true DGP at the equilibrium.

We are interested in what actions have a positive probability of becoming the long-run behavior of the agent. The following definition captures this requirement.

Definition 5. Action a is a Λ-*limit action* if there is a Λ-optimal policy Π such that $\mathbb{P}_{\Pi}\left[\sup\{\mathbf{t}\colon \mathbf{a_t} \neq a\} < \infty\right] > 0$.

Our first limit result is a consistency check: Concern for misspecification is irrelevant in environments with a finite number of outcomes if the agent is correctly specified about the consequences induced by the limit action.

Proposition 2. *If Y is finite, a^* is a Λ-limit action with $p^*_{a^*} \in int\{q_{a^*}\}_{q\in Q}$, and for every history sequence $(h_t)_{t\in\mathbb{N}} \in \times_{t\in\mathbb{N}}\mathcal{H}_t$*

$$\lim_{t\to\infty} \frac{LLR\left(h_t, Q\right)}{t} = 0 \implies \lim_{t\to\infty} \Lambda\left(h_t\right) = 0$$

then a^* is a self-confirming equilibrium.

Instead, how quickly the agent becomes unsatisfied with their model plays a key role when misspecified.

Theorem 1. *Let a^* be a Λ-limit action with $p^*_{a^*} \notin \{q_{a^*}\}_{q\in Q}$. We have:*

1. *If*

$$\Lambda\left(h_t\right) = o\left(\frac{LLR\left(h_t, Q\right)}{t}\right) \qquad \forall\left(h_t\right)_{t\in\mathbb{N}} \in \times_{t\in\mathbb{N}}\mathcal{H}_t, \tag{1.4}$$

   then a^* is a Berk-Nash equilibrium.

2. *If*

$$o\left(\Lambda\left(h_t\right)\right) = \frac{LLR\left(h_t, Q\right)}{t} \qquad \forall \left(h_t\right)_{t\in\mathbb{N}} \in \times_{t\in\mathbb{N}}\mathcal{H}_t, \tag{1.5}$$

then a^* is a maxmin equilibrium.

3. *If*

$$\Lambda\left(h_t\right) = \frac{LLR\left(h_t, Q\right)}{ct} \qquad \forall t \in \mathbb{N}, \forall h_t \in \mathcal{H}_t,$$

then a^* is a c-robust equilibrium.

The theorem characterizes the possible limit actions of all types of agents. At one extreme, the concept of Berk-Nash equilibrium, introduced for subjective expected utility maximizers, is still sufficient to describe the long-run behavior of lenient types. At the other extreme, the repeated failures in explaining the observed data lead demanding agents to a highly pessimistic behavior and consider the worst-case scenario among all the DGPs that are minimally consistent with the structured models.

Finally, if the behavior of the statistically sophisticated type converges, the limit action a^* is a best reply to beliefs that are supported on the relative entropy minimizers. Here the misspecification concern is determined by the relative entropy between the actual DGP and the best-fitting model.

1.3.3 Equilibrium Illustrations

In this section, we revisit two of the main biases that have been justified as a consequence of misspecified learning (see Esponda and Pouzo, 2016). Within each example, adding an endogenous concern for misspecification predicts a change in a clear direction. However, one bias is reduced while the other is enhanced. Both changes are broadly consistent with the documented evidence. The first example shows how the endogenous misspecification concern moderates the Berk-Nash equilibrium's prediction that a more complicated tax schedule induces a higher labor supply.

Example 2 (Bias Reduction under Misperceived Taxation, Sobel, 1984 and Esponda and Pouzo, 2016)**.** *An agent chooses effort $a \in A$ at cost $c(a)$ and obtains income*

$z = a + \omega_a$, where ω_a is a stochastic term with $\mathbb{E}_{p_a^*}[\omega_a] = 0$ for all $a \in A$. The agent pays taxes $t = \tau(z) + l\varepsilon_1$, where $\tau : \mathbb{R} \to \mathbb{R}$ is a convex tax schedule. Here $y = (z, t)$, and the payoff is $u(a, y) = z - t - c(a)$. The agent believes in a random coefficient model, $t = (\theta + \varepsilon_2) z$, in which the marginal and average tax rates are both equal to $\theta + \varepsilon_2$ and $\Theta \subseteq \mathbb{R}$. The stochastic terms $\varepsilon_1, \varepsilon_2 \sim N(0, 1)$ measure respectively actual and conjectured uncertain aspects of the tax schedule, and the $(\omega_a)_{a \in A}, \varepsilon_1$, and ε_2 are independent.[17] *See Liebman and Zeckhauser (2004) and Rees-Jones and Taubinsky (2020) for the empirical evidence supporting this "schmeduling" bias.*

Simple computations show that $Q(a) \sim \left\{\mathbb{E}_{p_a^*}\left[\frac{\tau(a+\omega_a)}{a+\omega_a}\right]\right\}$ for l small, i.e., the best fitting marginal taxation is equal to the (lower) average taxation.[18] *Therefore, as pointed out by Esponda and Pouzo (2016), in any Berk-Nash equilibrium, the agent ends up exerting higher effort than the optimal. Moreover, the more complex (i.e., convex) the tax code is, the more significant the gap between the average and marginal rate and the higher the excess effort of the agent.*

In every c-robust equilibrium, this bias is reduced. To see this observe that since the agent is not perfectly able to explain the equilibrium data, i.e., $\min_{\theta \in \Theta} R\left(p_a^* || q_a^\theta\right) > 0$, they maintain a positive level of concern for misspecification. However, higher efforts are perceived as more exposed to the uncertainty in the marginal rate (as the stochastic term $\theta + \varepsilon$ gets multiplied by an, on average, higher z).

Therefore, c-robust equilibrium provides a natural force that reduces the counter-intuitive prediction that complicated nonlinear taxation codes induce more effort: failures to rationalize the received tax bill reduce effort. Moreover, the more complicated the tax code is, i.e., the more nonlinear τ is, the larger the correction size. This set of predictions is consistent with Rees-Jones and Taubinsky (2020), where it is shown that around 40% of the agents have beliefs *(elicited in an incentive-compatible way) corresponding with the schmeduling heuristic but that there are 20% fewer agents who* act *accordingly to the heuristic.*[19]

[17] Formally, ε normally distributed implies that Y is not compact, in contrast with the primary analysis of the chapter. Still, the conclusions below are unaffected by considering ε with a symmetrically truncated normal distribution that allows remaining in our main framework.

[18] See Appendix .1.4 for the computations supporting the claims of the examples.

[19] In this discussion we followed Rees-Jones and Taubinsky (2020) preferred interpretation in terms

The second example shows that an endogenous concern for misspecification can enhance some biases. In particular, this is the case for Correlation Neglect, a bias that is indeed widely documented (see Enke and Zimmermann, 2019 and the references therein).

Example 3 (Bias Increase under Correlation Neglect, Esponda, 2008)**.** *A buyer with valuation $v \in V$ and a seller submit a (bid) price $a \in A$, and an ask price $s \in S \subseteq \mathbb{R}_+$, respectively. They play a double auction with price at the buyer's bid, so the seller sets their ask s equal to their value, and a sale occurs if the buyer's bid a is at least s. The payoff for the buyer is*

$$u(a, v, s) = \begin{cases} v - a & a \geq s \\ 0 & \textit{otherwise.} \end{cases}$$

The buyer mistakenly believes that the ask price and valuation are independent: $Q = \Delta(V) \times \Delta(S)$. Easy computations show that for every $a^* \in A$,

$$Q(a^*) = \{q \in Q : \forall a \in A, \forall s \in S, q_a(s) = p^*_a(s), q_a(v) = p^*_a(v)\}.$$

Therefore, in the Berk-Nash equilibrium, the agent makes a bid a^* lower than the optimal one, not realizing that higher successful bids are, on average, associated with higher quality goods. In this case, the bias is reinforced in a c-robust equilibrium: a complete unraveling of the market where the buyer bids 0 is easier to achieve with an endogenous concern for misspecification. The correlation between valuations and prices results in a positive $\min_{q \in Q} R(p^*_{a^*} || q_{a^*}) > 0$ and makes the agent less confident in their model. Since offering 0 gives a certain payoff, it is less sensitive to the misspecification concern, and, therefore, this positive concern makes market participation less desirable.

of an heterogeneous population. They observe that their data are also compatible with all the agents having beliefs induced by the schmeduling heuristic but under-responding to this biased estimation of the marginal tax rate. This explanation is consistent with a c-robust equilibrium and inconsistent with a Berk-Nash equilibrium, too.

1.4 Cycles

Part 3 of Theorem 1 provides a necessary condition for the limit actions of the statistically sophisticated type. However, as momentarily illustrated by the monetary policy application of Section 1.4.1, there is no guarantee that such an action exists. In these cases, we know by Theorem 1 that the agent behavior cannot stabilize. We now propose a generalization of c-robust equilibrium, show that it always exists, and prove that it characterizes a weaker form of behavior convergence. Formally, for every $\alpha \in \Delta(A)$, let

$$Q(\alpha) = \text{argmin}_{q \in Q} \sum_{a \in A} \alpha(a) R(p_a^* || q_a)$$

be the set of parameters with the lowest average relative entropy from the actual data-generating process, where the average is computed using α.

Definition 6. A mixed action $\alpha^* \in \Delta(A)$ is a *mixed c-robust equilibrium* if there exists $\nu \in \Delta(Q)$ with

$$\text{supp}\nu \subseteq Q(\alpha^*), \ \alpha^* \in \Delta\left(BR^{\lambda}(\nu)\right), \text{ and } \lambda = \min_{q \in Q} \sum_{a \in A} \alpha^*(a) R(p_a^* || q_a) / c.$$

A mixed robust equilibrium allows multiple actions to be played but requires that the beliefs and the concern for misspecification are determined by the probability assigned to each action. Intuitively, suppose actions for which the models in Q do not satisfactorily explain the consequences are played more often. In that case, the mixed action α^* must best reply to a more significant misspecification concern.

Proposition 3. *For every $c > 0$ there exists a mixed c-robust equilibrium.*

Existence is established by proving that the conditions characterizing a mixed c-robust (single-agent) equilibrium are equivalent to the ones of a Nash equilibrium in a game among the agent and two adversarial Nature players. The result is then obtained by showing that this game satisfies the conditions that guarantee existence in Reny (1999).

Theorem 1 assumes convergence and characterizes the possible limit actions. However, there are natural environments where the action process almost surely does not converge. In that case, it is important to study a weaker form of behavior stabilization, i.e., the convergence of the empirical distribution over actions, that allows for persistent changes in actions and misspecification concerns. Let $\alpha_t(h_t) \in \Delta(A)$ be the empirical action frequency in history h_t, defined as

$$\alpha_t(h_t)(a) = \frac{\sum_{\tau=1}^{t} \mathbb{I}_{\{a\}}(a_\tau)}{t} \qquad \forall a \in A, \forall t \in \mathbb{N}, \forall h_t \in \mathcal{H}_t.$$

Definition 7. A mixed action $\alpha \in \Delta(A)$ is a Λ-*limit frequency* if there is a Λ-optimal policy Π such that $\mathbb{P}_\Pi[\lim_{t\to\infty} \alpha_t(\mathbf{h}_t) = \alpha] > 0$.

The following result shows that mixed robust equilibria are the relevant equilibrium concept to capture the long-run stabilization of the average time spent playing each action.

Theorem 2. *If*

$$\Lambda(h_t) = \frac{LLR(h_t, Q)}{ct} \qquad \forall t \in \mathbb{N}, \forall h_t \in \mathcal{H}_t$$

and α^* is a Λ-limit frequency, then α^* is a mixed c-robust equilibrium.

To interpret Theorem 2, consider the case where α^* is supported over two actions a, a' such that Q explains very well the consequences of a, —i.e., $\min_{q\in Q} R(p^*_a||q_a)$ is low— but it explains poorly the consequences of a' —i.e., $\min_{q\in Q} R(p^*_{a'}||q_{a'})$ is high. Suppose also that a is a best reply to a high misspecification concern, while a' is a best reply to a low misspecification concern. Then, the agent oscillates between periods with great concern for misspecification, when they play a, and phases in which the excellent data fit leads them to experiment with action a'.

Whenever cycles are involved, a natural concern is whether the agent can predict them and whether they have the incentive to break them.[20] This is not the case in this model for two orders of reasons. First, the oscillations in behavior are stochastic, and the agent cannot predict and anticipate the changes perfectly. Second and more

[20]For example, this is what happens under fictitious play.

important, although the agent behavior does not converge, whenever $Q(\alpha)$ is a singleton, the agent's preferences converge.[21] They are approaching indifference between all the actions with positive frequency. This asymptotic indifference dramatically reduces the incentives to try to detect the probabilistic cycle and break them.

1.4.1 Application: Monetary Policy Cycles

Here we consider a monetary policy model taken from Sargent (1999), Cogley and Sargent (2005), and Sargent (2008) and in particular its adaptation in Battigalli, Cerreia-Vioglio, Maccheroni, Marinacci, and Sargent (2022).[22] A central bank is trying to control a two-dimensional consequence, $Y \subseteq \mathbb{R}^2$, where the y_U component is unemployment and the y_π component is inflation. The policy is aggressive $a = 1$ or conservative $a = 0$.[23] The central bank models are parametrized by the vector θ, with the following specification:

$$\begin{aligned} y_U &= \theta_0 + \theta_{1\pi} y_\pi + \theta_{1a} a + \theta_2 \varepsilon_U \\ y_\pi &= a + \theta_3 \varepsilon_\pi \end{aligned}$$

where ε_U and ε_π are independent, zero-mean random shocks normalized to have the same support $[-1, 1]$. Here $\theta_0 > 0$ is the natural unemployment level, $\theta_{1\pi} < 0$ is the impact of the actual inflation on unemployment, and $\theta_{1a} > 0$ is the impact of the planned inflation on unemployment, a reduced form of the fact that the market participants (partially) incorporate the central bank actions in their inflation expectations. In particular, if $\theta_{1\pi} + \theta_{1a} = 0$, this is a Lucas-Sargent model with no (structural) exploitable employment-inflation trade-off. If $\theta_{1\pi} + \theta_{1a}$ is negative, this is a Samuelson-Solow model with a structural exploitable employment-inflation trade-off.

[21] A singleton $Q(\alpha)$ is a mild requirement satisfied in many cases. See Fudenberg, Lanzani, and Strack (2021) for a discussion.

[22] Spiegler (2020) also considers the effect of a misspecified simpler model in the context of Philipps curve estimation, with the difference being that the misspecification is on the side of the market rather than the central bank.

[23] Two actions are assumed for simplicity, but a finite A is needed to apply our results.

The agent's model is misspecified in that it misses the fact that an aggressive monetary policy, beyond raising its baseline level, also increases the inflation variability:

$$\begin{aligned} y_U &= \theta_0^* + \theta_{1\pi}^* y_\pi + \theta_{1a}^* a + \theta_2^* \varepsilon_U \\ y_\pi &= a + \theta_3^* f_a(\varepsilon_\pi) \end{aligned}$$

where f_0 is the identity function, while f_1 is a continuous, strictly increasing, and odd function with $f_1(1) = 1$ that is strictly concave on $\mathbb{R}_{++}$, i.e., that amplifies the inflation-specific shocks.[24] This form of misspecification is motivated by the findings in Primiceri (2005) and Sims and Zha (2006) and recent inflation consequences of an aggressive monetary policy.

The central bank is endowed with standard quadratic preferences:

$$u(a, (y_U, y_\pi)) = -y_U^2 - y_\pi^2.$$

Assumption 3. i) Some trade-off is present: $(\theta_{1\pi}^* + \theta_{1a}^* + \theta_0^*)^2 + 1 < (\theta_0^*)^2$. ii) Inflation is more volatile than unemployment under the aggressive monetary policy: $\operatorname{essinf}_{p_1^*} u(1, y) < \operatorname{essinf}_{p_0^*} u(0, y)$. iii) Θ is a product set that includes θ^* and for all $\theta \in \Theta$, $(\theta_{1a}, \theta_2, \theta_3) = (\theta_{1a}^*, \theta_2^*, \theta_3^*)$.

Observe that the exploitable trade-off required by (i) may be so small that the reduced inflation variability under a conservative policy makes the latter optimal. Condition (ii) requires that the additional inflation volatility induced by the aggressive policy is enough to have the worst tail payoffs. Condition (iii) allows us to focus on the cycles induced by the oscillation in the concern for misspecification. Without that, one would get the same insights with other oscillations of beliefs that push even

[24] An alternative, more parametric specification would have

$$\begin{aligned} y_U &= \theta_0^* + \theta_{1\pi}^* y_\pi + \theta_{1a}^* a + \theta_2^* \varepsilon_U \\ y_\pi &= \left(1 + \sigma_{\pi a}^2 \varepsilon_{\pi a}\right) a + \theta_3^* \varepsilon_\pi \end{aligned}$$

where $\varepsilon_{\pi a}$ is an independent error and $\sigma_{\pi a}^2 > 0$. If we let the support of ε_U and ε_π be unbounded, nothing in the analysis below would be affected by a shift to this alternative specification. However, that change would bring us outside the compact Y setting study in the rest of the chapter, so we opted for preserving the consistency.

more towards cycles, a channel pointed out by Nyarko (1991) in a monopoly pricing setting.

Corollary 1. *There is $\bar{c} > 0$ such that for all $c \leq \bar{c}$*

1. *There is no c-robust equilibrium.*

2. *There exists a mixed c-robust equilibrium.*

3. *The maximal and minimal equilibria are such that $\alpha^*(0)$ is increasing in $\theta^*_{1\pi} + \theta^*_{1a}$.*

Playing the conservative policy is the best reply to a high misspecification concern and θ^* but induces a low concern as its consequences are well explained. In contrast, the aggressive policy is a best reply to a low misspecification concern and θ^* but induces a severe concern. Therefore, the policy cannot stabilize, consistently with the cyclical behavior of monetary policies documented in Sargent (1999), Clarida, Gali, and Gertler (2000), and Sargent (2008). We also have some natural comparative statics in the extremal robust equilibria, as a more significant exploitable trade-off between inflation and unemployment induces more time spent using an aggressive monetary policy.[25]

In this application, we purposefully chose one of the most straightforward macroeconomic frameworks to isolate and illustrate the effect of an endogenous concern for misspecification. However, incorporating an endogenous concern for misspecification in more elaborate models is a valuable enterprise. For example, the fact that evidence impacts the trust in the model may be used to explain the observed pattern of initial underreaction to information when only beliefs within a model are adjusted and medium-run overreaction when the belief adjustment compounds with a change in model trust (see Angeletos, Huo, and Sastry, 2021 and the references therein for a discussion of this pattern).

[25] It is well-known that non-extremal equilibria are less well-behaved in terms of comparative statics. See Diamond (1982) for a very early example. A supermodularity condition between the concern for misspecification and the conservative policy payoff guarantees equilibrium uniqueness.

1.5 Representation

We next move to characterize the average robust control model in terms of observable choices in an Anscombe-Aumann framework. In line with the literature on decision theory under uncertainty, our goal is to associate the decision criterion in equation (1.1) with axioms on a binary preference relation over acts.

Before jumping into the details of the axiomatization, we provide a high-level description of the steps involved and the intuitive meaning of the axioms we link to the representation. In terms of observability requirements, we allow the analyst to elicit preferences for bets both on the data-generating process, e.g., the urn composition, and on the actual realization, e.g., the color of the drawn ball.[26] The analysis then has two nested levels: 1) An axiomatization of the static decision criterion, 2) An axiomatization of the changes of the preference parameters, and in particular the speed of adjustment of the concern for misspecification.

The static decision criterion belongs to the variational class of Maccheroni, Marinacci, and Rustichini (2006a). More importantly, within this class, it is identified by a relaxed Sure-Thing Principle: the agent satisfies it for bets that involve the identity of the model (e.g., bets on an urn composition) and for bets on events conditional on the model (e.g., bets on the color after having revealed the urn composition). However, failures of the Sure-Thing principle can realize for acts that involve the realization of the outcome without conditioning on the model (e.g., bets on the color without knowing the urn composition, which are the ones involved in the classical Ellsberg's paradox). The final conceptual axiom involved in the representation of equation (1.1) is a notion of uniform conditional misspecification concern. It requires that conditional on being told the identity of their best-fitting model, the agent is equally concerned about it not being exact regardless of which one it is.

We consider a collection of binary relations indexed by the observed history to characterize the agent's dynamic preferences. Three other axioms identify the qual-

[26]This is standard when dealing with multiple sources of uncertainty, see for example Klibanoff, Marinacci, and Mukerji (2005) and Gul and Pesendorfer (2014), and Cerreia-Vioglio, Maccheroni, Marinacci, and Montrucchio (2013a) for a general framework and results. We discuss how to relax this requirement in Section 1.6.4.

itative changes of the preference parameters u, λ, μ. Constant Preference Invariance guarantees that the taste u for uncertain alternatives is stable over time. Dynamic Consistency over Models guarantees that the probability distribution over models is updated in a Bayesian fashion. Finally, we axiomatize the asymptotic speed of adjustment of the misspecification concern. To do so, we need a quantitative notion of how similar two preference relations are, which is defined using an event E and two deterministic and strictly ranked outcomes, x and y, as measuring rods. Loosely speaking, two relations are (x, y, E, ε) similar if their certain equivalents for the binary act that pays x if E realizes and y otherwise are ε close. With this, an Asymptotic Frequentism axiom singles out the statistically sophisticated type: for every (x, y, E, ε), the conditional preferences after sufficiently long sequences of outcomes sharing the same empirical frequency must be (x, y, E, ε)-similar. Conversely, a lenient type asymptotically becomes similar to those SEU preferences that are less misspecification concerned than the initial preference. The demanding type must approach the preferences of a maxmin agent, thus confirming in a decision-theoretic setting the insights of Theorem 1.

1.5.1 Notation and Preliminaries

The agent evaluates simple acts, i.e., measurable and finite ranged maps from a nonempty state space S into a convex set of outcomes X, where S is endowed with a σ-algebra of events Σ. The set of those acts is denoted as $\mathcal{F}$. Given any $x \in X$, $x \in \mathcal{F}$ is the act that delivers x in every state, and in this way, we identify X as the subset of constant acts in $\mathcal{F}$. If $f, g \in \mathcal{F}$, and $E \in \Sigma$, we denote as gEf the simple act that yields $g(s)$ if $s \in E$ and $f(s)$ if $s \notin E$. Since X is convex, for every $f, g \in \mathcal{F}$, and $\gamma \in (0, 1)$, we denote as $\gamma f + (1-\gamma) g \in \mathcal{F}$ the simple act that pays $\gamma f(s) + (1-\gamma) g(s)$ for all $s \in S$.

We model the agent's preference with a binary relation $\succsim$ on $\mathcal{F}$. As usual $\succ$ and $\sim$ denote the asymmetric and symmetric parts of $\succsim$. An event E is *null* if $fEh \sim gEh$ for every $f, g, h \in \mathcal{F}$. An event is *nonnull* if it is not null. For every $E \in \Sigma$, the conditional preference relation $\succsim_E$ is defined by $f \succsim_E g$ if $fEh \succsim gEh$ for some

$h \in \mathcal{F}$.

A key concept to understand the concern for misspecification evolution is a notion of being more misspecification concerned from Ghirardato and Marinacci (2002).

Definition 8. Given two preferences $\succsim_1$ and $\succsim_2$ on $\mathcal{F}$, we say that $\succsim_1$ is more concerned with misspecification than $\succsim_2$ if, for each $f \in \mathcal{F}$ and each $x \in X$, $f \succsim_1 x$ implies $f \succsim_2 x$.

1.5.2 Decision Criterion

When formalized in terms of a binary relation, the average robust control decision criterion reads as follows.

Definition 9. A tuple (u, Q, μ, λ) is an *average robust control representation* of the preference relation $\succsim$ if $u : X \to \mathbb{R}$ is a nonconstant affine function, $Q \subseteq \Delta(S)$ is a nonempty set, $\mu \in \Delta(Q)$, $\lambda \geq 0$, and for all $f, g \in \mathcal{F}$

$$f \succsim g \iff \mathbb{E}_\mu \left[\min_{p \in \Delta(S)} \left(\int_S u(f)\, \mathrm{d}p + \frac{R(p||q)}{\lambda} \right) \right] \geq \mathbb{E}_\mu \left[\min_{p \in \Delta(S)} \left(\int_S u(g)\, \mathrm{d}p + \frac{R(p||q)}{\lambda} \right) \right]. \tag{1.6}$$

The average robust control representation is the counterpart of (1.1) when expressed over acts. An apparent difference is that u here takes as input only outcomes instead of pair of actions and consequences. However, this discrepancy is inconsequential, as in Section 1.2 we can define a larger space of consequences $\hat{Y} = A \times Y$ that includes both actions and outcomes and transforming each model $p \in \Delta(Y)^A$ into an element of $\hat{p} \in \Delta\left(\hat{Y}\right)^A$ such that $\hat{p}_a(a', y) = 0$ if $a' \neq a$ and $\hat{p}_a(a, y) = p_a(y)$ for all $y \in Y$. Still, this embedding of actions into outcomes muddles the interpretation of the learning results significantly. Therefore we opted to maintain the distinction explicit at the cost of some visual discrepancy between equations (1.1) and (1.6).[27]

[27] See Fishburn (1970) Chapter 12.1 for a more detailed discussion of the equivalence of a formulation with exogenously given states and one where states are maps from actions into consequences.

1.5.3 Static Axioms

Our first axiomatic step is a static one. We characterize in terms of behavioral axioms an agent that evaluates accordingly to equation (1.6) the acts whose consequences are obtained in the same period and before any new information is received.

Axiom 1 (Variational Axiom)**.** ***Weak Order.***

Weak Certainty Independence. *If $f, g \in \mathcal{F}$, $x, x' \in X$, $\gamma \in (0,1)$, and $\gamma f + (1-\gamma) x \succsim \gamma g + (1-\gamma) x$, then $\gamma f + (1-\gamma) x' \succsim \gamma g + (1-\gamma) x'$.*

Continuity. *If $f, g, h \in \mathcal{F}$ the sets $\{\gamma \in [0,1] : \gamma f + (1-\gamma) g \succsim h\}$ and $\{\gamma \in [0,1] : h \succsim \gamma f + (1-\gamma) g\}$ are closed.*

Monotonicity. *If $f, g \in \mathcal{F}$, and $f(s) \succsim g(s)$ for all $s \in S$, then $f \succsim g$.*

Uncertainty Aversion. *If $f, g \in \mathcal{F}$, $\gamma \in (0,1)$, and $f \sim g$, then $g + \gamma(f-g) \succsim f$.*

Nondegeneracy. *$f \succ g$ for some $f, g \in \mathcal{F}$.*

Weak Monotone Continuity. *If $f, g \in \mathcal{F}$, $x \in X$, $(E_n)_{n \in \mathbb{N}} \in \Sigma^{\mathbb{N}}$ with $f \succ g$, $E_1 \supseteq E_2 \supseteq ...$ and $\cap_{n \in \mathbb{N}} E_n = \emptyset$, then there exists $n_0 \in \mathbb{N}$ such that $x E_{n_0} f \succ g$.*

Maccheroni, Marinacci, and Rustichini (2006a) shows that Axiom 1 characterizes the class of variational preferences. Weak Order, Continuity, and Nondegeneracy are standard technical requirements. Weak Monotone Continuity guarantees that the probabilistic scenarios considered by the agent are countably additive. Weak Certainty Independence allows the agent to perceive some advantage in hedging, but this cannot come from mixing with different constants using the same weights. Monotonicity requires that the preference over acts is minimally consistent with the preference over the outcomes they induce. Uncertainty Aversion leads to aversion for the acts that perform well for a postulated model but poorly for its perturbations.

Structured Preferences

We are considering agents who face two levels of uncertainty: the uncertainty on the best structured description of the data-generating process and whether each description is exact. A representation is structured if it allows separating these two layers.

In particular, to achieve this separation, we consider a state space S that admits the decomposition $S = \Omega \times \Delta(\Omega)$ for some finite Ω endowed with its Borel sigma-algebra.

Definition 10. An average robust control representation (u, Q, μ, λ) is *structured* if μ has finite support and there exists a map

$$\begin{aligned} Q &\to \Delta(\Omega) \\ q &\mapsto \rho_q \end{aligned}$$

such that for every $q \in Q$ and $\omega \in \Omega$, $q(\{\omega, \rho_q\}) = \rho_q(\omega)$.

The interpretation of a structured representation is that the state space can be factored in two components, the realization of the single period consequence $\omega \in \Omega$ and a component $\rho \in \Delta(\Omega)$ that pins down the distribution over states each period. An event E is *structured* if $E = \Omega \times B$ for some $B \in \mathcal{B}(\Delta(\Omega))$. The sigma-algebra generated by the structured events is denoted as Σ_s.[28]

We say that an event $E \subseteq S$ satisfies the sure-thing principle if, for all $f, g, h, h' \in \mathcal{F}$ we have that $fEh \succsim gEh$ implies $fEh' \succsim gEh'$. We denote by Σ_{st} the set of events that satisfy the sure-thing principle.

Axiom 2 (Structured Savage)**.** *i) There is a finite set $E \subseteq S$ such that $S \setminus E$ is null. ii)* ***P2.*** *$\Sigma_s \subseteq \Sigma_{st}$. iii)* ***P4.*** *If $E, E' \in \Sigma_s$ and $x, y, w, z \in X$ are such that $x \succ y$ and $w \succ z$, then*

$$xEy \succ xE'y \Rightarrow wEz \succ wE'z.$$

Structured Savage requires that (i) the agent posits a finite number of models and (ii) guarantees that when evaluating acts that only depend on the identity of the structured model, the agent satisfies the Sure-Thing Principle.[29] It also (iii) guarantees that when an agent faces alternatives whose outcomes depend only on

[28] With a slight abuse of notation for every $B \in \mathcal{B}(\Delta(\Omega))$ and $W \subseteq \Omega$ we denote as $\succsim_B$ and $\succsim_W$ the binary relations $\succsim_{\Omega \times B}$ and $\succsim_{W \times \Delta(\Omega)}$ and we write fBg and fWg for $f(\Omega \times B)g$ and $f(W \times \Delta(\Omega))g$.

[29] The extension to infinitely many models does not provide additional conceptual difficulties but makes the conditioning involved in the dynamic axioms much more cumbersome. Gul and Pesendorfer (2014) introduces the idea of sources of uncertainty for which the decision maker can quantify uncertainty and connects it with the Sure-Thing Principle.

whether the DGP belongs to two sets of models, their choices consistently reveal the one deemed more likely.

Axiom 3 (Intramodel Sure-Thing Principle)**.** *For every* $f, g, h, h' \in \mathcal{F}$*,*

$$fWh \succsim_\rho gWh \implies fWh' \succsim_\rho gWh' \qquad \forall W \subseteq \Omega, \forall \rho \in \Delta(\Omega).$$

Structured Savage's P2 and the Intramodel STP imply that bets *between* models and preference over acts *within* a model satisfy the STP. However, they admit violations of the STP for acts whose payoff depends on both the model's identity and the outcome realization within the model, as the ones of the original Ellsberg's paradox.

The case we study is when the relative likelihood of the structured models is only captured by the belief μ. In particular, the agent is equally concerned about how much each model departs from the actual data-generating process.

Axiom 4 (Uniform Misspecification Concern)**.** *For every* $\rho, \rho' \in \Delta(\Omega)$ *and* $f, g \in \mathcal{F}$ *such that*

$$\rho(\{\omega : f(\omega, \rho) = y\}) = \rho'(\{\omega : g(\omega, \rho') = y\}) \quad \forall y \in X$$

and $\Omega \times \{\rho\}$*,* $\Omega \times \{\rho'\}$ *are nonnull we have*

$$f \succsim_\rho x \iff g \succsim_{\rho'} x \qquad \forall x \in X.$$

This axiom requires that if acts f and g induce identical outcome distributions under ρ and ρ', they are compared with a safe alternative in the same way conditional on the best fitting model being revealed to be ρ or ρ'.

Definition 11. The state space is adequate if: (i) there exist $k \in (0,1)$ and $(W_\rho)_{\rho \in \Delta(\rho)} \in \left(2^\Omega\right)^{\Delta(\rho)}$ such that for all $\rho \in \Delta(\Omega)$ such that $\Omega \times \{\rho\}$ is nonnull, $\rho(W_\rho) = k$, (ii) for every $\omega, \omega' \in \Omega$, and $\rho \in \Delta(\Omega)$ such that $\{\omega\} \times \{\rho\}$ and $\{\omega'\} \times \{\rho\}$ are nonnull, $\rho(\omega) = \rho(\omega')$.

All the agent's structured models have an event with the same probability and are uniform over a model-specific set of outcomes. It is well-known that equal probability

requirements are essential for probabilistic sophistication with respect to a finite measure over states to have a bite (see, e.g., Chew and Sagi, 2006). They can be relaxed if we allow for a continuum Ω. The only role of (i) for us is to obtain a concern for misspecification λ that is not model dependent (i.e., not to have $(\lambda_q)_{q\in Q}$ in the representation) from Uniform Misspecification Concern, an axiom with an unmistakable flavor of probabilistic sophistication.

Axiom 5 (Uncertainty Neutrality Over Models)**.** *Let $x, y, w, z \in X$, $\rho \in \Delta(\Omega)$, and $\gamma \in (0,1)$. Then $[\gamma x + (1-\gamma) y]_\rho w \sim y_\rho z$ if and only if $x_\rho w \sim [(1-\gamma) x + \gamma y]_\rho z$.*

Uncertainty Neutrality over Models guarantees that at the level of bets over models, the agent is "risk-neutral", as changing the performance under ρ by $(x-y)\gamma$ has an impact that does not depend on the level of utility under that model. It is immediate from the proof of Theorem 3 that if dropped, it leads to a more general representation with a nonlinear utility index U over the performance of each robust control model.

Theorem 3. *Suppose that S is adequate, there at least three disjoint nonnull events in Σ_s, and every nonnull $E \in \Sigma_s$ contains at least three disjoint nonnull events. The following are equivalent:*

1. *$\succsim$ admits a structured average robust control representation (u, Q, μ, λ);*
2. *$\succsim$ satisfies Variational Axiom, Structured Savage, Uniform Misspecification Concern, Intramodel Sure-Thing Principle, and Uncertainty Neutrality over Models.*

Moreover, in this case, every two structured average robust control representations share the same μ.

The theorem characterizes the representation (u, Q, μ, λ) with probabilistic uncertainty *about* the model (Structured Savage), probabilistic sophistication *given* a model (Intramodel Sure-Thing Principle), and *incomplete trust* in any model (Uncertainty Aversion).

Corollary 2. *Suppose that $\succsim$ admits a structured average robust control representation (u, Q, μ, λ). Then $\succsim$ is more misspecification concerned than the subjective utility preference with utility index u and belief $\int_Q q d\mu(q)$.*

1.5.4 Dynamic Axioms

We next provide axioms that characterize the dynamic adjustment of preferences in the face of information. In particular, we look at joint axioms on a collection of history-dependent binary relations $\left(\succsim^h\right)_{h\in\mathcal{H}}$ indexed by the realized history. Recall that the relevant set of length $t \in \mathbb{N}$ histories for structured preferences is Ω^t.

Axiom 6 (Constant Preference Invariance)**.** *For every $x, x' \in X$ and $h \in \mathcal{H}$,*

$$x \succsim^h x' \Leftrightarrow x \succsim^{\emptyset} x'.$$

This axiom captures the fact that we are not considering the problem of an agent discovering their taste. The preferences over uncertain alternatives are fixed and do not react to new information.

Axiom 7 (Dynamic Consistency over Models)**.** *Let $f, g \in \mathcal{F}$ be Σ_s-measurable, $t \in \mathbb{N}$, $(\omega_1, ..., \omega_t) \in \Omega^t$ and $\bar{z}, \underline{z} \in X$ be such that $\bar{z} \succsim f(s) \succsim \underline{z}$ and $\bar{z} \succsim g(s) \succsim \underline{z}$ for all $s \in S$. Define h^0 as*

$$h^0(\omega, \rho) = \gamma_{h(\omega,\rho)} \prod_{i=1}^{t} \rho(\omega_i)\, \bar{z} + \left(1 - \gamma_{h(\omega,\rho)} \prod_{i=1}^{t} \rho(\omega_i)\right) \underline{z} \qquad \forall (\omega, \rho) \in S, \forall h \in \{f, g\}$$

where $\gamma_{h(\omega,\rho)}$ satisfies $h(\omega, \rho) \sim \bar{z}\gamma_{h(\omega,\rho)} + \left(1 - \gamma_{h(\omega,\rho)}\right)\underline{z}$. Then, we have

$$f \succsim^{(\omega_1,...,\omega_t)} g \Longleftrightarrow f^0 \succsim g^0.$$

The second dynamic axiom requires Bayesian rationality when considering acts whose consequences only depend on the structured model. Formally, it requires that when comparing acts that only bet on the identity of the model, at a given history, we can reduce the comparison to acts evaluated ex-ante. To do so, the payoff conditional

to each model must be scaled proportionally to the amount of evidence that has been generated in favor of that model.[30]

To single out the *quantitative* speed at which the concern for misspecification is adjusted, we need a quantitative measure of similarity. For every $x, y \in X$ with $x \succ y$ and $E \in \Sigma$ let $\gamma_{\succsim}^{xEy}$ be defined by

$$\gamma_{\succsim}^{xEy} x + \left(1 - \gamma_{\succsim}^{xEy}\right) y \sim xEy.$$

That is, $\gamma_{\succsim}^{xEy}$ is the weight to alternative x in the certain equivalent to act xEy. It captures both the confidence in event E and the attitudes towards uncertainty. It is easy to see that under the Variational Axiom $\gamma_{\succsim}^{xEy}$ always exists and is unique.

For every $x, y \in X$, $E \in \Sigma$, $\varepsilon \in (0,1)$, and $\succsim$ and $\succsim'$ that satisfy the Variational Axiom, we say that $\succsim$ is (x, y, E, ε)-*similar* to $\succsim'$ if

$$\left|\gamma_{\succsim}^{xEy} - \gamma_{\succsim'}^{xEy}\right| \leq \varepsilon.$$

That is, the certain equivalent of the binary act xEy is ε close under preferences $\succsim$ and $\succsim'$.

Axiom 8 (Asymptotic Frequentism)**.** *For every $\rho \in \Delta(\Omega)$, $x, y \in X$ with $x \succ^{\emptyset} y$, $\varepsilon \in (0,1)$, and $E \in \Sigma$ there is $\tau \in \mathbb{N}$ such that if $t, t' \geq \tau$ and $h_t, h_{t'}$ have outcome frequency ρ then $\succsim^{h_t}$ is (x, y, E, ε)-similar to $\succsim^{h_{t'}}$.*

The axiom requires that for every binary act xEy, a sufficiently long sequence of outcomes with the same empirical frequency stabilize the certain equivalent.

Proposition 4. *Let $\left(\succsim^h\right)_{h \in \mathcal{H}}$ be such that*

1. *For every $h \in \mathcal{H}$, $\succsim^h$ satisfies the axioms of Theorem 3,*

[30] This axiom can lead to fruitful implications beyond our average robust control decision criterion, as it implies Bayesian updating for each decision criteria that performs an average of model-specific evaluations (that could, for example, take the form of other divergence preferences or rank-dependent utility evaluations). In this way, it would complement the elegant theory of subjective learning developed in Dillenberger, Lleras, Sadowski, and Takeoka (2014), which does not require that the analyst observes the same information as the agent.

2. $\left(\succsim^h\right)_{h\in\mathcal{H}}$ *satisfies Constant Preference Invariance, Dynamic Consistency over Models, and Asymptotic Frequentism.*

Then for every sequence $(h_{t_n})_{n\in\mathbb{N}}$ *with a constant outcome frequency not in* $\{\rho_q : q \in Q\}$,

$$\lim_{n\to\infty} \lambda_{h_{t_n}} / \left(\frac{LLR\left(h_{t_n}, Q\right)}{t_n}\right) \tag{1.7}$$

exists. Moreover, if for some $q \in Q$, $x \succ^{\emptyset} y$, *and* $E \subseteq \Omega$ *with* $\rho_q(E) > 0$

$$\liminf_{n\to\infty} \gamma_{\succsim^{h_{t_n}}}^{x(E\times\{\rho_q\})y} > 0,$$

the limit is finite, and if

$$\limsup_{n\to\infty} \gamma_{\succsim^{h_{t_n}}}^{x(E\times\{\rho_q\})y} < \rho_q(E)\,\mu(q),$$

it is strictly positive.

The proof of the result has two main steps. First, we show that the likelihood ratio statistic of the models Q is growing linearly in t_n along the sequence of histories $(h_{t_n})_{n\in\mathbb{N}}$, so that the denominator in equation (1.7) converges.[31] Because the outcome frequency does not correspond to a model in Q, this limit is not 0. With this, the proof amounts to showing that the revealed concern for misspecification also converges. The second step rules out the existence of different finite limit points for $(\lambda_{t_n})_{n\in\mathbb{N}}$ by contradiction. If these points exist, then for every pair of strictly ranked outcomes $x \succ y$, we construct an event E for which the DM does not satisfy the Sure-Thing Principle such that the preference with the high concern for misspecification has a strictly higher certain equivalent than the one with the low concern.

Axiom 9 (Asymptotic Concern)**.** *Let* $f \in \mathcal{F}$, $x \in X$, *and* $\hat{\rho}, \rho \in \Delta(\Omega)$ *be such that* $\Omega\times\{\hat{\rho}\}$ *is* $\succsim^{\emptyset}$*-null,* $\Omega\times\{\rho\}$ *is* $\succsim^{\emptyset}$*-nonnull, and* $\rho \gg \hat{\rho}$. *If* $\rho(\{\omega \in \Omega : x \succ f(\omega, \rho)\}) >$

[31] Given the finiteness of Ω, we can focus on the case in which the alternative set of models $N(Q) = \Delta(\Omega)$, discussed in Section 1.2, i.e., $LLR\left(h_t, Q\right) = -\log\left(\frac{\max_{q\in Q}\prod_{\tau=1}^{t}\rho_q(\omega_\tau)}{\max_{\rho\in\Delta(\Omega)}\prod_{\tau=1}^{t}\rho(\omega_\tau)}\right)$. The extension to general sets of alternative models is straightforward.

0, *then there exists* $\tau \in \mathbb{N}$ *such that for all* $t \geq \tau$ *and all* h_t *with outcome frequency* $\hat{\rho}$, $x \succsim_{\rho}^{h_t} f$.

Asymptotic Concern requires that long-run failures in explaining the data (i.e., an empirical frequency $\hat{\rho}$ that is not among the agent's structured models) increase the concern so that every certain outcome is preferable to an act with worse payoffs under a relevant model ρ.

Proposition 5. *Let* $\left(\succsim^h\right)_{h\in\mathcal{H}}$ *be such that*

1. *For every* $h \in \mathcal{H}$, $\succsim^h$ *satisfies the axioms of Theorem 3,*

2. $\left(\succsim^h\right)_{h\in\mathcal{H}}$ *satisfies Constant Preference Invariance, Dynamic Consistency over Models, and Asymptotic Concern.*

Then for every sequence $(h_{t_n})_{n\in\mathbb{N}}$ *with a constant outcome frequency that is not in* $\{\rho_q : q \in Q\}$ *we have*

$$\lim_{t\to\infty} \frac{LLR\left(h_{t_n}, Q\right)}{\lambda_{h_{t_n}} t_n} = 0.$$

This result shows that Asymptotic Concern characterizes agents who apply an excessively demanding time discount to the likelihood ratio test statistic (see equation (1.2)). Indeed, the elicited ratio between the LRT and the concern for misspecification revealed by the choices grows sublinearly time, the condition that defines demanding agents.

Axiom 10 (Asymptotic Leniency)**.** *Let* $x, y \in X$, $E \in \Sigma$, $\rho \in \Delta(\Omega)$, $\varepsilon \in (0,1)$, *be such that* $x \succ y$. *For every Bayesian SEU preferences* $\left(\succeq^h\right)_{h\in\mathcal{H}}$ *such that* $\succeq^{\emptyset}$ *is less misspecification averse than* $\succsim^{\emptyset}$, *there exists* $\tau \in \mathbb{N}$ *such that for every* $t \geq \tau$ *and* h_t *with outcome frequency* ρ, $\succsim^{h_t}$ *is* (x, y, E, ε)*-similar to* $\succeq^{h_t}$.

Asymptotic Leniency requires that if the empirical distribution converges to some $\rho \in \Delta(\Omega)$, the preferences of the agents approximate, i.e., are eventually (x, y, E, ε)-similar to the updated preferences of an SEU whose model contingent preferences were initially less misspecification averse than the agent.

Proposition 6. *Let* $\left(\succsim^h\right)_{h\in\mathcal{H}}$ *be such that*

1. *For every* $h \in \mathcal{H}$, $\succsim^h$ *satisfies the axioms of Theorem 3,*

2. $\left(\succsim^h\right)_{h\in\mathcal{H}}$ *satisfies Constant Preference Invariance, Dynamic Consistency over Models, and Asymptotic Leniency.*

Then for every sequence $(h_{t_n})_{n\in\mathbb{N}}$ *with constant outcome frequency not in* $\{\rho_q : q \in Q\}$

$$\lim_{t\to\infty} \frac{LLR\left(h_{t_n}, Q\right)}{t_n \lambda_{h_{t_n}}} = \infty.$$

This proposition shows that convergence to subjective expected utility maximization (in the form of Asymptotic Leniency) characterizes excessively lenient time normalizations.

1.6 Discussion

1.6.1 Related Literature

A few papers allow the agents to realize that they are misspecified. In particular, in He and Libgober (2022), Ba (2022), Fudenberg and Lanzani (2022), and Gagnon-Bartsch, Rabin, and Schwartzstein (2022) misspecification can be eliminated either by "light bulb realizations" or evolutionary pressure. The key difference with our approach is that in these papers, as well in the earlier Foster and Young (2003), Cho and Kasa (2015), and Giacomini, Skreta, and Turén (2015), where agents switch between models on the basis of a specification test, the agents act as if they have complete trust in the set of models currently entertained and are never concerned about being misspecified. Still, there is a tight connection between the robust control decision criterion and a maxmin decision criterion where the set of models expands as the penalization term in the robust control increases (see Hansen and Sargent, 2011, for a textbook treatment). In light of this, compared to the previous set of papers, our work can additionally be interpreted as providing the first *smooth* framework for

expanding (or restricting) the set of considered models as a function of the evidence. Farther afield, Ortoleva (2012) proposes and axiomatizes a model where a decision maker can reject their model in favor of a backup one when faced with events with sufficiently low probability. Karni and Vierø (2013) proposes and axiomatizes a model where the agent becomes progressively aware of more states and acts. However, their decision maker trusts their probability over states completely when making decisions. Banerjee, Chassang, Montero, and Snowberg (2020) studies a Wald problem where the agent trade-offs between robustness and the subjective expected utility performance of the experiment. Differently from us, the concern in this model does not evolve, and the agent makes a single decision. Epstein and Ji (2022) characterizes optimal stopping with a concern for robustness captured by maxmin preferences, showing that the robustness concern, in general, induces earlier stopping.

There is fast-growing literature on learning under misspecification with subjective expected utility preferences. Arrow and Green (1973) gives the first general framework for this problem, and Nyarko (1991) points out that the combination of misspecification and endogenous data can lead to cycles. This literature has been revived by the more recent Esponda and Pouzo (2016); see Bohren and Hauser (2021), Esponda, Pouzo, and Yamamoto (2021a), Fudenberg, Lanzani, and Strack (2021), and Frick, Iijima, and Ishii (2023) for analyses of more closely related settings.

The identification of an agent who is disappointed with minor discrepancies between the empirical and the theoretical distributions as a believer in the Law of Small Numbers follows the formalization of this bias proposed by Rabin (2002). The normative role of the likelihood ratio that makes it proportional to the relative entropy (cf. Proposition 1 and Theorem 1) is somewhat reminiscent of the normative role of (absolute) entropy as a measure of informativeness found by Cabrales, Gossner, and Serrano (2013).

Hansen and Sargent (2007) mention a time-varying penalization parameter as a way to maintain dynamic consistency in the robust control model. Maenhout (2004) also uses a time-varying penalization parameter in a portfolio selection problem to keep the recursive discounted preferences homothetic at any history. See Pathak et al.

(2002) for a critical perspective on the latter paper and the subsequent literature. In both cases, the parameter evolution does not capture the fit of the models to the observed data. Anderson, Hansen, and Sargent (2003) and Barillas, Hansen, and Sargent (2009) pioneer a literature that calibrates the (time-invariant) concern for misspecification from the acceptable error probability in likelihood ratio test between the unperturbed model and the worst-case model (that does not depend on the action there). See Hansen and Sargent (2011) for a textbook treatment.

In the evolutionary game theory literature studying preferences formation, the closer papers are Dekel, Ely, and Yilankaya (2007) and Robatto and Szentes (2017). In the former, players may be "misspecified" in the sense of having a vN-M utility different from the one determining reproductive finiteness. In the latter, the evolutionary pressure determines the risk attitudes of the players.

On the axiomatic side, the static decision criterion considered here is due to Cerreia-Vioglio, Hansen, Maccheroni, and Marinacci (2022).[32] The explicit use of a state space where every state describes both the single-period outcome realization and the probability distribution over outcomes follows the approach introduced in Cerreia-Vioglio, Maccheroni, Marinacci, and Montrucchio (2013a) as a two-stage "statistical" interpretation and axiomatization of some of the decision criteria under ambiguity, in particular the smooth ambiguity one. For this criterion, this approach has been recently extended by Denti and Pomatto (2022). They allow for a fully revealed-preference elicitation of the relevant probability distributions, viewed as subjective statistical models. See also Dean and Ortoleva (2017) for a less related decision criterion where the agent has a prior over multiple data-generating processes and evaluates each of them with rank-dependent utility and Gilboa, Minardi, and Samuelson (2020) for a different quasi-Bayesian criterion that combines Bayesian updating with case-reasoning rather than misspecification considerations.

[32] Given the Donsker–Varadhan variational formula, our decision criterion can also be seen as the average of CARA certain equivalents, an object studied and characterized from a statistical perspective in Mu, Pomatto, Strack, and Tamuz (2021).

1.6.2 Experimental Evidence

We are unaware of experiments that explicitly test the positive relation between misspecification and the belief in the Law of Small Numbers with uncertainty aversion. However, the findings in Esponda, Vespa, and Yuksel (2022) suggest that a mechanism similar to the one outlined in this chapter is actually at play. The paper studies the repeated behavior of two groups of agents, one with an agnostic (full support) belief about the possible data-generating process faced and one that is misspecified because of base rate neglect. The long-run average play of the misspecified agents is in between the best reply to the misspecified model and the uniform distribution over outcomes. Notably, this behavior is not the best reply to the observed empirical frequency, which suggests that, as in our model, even in the medium run (200 repetitions in the experiment above), the agents do not altogether drop their models; they rely less on it to make their choices. Instead, the correctly specified agents converge to making choices that are optimal only under the actual data-generating process, i.e., they behave as a subjective expected utility maximizer with a belief concentrated on the true DGP. More indirectly De Filippis, Guarino, Jehiel, and Kitagawa (2022) show that there is an overreaction of beliefs to consistent signal than to inconsistent signal, suggesting that agents who find a model consistently validated may rely more on it to make decisions.

1.6.3 Forward-looking Agents

One key generalization to our model would be to allow for forward-looking agents. Of course, as for many decision criteria that depart from SEU, the main complication is dealing with the fact that the most immediate extension of the criterion to forward-looking agents would induce dynamic inconsistencies under some information structures (see Appendix .1.4 for simple explicit example). One approach would be to directly impose a recursive formulation for the preferences, as in Maccheroni, Marinacci, and Rustichini (2006b) and Klibanoff, Marinacci, and Mukerji (2009). Since the decision criterion belongs to the variational class, we know from the

first reference that a recursive formulation can be obtained. A complementary approach does not impose recursivity and allows for dynamic inconsistency. Preliminary analysis suggests that if we consider agents who do not anticipate their future taste variations, little is changed. However, analyzing an uncommitted, forward-looking, and sophisticated agent playing an intra-personal equilibrium with their future selves would require combining the insights of this chapter with the approach developed in Battigalli, Francetich, Lanzani, and Marinacci (2019). Analogously, to extend the axiomatic exercise to sophisticated agents, the techniques of this chapter should be combined with the consistent planning approach of Siniscalchi (2011).

1.6.4 Endogenous Structured Models

The more natural extension for the decision-theoretic part of the chapter involves using axioms that do not explicitly allow the agent to bet on the identity of the structured model. Allowing such bets is relatively standard when dealing with two levels of uncertainty for which the agent has different attitudes (Klibanoff, Marinacci, and Mukerji, 2005 being the most prominent example). However, Denti and Pomatto (2022) proposed an identifiability condition that avoids the need for explicit bets on the structured models. Identifiability requires a way to partition S that singles out the probabilistic model. In particular, each probabilistic model assigns probability one to its corresponding partition element.

When considering a structured environment, we required that this identification is spelled out in the description of the states, with the second component being the distribution ρ. Although in light of the results of Denti and Pomatto (2022), the axiomatization of this static criterion without this restriction does not generate conceptual complications, the dynamic characterization becomes significantly more involved. In particular, the challenge is created by the conditioning with respect to the endogenously identified model. This substantial extension is left for future work.

1.7 Conclusion

In this work, we propose a novel model of agents actively learning about the environment and dynamically adjusting their concern for misspecification based on the evidence they face. We show that the agents develop different long-run uncertainty attitudes depending on their understanding of how quickly evidence in favor or against a model is accumulated. Statistically sophisticated agents converge to robust control preferences a la Hansen and Sargent (2001), with the misspecification concern endogenously determined by their models fit with the true DGP at the equilibrium action. In contrast, an agent who is too demanding in evaluating their model converges to behave as a maxmin agent a la Gilboa and Schmeidler (1989), while a lenient agent eventually becomes a standard subjective expected utility maximizer. These results provide the first learning foundation for nonstandard decision criteria.

We then point out that in natural environments, the behavior of the statistically sophisticated type need not converge, and we characterize the limit frequency of time spent playing each action. We apply this result to a simple macroeconomic model and obtain a new rationale for the periodic switches in monetary policies.

We also provide an axiomatization of the proposed decision criterion and its evolution in the face of evidence. We introduce a new axiom type, Asymptotic Frequentism, requiring long streams of outcomes with the same empirical frequency to induce similar preferences. We prove that this axiom induces the statistically sophisticated behavior studied in the learning part of the chapter.

.1 Appendix

.1.1 Learning Results

Preliminaries

By Assumption 1, there exists $K \in \mathbb{R}_{++}$ such that

$$-\ln \tilde{q}_a(y) \leq K \qquad \forall a \in A, \forall y \in Y, \forall q \in Q.$$

Throughout Appendix .1.1, the symbol K will denote such strictly positive real number.

For an arbitrary Borel measurable subset C of a metric space, we endow the space $C^{\mathbb{N}}$ with the Borel σ-algebra, $\mathcal{B}\left(C^{\mathbb{N}}\right)$, corresponding to the product topology on $C^{\mathbb{N}}$. For $k_1, .., k_t \in C$, $t \in \mathbb{N}$, we denote by $k^t = (k_1, ..., k_t)$ both the finite sequence in C^t and the elementary cylinder in $C^{\mathbb{N}}$ that it identifies. For every policy $\Pi \in A^{\mathcal{H}}$, the density of the objective probability distribution over infinite histories is defined over a finite number of periods $I \subseteq \mathbb{N}$ with $t_I = \max I$ as

$$\tilde{\mathbb{P}}_{\Pi}\left((a_\tau, y_\tau)_{\tau \in I}\right) = \begin{cases} 1 & \text{if } \exists \left(\hat{a}_\tau, \hat{y}_\tau\right)_{\tau=1}^{t_I} \in (A \times Y)^{t_I} : \\ & \hat{a}_{\tau+1} = \Pi\left(\hat{a}^\tau, \hat{y}^\tau\right), \forall \tau \in \{0, ..., t_I - 1\} \\ & \text{and } \left(\hat{a}_\tau, \hat{y}_\tau\right) = \left(a_\tau, y_\tau\right), \forall \tau \in I, \\ 0 & \text{otherwise,} \end{cases} \tag{8}$$

and

$$\mathbb{P}_{\Pi}\left((a_\tau)_{\tau \in I}, C\right) = \int_C \tilde{\mathbb{P}}_{\Pi}\left((a_\tau, \cdot)_{\tau \in I}\right) d\left(\prod_{\tau \in I} p^*_{a_\tau}\right) \qquad \forall C \in \mathcal{B}\left(Y^I\right).$$

Since the corresponding set of finite-dimensional probability measures is consistent, there is a unique probability measure over infinite sequences of action-outcome pairs with these marginals, defined through the Kolmogorov extension theorem (see Theorem V.5.1 in Parthasarathy, 2005 for the version for standard Borel spaces used here).

For every $t \in \mathbb{N}$ and history $h_t = (a^t, y^t) \in \mathcal{H}$ let $p^{h_t} \in \Delta\left(Y\right)^A$ be the action contingent (finite support) probability measure over outcomes corresponding to the empirical frequency: for all $a \in A$ such that $\sum_{\tau=1}^t \mathbb{I}_{\{a\}}\left(a_\tau\right) > 0$,

$$p_a^{h_t}\left(C\right) = \frac{\sum_{\tau=1}^t \mathbb{I}_{\{(a,y): y \in C\}}\left(a_\tau, y_\tau\right)}{\sum_{\tau=1}^t \mathbb{I}_{\{a\}}\left(a_\tau\right)} \qquad \forall C \subseteq \mathcal{B}\left(Y\right)$$

and $p_a^{h_t} = \delta_{\bar{y}}$ for some arbitrary fixed $\bar{y} \in Y$ if $\sum_{\tau=1}^t \mathbb{I}_{\{a\}}\left(a_\tau\right) = 0$. For every two histories $h_t, h_\tau \in \mathcal{H}$ we write $h_t \succ h_\tau$ if there is $n \in \mathbb{N}$ and $(a_i, y_i)_{i=1}^n$ such that $h_t = (h_\tau, (a_i, y_i)_{i=1}^n)$. For all $b \in A$ let Π^b the policy that prescribes b at every

period. Define the set $Q^{\varepsilon}(a)$ as all parameters at most ε away from a relative entropy minimizer given action a,

$$Q^{\varepsilon}(a) = \{q \in Q \colon \exists q' \in Q(a) \cap B_{\varepsilon}(q)\}. \tag{9}$$

Results

Our first lemma justifies the repeated use of min and max rather than inf and sup in the definitions of the average robust criterion and LRT.

Lemma 1. *1. For every $a \in A$, $\lambda \in \mathbb{R}_{++}$, and $q \in Q$*

$$\emptyset \neq argmin_{p_a \in \Delta(Y)} \left(\mathbb{E}_{p_a}[u(a,y)] + \frac{R(p_a||q_a)}{\lambda} \right).$$

2. For every $t \in \mathbb{N}$, $h_t \in \mathcal{H}_t$,

$$\emptyset \neq argmax_{p \in N(Q)} \prod_{\tau=1}^{t} \tilde{p}_{a_\tau}(y_\tau).$$

Proof. 1) Fix $a \in A$, $\lambda \in \mathbb{R}_{++}$, and $q \in Q$. Since u is continuous and Y is compact, $u(a,\cdot)$ is bounded, $\mathbb{E}_{p_a}[u(a,y)] \geq \min_{y \in Y} u(a,y) \in \mathbb{R}$ for all $p_a \in \Delta(Y)$, and $p_a \mapsto \mathbb{E}_{p_a}[u(a,y)]$ is continuous. Since Y is a compact metric space, by, e.g., Royden and Fitzpatrick (1988), it is a Polish space and so $p_a \mapsto R(p_a||q_a)$ is lower semicontinuous by Lemma 1.4.3 in Dupuis and Ellis (2011). Therefore, the set

$$E \colon = \left\{ p_a \in \Delta(Y) : \mathbb{E}_{p_a}[u(a,y)] + \frac{1}{\lambda} R(p_a||q_a) \leq \mathbb{E}_{q_a}[u(a,y)] \right\}$$

is closed, and as $R(q_a||q_a) = 0$ we clearly have

$$\text{argmin}_{p_a \in \Delta(Y)} \mathbb{E}_{p_a}[u(a,y)] + \frac{1}{\lambda} R(p_a||q_a) = \text{argmin}_{p_a \in E} \mathbb{E}_{p_a}[u(a,y)] + \frac{1}{\lambda} R(p_a||q_a).$$

Since Y is a compact metric space, by Theorem 15.11 in Aliprantis and Border (2013) and Proposition 11.15 in Royden and Fitzpatrick (1988) so are $\Delta(Y)$ and E endowed

with the topology of weak convergence of measures. Since

$$p_a \mapsto \mathbb{E}_{p_a}\left[u\left(a,y\right)\right] + \frac{1}{\lambda} R\left(p_a||q_a\right)$$

is real-valued (with values in $[\min_{y\in Y} u\left(a,y\right), \mathbb{E}_{q_a}\left[u\left(a,y\right)\right]]$) and lower semicontinuous on the compact E, it admits a minimizer by the generalized Weierstrass' theorem (see, e.g., Theorem 2.43 in Aliprantis and Border, 2013).

2) Let $t \in \mathbb{N}$, $h_t \in \mathcal{H}_t$. It follows from Assumption 2 (ii) and Theorem 1 in Sweeting (1986) that

$$\begin{array}{ccc} N\left(Q\right) & \to & \mathbb{R} \\ p & \mapsto & \prod_{\tau=1}^{t} \tilde{p}_{a_\tau}(y_\tau) \end{array}$$

is continuous.[33] Then, the maximum is attained since $N\left(Q\right)$ is closed by Assumption 2 (i), and thus compact since Y is a compact metric space, and by Theorem 15.11 in Aliprantis and Border (2013) and Proposition 11.15 in Royden and Fitzpatrick (1988). ■

The next lemma provides a useful rewriting of the LRT as a weighted average of the empirical log-likelihood ratio when playing the different actions, with weights proportional to how frequently each action has been used in the past.

Lemma 2. *For every $t \in \mathbb{N}$ and $h_t = (a^t, y^t) \in \mathcal{H}_t$, if $q' \in argmax_{q\in Q} \prod_{\tau=1}^{t} \tilde{q}_{a_\tau}(y_\tau)$, and $p \in argmax_{r\in N(Q)} \prod_{\tau=1}^{t} \tilde{r}_{a_\tau}(y_\tau)$ then*

$$LLR\left(h_t, Q\right) = \sum_{a\in A} \sum_{\tau=1}^{t} \mathbb{I}_{\{a\}}\left(a_\tau\right) \int_Y \log\left(\frac{dp_a\left(y\right)}{dq'_a\left(y\right)}\right) dp_a^{h_t}\left(y\right).$$

[33] Observe that although the theorem in Sweeting (1986) is stated for densities with respect to the Lebesgue measure, both Scheffe's Theorem and Theorem 2.18 in Rudin (1970), that are used to prove the result, work for densities with respect to any regular Borel measure, as the $(p_a^*)_{a\in A}$ we consider are (by, Y being metric and, e.g., Theorem II.1.2 in Parthasarathy, 2005).

Proof. We have

$$\begin{aligned}
& LLR\left(h_t, Q\right) = -\log\left(\frac{\max_{q\in Q}\prod_{\tau=1}^{t} q'_{a_\tau}(y_\tau)}{\max_{r\in N(Q)}\prod_{\tau=1}^{t}\tilde{r}_{a_\tau}(y_\tau)}\right) \\
= & \log\left(\frac{\max_{r\in N(Q)}\prod_{\tau=1}^{t}\tilde{r}_{a_\tau}(y_\tau)}{\max_{q\in Q}\prod_{\tau=1}^{t}\tilde{q}_{a_\tau}(y_\tau)}\right) = \log\left(\frac{\prod_{\tau=1}^{t}\tilde{p}_{a_\tau}(y_\tau)}{\prod_{\tau=1}^{t}\tilde{q}'_{a_\tau}(y_\tau)}\right) \\
= & \log\left(\prod_{\tau=1}^{t}\tilde{p}_{a_\tau}(y_\tau)\right) - \log\left(\prod_{\tau=1}^{t}\tilde{q}'_{a_\tau}(y_\tau)\right) \\
= & \log\left(\prod_{y\in Y}\prod_{a\in A}\tilde{p}_a(y)^{\sum_{\tau=1}^{t}\mathbb{I}_{\{a\}}(a_\tau)p_a^{h_t}(\{y\})}\right) - \log\left(\prod_{y\in Y}\prod_{a\in A}\tilde{q}'_a(y)^{\sum_{\tau=1}^{t}\mathbb{I}_{\{a\}}(a_\tau)p_a^{h_t}(\{y\})}\right) \\
= & \sum_{a\in A}\sum_{\tau=1}^{t}\mathbb{I}_{\{a\}}(a_\tau)\sum_{y\in Y}p_a^{h_t}(\{y\})\log\left(\tilde{p}_a(y)\right) - \sum_{a\in A}\sum_{\tau=1}^{t}\mathbb{I}_{\{a\}}(a_\tau)\sum_{y\in Y}p_a^{h_t}(\{y\})\log\left(\tilde{q}'_a(y)\right) \\
= & \sum_{a\in A}\sum_{\tau=1}^{t}\mathbb{I}_{\{a\}}(a_\tau)\left(\sum_{y\in Y}p_a^{h_t}(\{y\})\log\left(\tilde{p}_a(y)\right) - \sum_{y\in Y}p_a^{h_t}(\{y\})\log\left(\tilde{q}'_a(y)\right)\right) \\
= & \sum_{a\in A}\sum_{\tau=1}^{t}\mathbb{I}_{\{a\}}(a_\tau)\int_Y \log\left(\frac{\tilde{p}_a(y)}{\tilde{q}'_a(y)}\right)\mathrm{d}p_a^{h_t}(y).
\end{aligned}$$

■

The next lemma shows that a robust control evaluation with respect to a structured model $q \in Q$ converges to a subjective expected utility evaluation as λ tends to 0, generalizing previous results in the decision-theoretic literature, where the function evaluated was a finite range one, to continuous utility functions.

Lemma 3. *For every $a \in A$, $q \in Q$, and $(q_n, \lambda_n)_{n\in\mathbb{N}} \in (Q \times \mathbb{R}_{++})^{\mathbb{N}}$ with*

$$\lim_{n\to\infty}(q_n, \lambda_n) = (q, 0)$$

we have

$$\lim_{n\to\infty}\min_{p_a\in\Delta(Y)}\left(\mathbb{E}_{p_a}\left[u\left(a,y\right)\right] + \frac{R\left(p_a||q_a\right)}{\lambda_n}\right) = \mathbb{E}_{q_a}\left[u\left(a,y\right)\right].$$

Proof. Fix $a \in A$ and define $\bar{u} = \max_{y\in Y} u(a,y) - \min_{y\in Y} u(a,y)$. For every $n \in \mathbb{N}$,

$$\min_{p_a\in\Delta(Y)}\left(\mathbb{E}_{p_a}\left[u\left(a,y\right)\right] + \frac{R\left(p_a||q_{a,n}\right)}{\lambda_n}\right) \in \left[\min_{y\in Y} u\left(a,y\right), \mathbb{E}_{q_{a,n}}\left[u\left(a,y\right)\right]\right] \subseteq \left[\min_{y\in Y} u\left(a,y\right), \max_{y\in Y} u\left(a,y\right)\right],$$

so possibly restricting to a subsequence, we can assume that the limit in the LHS of the statement is well defined. The statement is then proved by showing that any such subsequence converges to the RHS. In particular, we show that we cannot have

$$\lim_{n\to\infty} \min_{p_a\in\Delta(Y)} \left(\mathbb{E}_{p_a}\left[u\left(a,y\right)\right] + \frac{R\left(p_a||q_{a,n}\right)}{\lambda_n} \right) < \mathbb{E}_{q_a}\left[u\left(a,y\right)\right]. \tag{10}$$

This is sufficient as $\lim_{n\to\infty} \mathbb{E}_{q_{a,n}}\left[u\left(a,y\right)\right] = \mathbb{E}_{q_a}\left[u\left(a,y\right)\right]$ and therefore we know by the lower semicontinuity of R (see Lemma 1.4.3 in Dupuis and Ellis (2011)) that

$$\lim_{n\to\infty} \min_{p_a\in\Delta(Y)} \left(\mathbb{E}_{p_a}\left[u\left(a,y\right)\right] + \frac{R\left(p_a||q_{a,n}\right)}{\lambda_n} \right) \leq \mathbb{E}_{q_a}\left[u\left(a,y\right)\right].$$

If equation (10) held, there would be an $\varepsilon \in \mathbb{R}_{++}$ with

$$\lim_{n\to\infty} \min_{p_a\in\Delta(Y)} \left(\mathbb{E}_{p_a}\left[u\left(a,y\right)\right] + \frac{R\left(p_a||q_{a,n}\right)}{\lambda_n} \right) = \mathbb{E}_{q_a}\left[u\left(a,y\right)\right] - \varepsilon. \tag{11}$$

For every $n \in \mathbb{N}$, let $p_a^n \in \Delta\left(Y\right)$ be an arbitrary element of

$$\operatorname{argmin}_{p_a\in\Delta(Y)} \left(\mathbb{E}_{p_a}\left[u\left(a,y\right)\right] + \frac{R\left(p_a||q_{a,n}\right)}{\lambda_n} \right).$$

Since Y is a compact metric space, by Theorem 15.11 in Aliprantis and Border (2013) so is $\Delta\left(Y\right)$, and therefore, we can assume (by restricting to a subsequence) that p_a^n converges to some $\hat{p}_a \in \Delta\left(Y\right)$. By equation (11) and the fact that $\lim_{n\to\infty} p_a^n = \hat{p}_a$, we have

$$\mathbb{E}_{\hat{p}_a}\left[u\left(a,y\right)\right] \leq \mathbb{E}_{q_a}\left[u\left(a,y\right)\right] - \varepsilon.$$

Therefore,

$$\begin{aligned}
&\int_0^{\bar{u}} 1 - \hat{p}_a\left(\left\{y \in Y : u\left(a,y\right) - \min_{\bar{y}\in Y} u\left(a,\bar{y}\right) \leq x\right\}\right) \mathrm{d}x + \frac{3}{4}\varepsilon \\
= \;& \mathbb{E}_{\hat{p}_a}\left[u\left(a,y\right)\right] + \frac{3}{4}\varepsilon + \min_{\bar{y}\in Y} u\left(a,\bar{y}\right) \leq \mathbb{E}_{q_a}\left[u\left(a,y\right)\right] + \min_{\bar{y}\in Y} u\left(a,\bar{y}\right) \\
= \;& \int_0^{\bar{u}} 1 - q_a\left(\left\{y \in Y : u\left(a,y\right) - \min_{\bar{y}\in Y} u\left(a,\bar{y}\right) \leq x\right\}\right) \mathrm{d}x.
\end{aligned} \tag{12}$$

Claim 1. *There exist $M \in \mathbb{R}$ and $L \in \mathbb{R}_{++}$ such that*

$$\hat{p}_a\left(\{y \in Y : u(a,y) \leq M - L\}\right) - q_a\left(\{y \in Y : u(a,y) \leq M\}\right) \geq \frac{\varepsilon}{2\bar{u}}. \tag{13}$$

Proof of the Claim. Suppose that for every $M \in \mathbb{R}$ and $L \in \mathbb{R}_{++}$ equation (13) does not hold. Then for every $L \in \mathbb{R}_{++}$

$$\begin{aligned}
& \int_0^{\bar{u}} 1 - \hat{p}_a\left(\left\{y \in Y : u(a,y) - \min_{\bar{y}\in Y} u(a,\bar{y}) \leq x\right\}\right) \mathrm{d}x \\
= & \int_0^{\bar{u}+L} 1 - \hat{p}_a\left(\left\{y \in Y : u(a,y) - \min_{\bar{y}\in Y} u(a,\bar{y}) + L \leq x\right\}\right) \mathrm{d}x - L \\
= & \int_0^{\bar{u}+L} 1 - \hat{p}_a\left(\left\{y \in Y : u(a,y) - \min_{\bar{y}\in Y} u(a,\bar{y}) + L \leq x\right\}\right) \mathrm{d}x - L \\
\geq & \int_0^{\bar{u}+L} 1 - q_a\left(\left\{y \in Y : u(a,y) - \min_{\bar{y}\in Y} u(a,\bar{y}) \leq x\right\}\right) - \frac{\varepsilon}{2\bar{u}}\mathrm{d}x - L \\
= & \int_0^{\bar{u}} 1 - q_a\left(\left\{y \in Y : u(a,y) - \min_{\bar{y}\in Y} u(a,\bar{y}) \leq x\right\}\right) - \frac{\varepsilon}{2\bar{u}}\mathrm{d}x - L \\
= & \int_0^{\bar{u}} 1 - q_a\left(\left\{y \in Y : u(a,y) - \min_{\bar{y}\in Y} u(a,\bar{y}) \leq x\right\}\right) - \varepsilon/2 - L\frac{\varepsilon}{2\bar{u}} - L.
\end{aligned}$$

Since L can be chosen to be arbitrarily small, we have

$$\begin{aligned}
& \int_0^{\bar{u}} 1 - \hat{p}_a\left(\left\{y \in Y : u(a,y) - \min_{\bar{y}\in Y} u(a,\bar{y}) \leq x\right\}\right) \mathrm{d}x \\
\geq & \int_0^{\bar{u}} 1 - q_a\left(\left\{y \in Y : u(a,y) - \min_{\bar{y}\in Y} u(a,\bar{y}) \leq x\right\}\right) \mathrm{d}x - \varepsilon/2,
\end{aligned}$$

a contradiction with equation (12). □

The claim, in turn, implies that there exists $N \in \mathbb{N}$ such that for all $n \geq N$

$$p_a^n\left(\left\{y \in Y : u(a,y) \leq M - \frac{L}{2}\right\}\right) - q_{a,n}\left(\left\{y \in Y : u(a,y) \leq M - \frac{L}{2}\right\}\right) \geq \frac{\varepsilon}{4\bar{u}}.$$

But then

$$\begin{aligned}
&\min_{p_a\in\Delta(Y)} \mathbb{E}_{p_a}\left[u\left(a,y\right)\right]+\frac{1}{\lambda_n}R\left(p_a||q_{a,n}\right)\\
=\;& \mathbb{E}_{p_a^n}\left[u\left(a,y\right)\right]+\frac{1}{\lambda_n}R\left(p_a^n||q_{a,n}\right)\\
\geq\;& \min_{y\in Y} u\left(a,y\right)+\left(p_a^n\left(\left\{y\in Y:u\left(a,y\right)\leq M-\frac{L}{2}\right\}\right)\log\frac{p_a^n\left(\left\{y\in Y:u\left(a,y\right)\leq M-\frac{L}{2}\right\}\right)}{q_{a,n}\left(\left\{y\in Y:u\left(a,y\right)\leq M-\frac{L}{2}\right\}\right)}\right)/\lambda_n\\
&+\left(p_a^n\left(\left\{y\in Y:u\left(a,y\right)> M-\frac{L}{2}\right\}\right)\log\frac{p_a^n\left(\left\{y\in Y:u\left(a,y\right)> M-\frac{L}{2}\right\}\right)}{q_{a,n}\left(\left\{y\in Y:u\left(a,y\right)> M-\frac{L}{2}\right\}\right)}\right)/\lambda_n
\end{aligned}$$

where the inequality follows from Theorem 1.24 in Liese and Vajda (1987). But the last term diverges to $+\infty$ as n goes to infinity, a contradiction with

$$\min_{p_a\in\Delta(Y)} \mathbb{E}_{p_a}\left[u\left(a,y\right)\right]+\frac{1}{\lambda_n}R\left(p_a||q_{a,n}\right)\leq\max_{y\in Y} u\left(a,y\right)<\infty.$$

■

Lemma 4. *1. For every $a\in A$, the function $G:\Delta\left(Q\right)\times\mathbb{R}_+\to\mathbb{R}$ defined by*

$$G\left(\nu,\lambda\right)=\int_Q\min_{p_a\in\Delta(Y)}\left(\mathbb{E}_{p_a}\left[u\left(a,y\right)\right]+\frac{R\left(p_a||q_a\right)}{\lambda}\right)d\nu\left(q\right)\quad\forall\nu\in\Delta\left(Q\right),\forall\lambda\in\mathbb{R}_{++}$$

and

$$G\left(\nu,0\right)=\int_Q\mathbb{E}_{q_a}\left[u\left(a,y\right)\right]d\nu\left(q\right)\qquad\forall\nu\in\Delta\left(Q\right)$$

is continuous.

2. The correspondence $BR^{(\cdot)}\left(\cdot\right):\mathbb{R}_+\times\Delta\left(Q\right)\rightrightarrows A$ where

$$BR^{(0)}\left(\nu\right):=BR^{Seu}\left(\nu\right)\qquad\forall\nu\in\Delta\left(Q\right)$$

is upper hemicontinuous.

Proof. (1) Fix $a\in A$. For every $q\in Q$, let $F\left(q,0\right):=\mathbb{E}_{q_a}\left[u\left(a,y\right)\right]$ and observe that

for each $\lambda \in \mathbb{R}_{++}$, by Proposition 1.4.2 in Dupuis and Ellis (2011) we have

$$F(q,\lambda) : = \min_{p_a \in \Delta(Y)} \left(\mathbb{E}_{p_a}\left[u(a,y)\right] + \frac{R(p_a||q_a)}{\lambda} \right) = \frac{-\log\left(\int_Y \exp\left(-\lambda u(a,y)\right) \mathrm{d}q_a(y)\right)}{\lambda}.$$

Since Y is compact and u is continuous, for all $\lambda \in \mathbb{R}_{++}$ and $q \in Q$, the RHS belongs to

$$\left[\min_{y \in Y} u(a,y), \max_{y \in Y} u(a,y)\right].$$

For every $\lambda \in \mathbb{R}_{++}$, $\exp\left(-\lambda u(a,\cdot)\right)$ is a continuous and bounded function that is bounded away from 0. Therefore,

$$q \mapsto \int_Y \exp\left(-\lambda u(a,y)\right) \mathrm{d}q_a(y)$$

is continuous by definition of the weak convergence of measures, and F is continuous by Lemma 3 (at $\lambda = 0$) and Theorem 15.7.3 in Kallenberg (1973) (at $\lambda \neq 0$).

Let $(\nu_n, \lambda_n)_{n \in \mathbb{N}} \in \Delta(Q) \times \mathbb{R}_{++}$ be a convergent sequence with limit (ν, λ). Suppose first that $\lambda > 0$. Then

$$\lim_{n \to \infty} \int_Q \frac{\log\left(\int_Y \exp\left(-\lambda_n u(a,y)\right) \mathrm{d}q_a(y)\right)}{-\lambda_n} \mathrm{d}\nu_n(q) = \int_Q \frac{\log\left(\int_Y \exp\left(-\lambda u(a,y)\right) \mathrm{d}q_a(y)\right)}{-\lambda} \mathrm{d}\nu(q)$$

since weak convergence implies vague convergence, by Theorem 15.7.3 in Kallenberg (1973) and the joint continuity of F established above. Next, suppose that $\lambda = 0$. Then

$$\lim_{n \to \infty} \int_Q -\frac{\log\left(\int_Y \exp\left(-\lambda_n u(a,y)\right) \mathrm{d}q_a(y)\right)}{\lambda_n} \mathrm{d}\nu_n(q) = \int_Q \int_Y u(a,y) \, \mathrm{d}q_a(y) \, \mathrm{d}\nu(q)$$

again by Theorem 15.7.3 in Kallenberg (1973) and the joint continuity of F established above. This proves (i).

(2) Follows by (1) and Theorem 17.31 in Aliprantis and Border (2013). ■

Lemma 5. *1. For every $a \in A$, if $(q_n, p_a^n)_{n\in\mathbb{N}} \in (Q \times \Delta(Y))^{\mathbb{N}}$ is such that*

$$\lim_{n\to\infty} (q_n, p_a^n)_{n\in\mathbb{N}} = (q', \bar{p}_a)$$

and $\text{supp}p_a^n \subseteq \{y \in Y : -\ln \tilde{q}_{a,n}(y) \leq K\}$ *for all $n \in \mathbb{N}$ then*

$$\lim_{n\to\infty} -\int_Y \log\left(\tilde{q}_{a,n}(y)\right) dp_a^n(y) = -\int_Y \log\left(\tilde{q}'_a(y)\right) d\bar{p}_a(y).$$

2. For every $a \in A$, if $(q_n, q'_n, p_a^n)_{n\in\mathbb{N}} \in (Q \times Q \times \Delta(Y))^{\mathbb{N}}$ is such that

$$\lim_{n\to\infty} (q_n, q'_n, p_a^n)_{n\in\mathbb{N}} = \left(\underline{q}, q', \bar{p}_a\right)$$

and $\text{supp}p_a^n \subseteq \{y \in Y : -\ln \tilde{q}_{a,n}(y) \leq K\}$, $\text{supp}p_a^n \subseteq \left\{y \in Y : -\ln \tilde{q}'_{a,n}(y) \leq K\right\}$, *for all $n \in \mathbb{N}$ then*

$$\lim_{n\to\infty} -\int_Y \log\left(\frac{\tilde{q}_{a,n}(y)}{\tilde{q}'_{a,n}(y)}\right) dp_a^n(y) = -\int_Y \log\left(\frac{\underline{q}_a(y)}{\tilde{q}'_a(y)}\right) d\bar{p}_a(y).$$

Proof. By Assumption 1 (i-ii) the assumptions of Theorem 15.7.3 in Kallenberg (1973) are satisfied for the sequences of integrand functions and probability measures $\left(\log\left(\tilde{q}_{a,n}\right), p_a\right)_{n\in\mathbb{N}}$ and $\left(\log\left(\tilde{q}'_{a,n}\right), p_{a,n}\right)_{n\in\mathbb{N}}$. ■

The following lemma shows that under every policy, almost surely the infinite sequence of observations do not contain a realization that provides an arbitrarily large evidence against a structured model.

Lemma 6. *For every $\Pi \in A^{\mathcal{H}}$ and $q \in Q$,*

$$\mathbb{P}_\Pi\left(\left\{(a_i, y_i)_{i\in\mathbb{N}} \in (A \times Y)^{\mathbb{N}} : \forall t \in \mathbb{N}, -\ln \tilde{q}_{a_t}(y_t) \leq K\right\}\right) = 1.$$

Proof. By Assumption 1 (i) for every $a \in A$, $\{y \in Y : -\ln \tilde{q}_a(y) \leq K\} \in \mathcal{B}(Y)$. By the definition of Radon-Nykodim derivative and equation (8), for every $t \in \mathbb{N}$,

$$\mathbb{P}_\Pi\left(\left\{(a_i, y_i)_{i\in\mathbb{N}} \in (A \times Y)^{\mathbb{N}} : -\ln \tilde{q}_{a_t}(y_t) > K\right\}\right) = 0.$$

Since $\mathbb{P}_\Pi$ is a measure, it is countably subadditive and so

$$\begin{aligned}
& \mathbb{P}_\Pi\left(\left\{(a_i,y_i)_{i\in\mathbb{N}}\in(A\times Y)^{\mathbb{N}}:\forall t\in\mathbb{N},-\ln\tilde{q}_{a_t}(y_t)\leq K\right\}\right)\\
= & \ 1-\mathbb{P}_\Pi\left(\left\{(a_i,y_i)_{i\in\mathbb{N}}\in(A\times Y)^{\mathbb{N}}:\exists t,-\ln\tilde{q}_{a_t}(y_t)>K\right\}\right)\\
\geq & \ 1-\sum_{t=1}^{\infty}\mathbb{P}_\Pi\left(\left\{(a_i,y_i)_{i\in\mathbb{N}}\in(A\times Y)^{\mathbb{N}}:-\ln\tilde{q}_{a_t}(y_t)>K\right\}\right)=1,
\end{aligned}$$

proving the statement. ■

The following lemma shows that, on every history where the empirical action process stabilizes on α^*, and the empirical outcome distribution contingent on the actions played infinitely often converges to the true distribution, the limit LRT can be rewritten as the minimum of an α^* weighted average of the relative entropy from the true DGP.

Lemma 7. *Let $\alpha^*\in\Delta(A)$ and $(a_t,y_t)_{t\in\mathbb{N}}\in(A\times Y)^{\mathbb{N}}$ be such that there exists $q\in Q$ with $-\ln\tilde{q}_{a_t}(y_t)\leq K$ for all $t\in\mathbb{N}$. For every $t\in\mathbb{N}$, set $h_t=(a^t,y^t)$, and let $q(h_t)$ and $r(h_t)$ be two arbitrary elements of $argmax_{q\in Q}\prod_{\tau=1}^{t}\tilde{q}_{a_\tau}(y_\tau)$ and $argmax_{p\in N(Q)}\prod_{\tau=1}^{t}\tilde{p}_{a_\tau}(y_\tau)$, respectively. If*

$$\lim_{t\to\infty}\left(\alpha_t(h_t),\left(p_a^{h_t}\right)_{a\in\mathrm{supp}\alpha^*}\right)=\left(\alpha^*,\left(p_a^*\right)_{a\in\mathrm{supp}\alpha^*}\right)$$

then

$$\frac{LLR(h_t,Q)}{t}=\lim_{t\to\infty}\sum_{a\in A}\sum_{\tau=1}^{t}\mathbb{I}_{\{a\}}(a_\tau)\int_Y\log\left(\frac{\tilde{r}_a(h_t)(y)}{\tilde{q}_a(h_t)(y)}\right)dp_a^{h_t}(y)/t=\min_{q\in Q}\sum_{a\in A}\alpha^*(a)R(p_a^*||q_a).$$

Proof. By assumption of the lemma, for all $t\in\mathbb{N}$, we have

$$\begin{aligned}
& \sum_{a\in A}\sum_{\tau=1}^{t}\mathbb{I}_{\{a\}}(a_\tau)\int_Y\log\left(\frac{\tilde{r}_a(h_t)(y)}{\tilde{q}_a(h_t)(y)}\right)\mathrm{d}p_a^{h_t}(y)/t\\
= & \ \frac{\sum_{a\in A}\sum_{\tau=1}^{t}\mathbb{I}_{\{a\}}(a_\tau)\left(\int_Y\log\left(\tilde{r}_a(h_t)(y)\right)\mathrm{d}p_a^{h_t}(y)-\int_Y\log\left(\tilde{q}_a(h_t)(y)\right)\mathrm{d}p_a^{h_t}(y)\right)}{t}.
\end{aligned}$$

Let $a \in \text{supp}\alpha^*$. By Assumption 2 (i), we have

$$\int_Y -\log\left(\tilde{r}_a\left(h_t\right)\left(y\right)\right) \mathrm{d}p_a^{h_t}\left(y\right) \leq 0 \qquad \forall t \in \mathbb{N}.$$

Also, take any subsequence h_{t_n} in which $r_a\left(h_t\right)$ converges to some r_a

$$0 \leq \int_Y -\log\left(\tilde{r}_a\left(y\right)\right) \mathrm{d}p_a^*\left(y\right) \leq \liminf_{n\to\infty} \int_Y -\log\left(\tilde{r}_a\left(h_{t_n}\right)\left(y\right)\right) \mathrm{d}p_a^{h_{t_n}}\left(y\right)$$

where the first inequality follows from Gibbs inequality and the second since by Assumption 2 (ii), there exists $K' \in \mathbb{R}_{++}$ such that $-\log\left(\tilde{r}_a\left(h_{t_n}\right)\left(y\right)\right) \geq -K'$, p_a^*-almost surely and so we can apply Lemma 3.2 in Serfozo (1982). Therefore, we have

$$\lim_{t\to\infty} \sum_{a\in A} \sum_{\tau=1}^{t} \mathbb{I}_{\{a\}}\left(a_\tau\right) \int_Y -\log\left(\tilde{r}_a\left(h_t\right)\left(y\right)\right) \mathrm{d}p_a^{h_t}\left(y\right)/t = 0.$$

So

$$\begin{aligned}
& \lim_{t\to\infty} \sum_{a\in A} \sum_{\tau=1}^{t} \mathbb{I}_{\{a\}}\left(a_\tau\right) \int_Y \log\left(\frac{\tilde{r}_a\left(h_t\right)\left(y\right)}{\tilde{q}_a\left(h_t\right)\left(y\right)}\right) \mathrm{d}p_a^{h_t}\left(y\right)/t \\
= & -\lim_{t\to\infty} \sum_{a\in A} \sum_{\tau=1}^{t} \mathbb{I}_{\{a\}}\left(a_\tau\right) \int_Y \log\left(\tilde{q}_a\left(h_t\right)\left(y\right)\right) \mathrm{d}p_a^{h_t}\left(y\right)/t \\
= & -\lim_{t\to\infty} \min_{q\in Q} \sum_{a\in A} \sum_{\tau=1}^{t} \mathbb{I}_{\{a\}}\left(a_\tau\right) \int_Y \log\left(\tilde{q}_a\left(h_t\right)\left(y\right)\right) \mathrm{d}p_a^{h_t}\left(y\right)/t.
\end{aligned}$$

Therefore the result follows from Lemma 5 and Theorem 17.31 in Aliprantis and Border (2013). ■

Lemma 8. *For every $a \in A$,*

$$\lim_{k\to\infty} \sup_{q\in Q} \min_{p_a\in\Delta(Y)} \int_Y u\left(a,y\right) dp_a\left(y\right) + \frac{R\left(p_a||q_a\right)}{k} = \min_{y\in\cup_{q\in Q}\text{supp}q_a} u\left(a,y\right).$$

Proof. Let $\hat{y} \in \text{argmin}_{y\in\cup_{q\in Q}\text{supp}q_a} u\left(a,y\right)$. If $\max_{y\in Y} u\left(a,y\right) = u\left(a,\hat{y}\right)$ the statement is trivially true, so suppose that $\max_{y\in Y} u\left(a,y\right) > u\left(a,\hat{y}\right)$. By Assumption 1 (i) we

have that

$$\inf_{q\in Q} q_a\left(B_\varepsilon\left(\hat{y}\right)\right) > 0 \qquad \forall \varepsilon \in \mathbb{R}_{++}.$$

Otherwise, by the compactness of Q the portmanteau theorem (see, e.g., Theorem 11.1.1 Dudley, 2018) would imply that there exists $\hat{q} \in Q$ with $\hat{q}_a\left(B_{\varepsilon/2}\left(\hat{y}\right)\right) = 0$. But then, since there exists $\bar{q} \in Q$ with $\hat{y} \in \text{supp}\bar{q}_a = \text{supp}p_a^*$, and so $p_a^*\left(B_{\varepsilon/2}\left(\hat{y}\right)\right) > 0$, we would obtain a contradiction with $p_a^* \sim \hat{q}_a$. Fix $\bar{\varepsilon} \in \left(0, \frac{\max_{y\in Y} u(a,y) - u(a,\hat{y})}{2}\right)$. Since $u\left(a,\cdot\right)$ is continuous, there exists ε such that

$$y \in B_\varepsilon\left(\hat{y}\right) \implies u\left(a,y\right) \leq u\left(a,\hat{y}\right) + \bar{\varepsilon}.$$

Then, for all $q \in Q$

$$\begin{aligned} u\left(a,\hat{y}\right) &\leq \min_{p_a\in\Delta(Y)} \int_Y u\left(a,y\right) \mathrm{d}p_a + \frac{R\left(p_a||q_a\right)}{k} \\ &= -\frac{1}{k}\log\left(\int_Y \exp\left(-ku\left(a,y\right)\right) \mathrm{d}q_a\left(y\right)\right) \\ &\leq -\frac{1}{k}\log\left(\begin{array}{c} \exp\left(-k\left(u\left(a,\hat{y}\right)+\bar{\varepsilon}\right)\right) \inf_{\hat{q}\in Q} \hat{q}_a\left(B_\varepsilon\left(\hat{y}\right)\right) \\ + \left(1 - \inf_{\hat{q}\in Q} \hat{q}_a\left(B_\varepsilon\left(\hat{y}\right)\right)\right) \exp\left(-k \max_{y\in Y} u\left(a,y\right)\right) \end{array}\right) \end{aligned}$$

where the equality follows from Proposition 1.4.2 in Dupuis and Ellis (2011). Moreover, the last term converges to $u\left(a,\hat{y}\right) + \bar{\varepsilon}$ as k goes to infinity by a simple application of L'Hôpital's rule. Since $\bar{\varepsilon} < \frac{\max_{y\in Y} u(a,y) - u(a,\hat{y})}{2}$ was arbitrarily chosen, and the last term does not depend on q this proves the desired uniformity of the convergence. ■

Proof of Proposition 1. Let (u, a, Y) be a decision problem with $\{q^*\} = \text{argmin}_{q\in Q} Q\left(a\right)$ for all $a \in A$ and $\varepsilon \in \mathbb{R}_{++}$. We start by showing that there exists $c \in \mathbb{R}_{++}$ such that adjusting λ according to

$$\Lambda\left(\mathbf{h}_t\right) = \frac{LLR\left(\mathbf{h}_t, Q\right)}{ct} \tag{14}$$

is ε-safe and ε-consistent under almost correct specification. This is done by first deriving a $c \in \mathbb{R}_{++}$ such that ε-safety is satisfied, and then showing that there exists a δ that delivers ε-consistency under almost correct specification. Safety is trivially

satisfied by every policy if

$$\max_{a\in A}\min_{y\in Y} u\left(a,y\right)=\min_{a\in A,y\in Y} u\left(a,y\right),$$

so in that case pick an arbitrary $c\in\mathbb{R}_{++}$. Suppose instead that we have

$$\max_{a\in A}\min_{y\in Y} u\left(a,y\right)>\min_{a\in A,y\in Y} u\left(a,y\right).$$

Let $\hat{P}\subseteq\Delta\left(Y\right)^{A}$ be the set of p^{*} that satisfy Assumption 1 jointly with Q, and define

$$\underline{A}\left(p^{*}\right):=\left\{a'\in A:\max_{a\in A}\min_{y\in Y} u\left(a,y\right)>\mathbb{E}_{p^{*}_{a'}}\left[u\left(a',y\right)\right]+\frac{\varepsilon}{2}\right\}.$$

Claim 2. *There exists $\varphi^{*}>0$ such that for every $\Pi\in A^{\mathcal{H}}$ and $p^{*}\in\hat{P}$,*

$$\mathbb{P}_{\Pi}\left(\begin{array}{c}\left\{\liminf_{t\to\infty}\frac{\sum_{i=1}^{t}u(\mathbf{a}_i,\mathbf{y}_i)}{t}-\max_{a\in A}\min_{y\in Y}u\left(a,y\right)-\varepsilon<0\right\}\\ \cap\left\{\limsup_{t\to\infty}\alpha_t\left(\mathbf{h}_t\right)\left(a'\right)<\varphi^{*},\forall a'\in\underline{A}\left(p^{*}\right)\right\}\end{array}\right)=0.$$

That is, almost surely the payoff is at most ε-lower than the safe guarantee if the actions whose objective expected performance is lower than the guarantee are played sufficiently rarely (i.e., each of them has an average frequency lower than φ^{*}).

Proof of the Claim. Consider the stochastic process defined by

$$\mathbf{X}_t=u\left(\Pi\left(\mathbf{h}_{t-1}\right),\mathbf{y}_t\right)-\mathbb{E}_{p^{*}_{\Pi(\mathbf{h}_{t-1})}}\left[u\left(\Pi\left(\mathbf{h}_{t-1}\right),y\right)\right]\quad\forall t\in\mathbb{N}$$

with the sequence of sigma-algebras $\left(\mathcal{F}_t\right)_{t\in\mathbb{N}}$ generated by the stochastic process of histories $\left(\mathbf{h}_t\right)_{t\in\mathbb{N}}$. The stochastic process is not i.i.d., as previous utility realizations affect current period choices. Nevertheless it is a martingale difference sequence, as u is continuous in y on the compact Y, so $\mathbb{E}\left[\left|\mathbf{X}_t\right|\right]\leq 2\max_{a,y}\left|u\left(a,y\right)\right|<\infty$ and $\mathbb{E}\left[\mathbf{X}_t|\mathcal{F}_{t-1}\right]=0$ by equation (8). A fortiori, $\left(\mathbf{X}_t\right)_{t\in\mathbb{N}}$ is a mixingale difference sequence, and by the strong law of large numbers for mixingale sequences (see Theorem 2.7 in

Hall and Heyde, 2014 for the version that applies here), we have

$$\lim_{n\to\infty}\frac{\sum_{t=1}^{n}\mathbf{X}_t}{n}=0\qquad\mathbb{P}_{\Pi}\text{-a.s.}$$

so that

$$\begin{aligned}
&\liminf_{t\to\infty}\frac{\sum_{i=1}^{t}u\left(\mathbf{a}_i,\mathbf{y}_i\right)}{t}=\liminf_{t\to\infty}\sum_{i=1}^{t}\frac{\mathbf{X}_i+\mathbb{E}_{p^*_{\mathbf{a}_t}}\left[u\left(\mathbf{a}_t,\cdot\right)\right]}{t}\\
\geq\;&\left(1-\limsup_{t\to\infty}\alpha_t\left(\mathbf{h}_t\right)\left(\underline{A}\left(p^*\right)\right)\right)\left(\max_{\bar{a}\in A}\min_{y\in Y}u\left(\bar{a},y\right)-\frac{\varepsilon}{2}\right)\\
&+\limsup_{t\to\infty}\alpha_t\left(\mathbf{h}_t\right)\left(\underline{A}\left(p^*\right)\right)\min_{a\in a,y\in Y}u\left(a,y\right)\\
\geq\;&\left(1-\sum_{a\in\underline{A}(p^*)}\limsup_{t\to\infty}\alpha_t\left(\mathbf{h}_t\right)\left(a\right)\right)\max_{\bar{a}\in A}\min_{y\in Y}u\left(\bar{a},y\right)-\frac{\varepsilon}{2}\\
&+\sum_{a\in\underline{A}(p^*)}\limsup_{t\to\infty}\alpha_t\left(\mathbf{h}_t\right)\left(a\right)\min_{a\in a,y\in Y}u\left(a,y\right)\\
\geq\;&\left(1-|A|\max_{a\in\underline{A}(p^*)}\limsup_{t\to\infty}\alpha_t\left(\mathbf{h}_t\right)\left(a\right)\right)\max_{\bar{a}\in A}\min_{y\in Y}u\left(\bar{a},y\right)-\frac{\varepsilon}{2}\\
&+\left(|A|\max_{a\in\underline{A}(p^*)}\limsup_{t\to\infty}\alpha_t\left(\mathbf{h}_t\right)\left(a\right)\right)\min_{a\in a,y\in Y}u\left(a,y\right)
\end{aligned}$$

and therefore the claim follows from setting

$$\frac{\varepsilon}{2\left(\max_{\bar{a}\in A}\min_{y\in Y}u\left(\bar{a},y\right)-\min_{a\in a,y\in Y}u\left(a,y\right)\right)|A|}=\varphi^*.$$

□

Claim 3. *There exists $\bar{\lambda}\in\mathbb{R}_{++}$ such that if $\lambda\geq\bar{\lambda}$ then for every $p^*\in\hat{P}$, $a'\in\underline{A}\left(p^*\right)$, $\nu\in\Delta\left(Q\right)$, we have $a'\notin BR^{\lambda}\left(\nu\right)$.*

That is, if the agent is sufficiently misspecification concerned, they do not play actions that can perform worse than the safe guarantee.

Proof of the Claim. Observe that if $\underline{A}\left(p^*\right)\neq\emptyset$, then by Assumption 1 (i) for all $q\in Q$, there is $y\in\operatorname{supp}q_{a'}$ with $u\left(a',y\right)\leq\max_{\bar{a}\in A}\min_{y\in Y}u\left(\bar{a},y\right)-\frac{\varepsilon}{2}$. But then the claim follows from Lemma 8. □

Claim 4. *There exists $J \in (0,1)$ such that for every $p^* \in \hat{P}$, $a' \in \underline{A}(p^*)$, $\mu \in \Delta(Q)$, and $\lambda \in \mathbb{R}_+$,*

$$\mu\left(\{q \in Q : R\left(p^*_{a'} || q_{a'}\right) > J\}\right) \leq J \implies a' \notin BR^{\lambda}(\mu). \tag{15}$$

That is, if the beliefs are sufficiently concentrated on the parameters that are close to the true DGP, and under the true DGP a' performs worse than the safe guarantee, a' cannot be chosen regardless of the level of misspecification concern.

Proof of the Claim. Observe that given Claim 3, the statement immediately holds for $\lambda > \bar{\lambda}$. Suppose by contradiction that equation (15) does not hold true. This means that there exists a convergent $(p^*_n, \mu_n, \lambda_n)_{n\in\mathbb{N}} \in \hat{P} \times \Delta(Q) \times \left[0, \bar{\lambda}\right]$ and $a' \in A$ with

$$\mu_n\left(\left\{q \in Q : R\left(p^*_{a',n} || q_{a'}\right) > \frac{1}{n}\right\}\right) \leq \frac{1}{n},\ a' \in \underline{A}\left(p^*_n\right), \text{ and } a' \in BR^{\lambda_n}\left(\mu_n\right).$$

By the lower semicontinuity of R and the fact that $R\left(p_{a'} || q_{a'}\right) = 0$ if and only if $p_{a'} = q_{a'}$, (see, e.g., Lemma 1.4.3 in Dupuis and Ellis (2011)), as well as Lemma 4 this in turn implies that there exists $q \in Q$ with

$$a' \in \underline{A}(q) \text{ and } \mathbb{E}_{q_{a'}}\left[u\left(a', y\right)\right] \geq \max_{\bar{a}\in A} \min_{y\in Y} u\left(\bar{a}, y\right),$$

a contradiction. □

Let $c = \frac{J\varphi^*}{4\bar{\lambda}}$. Take an arbitrary $p^* \in \hat{P}$, If for all $a' \in \underline{A}(p^*)$

$$\limsup_{t\to\infty} \alpha_t\left(\mathbf{h}_t\right)\left(a'\right) = 0 < \varphi^* \qquad \mathbb{P}_{\Pi}\text{-a.s.}$$

ε-safety follows by Claim 2. Suppose by contradiction that there is an action $a' \in \underline{A}(p^*)$ with $\limsup_{t\to\infty} \alpha_t\left(\mathbf{h}_t\right)\left(a'\right) > \varphi^*$. ByClaim 4, it must be the case that $\min_{q\in Q} R\left(p^*_{a'} || q_{a'}\right) \geq J$.

But then, by Lemmas 6 and 7 we have that

$$\liminf_{t\to\infty} \Lambda\left(\mathbf{h}_t\right) \geq \frac{\min_{q\in Q} R\left(p^* || q\right)}{c} \geq \frac{J}{c} = 2\bar{\lambda} \qquad \mathbb{P}_{\Pi}\text{-a.s.}$$

Then by Claim 3, we have that for all $a' \in \underline{A}(p^*)$

$$\limsup_{t\to\infty} \alpha_t(\mathbf{h}_t)(a') = 0 \qquad \mathbb{P}_\Pi\text{-a.s.}$$

a contradiction.

Since Q is compact, for every $\epsilon \in (0,1)$ we can pick $\delta < 0$ such that for all $p^* \in \hat{P}$,

$$\min_{q\in Q} R(p^*||q) < \delta$$

implies that for $q \in Q^\epsilon(a)$

$$\mathbb{E}_{p^*}[u(a',y)] - \min_{p\in\Delta(Y)} \mathbb{E}_p[u(a',y)] - \frac{R(p||q_a)}{\lambda} \leq \frac{\varepsilon}{4} \qquad \forall a' \in A, \forall \lambda \in [0,2\delta].$$

But this is ε-consistent with this δ by Lemmas 6, 7, and Berk (1966), page 54.

We show that there is a decision problem (u,a,Y) such that if the concern for misspecification of the agent is such that

$$\Lambda(\mathbf{h}_t) = o\left(\frac{LLR(\mathbf{h}_t,Q)}{t}\right) \qquad \mathbb{P}_\Pi\text{-a.s.}$$

then the decision rule is not $\frac{1}{10}$-safe. Suppose that

$$A = \{1,-1,0\} \qquad \text{and} \qquad Y = \{-1,1\}.$$

The utility function is $u(a,y) = ay$. Each model q considered by the agent is described by $q_a(1)$ for some arbitrary $a \in A$. With this, let $Q = \{0.9, 0.4\}$, $p_a^*(1) = 0.6$, and

$$\mu(0.9) = \frac{1}{2} = \mu(0.4).$$

Let $N(Q) = [0,1]$, i.e., the unstructured models include all the action-independent data-generating processes. We have

$$\max_{\bar{a}\in A} \min_{y\in Y} u(\bar{a},y) = \min_{y\in Y} u(0,y) = 0.$$

However, by the Strong Law of Large Numbers it follows that $\mathbb{P}_\Pi$-almost surely

$$\lim_{t\to\infty} \frac{\sum_{\tau=1}^{t} \mathbb{I}_{\{1\}}(\mathbf{y}_\tau)}{t} = 0.6.$$

Therefore, by Lemma 2 we have that

$$\lim_{t\to\infty} \frac{LLR(\mathbf{h}_t, Q)}{t} = R(0.6||0.4) \qquad \mathbb{P}_\Pi\text{-a.s.}$$

and so

$$\lim_{t\to\infty} \Lambda(\mathbf{h}_t) = 0 \qquad \mathbb{P}_\Pi\text{-a.s.}$$

Moreover, for the constant function $\phi(\varepsilon) = \frac{1}{2}$ for all $\varepsilon \in \mathbb{R}_{++}$ the prior is ϕ-positive on Q in the sense of Fudenberg, Lanzani, and Strack (2022a), and by their Lemma 1

$$\mu(0.4|\mathbf{h}_t) \to 1 \qquad \mathbb{P}_\Pi\text{-a.s.}$$

But then by the upperhemicontinuity of $BR^{(\cdot)}(\cdot)$ established in Lemma 4

$$\liminf_{t\to\infty} \frac{\sum_{i=1}^{t} u(\mathbf{a}_i, \mathbf{y}_i)}{t} = -0.2 < 0 = \max_{\bar{a}\in A} \min_{y\in Y} u(\bar{a}, y) \qquad \mathbb{P}_\Pi\text{-a.s.}$$

proving the desired result.

Finally, we show that there is a decision problem (u, a, Y) such that if the concern for misspecification of the agent is such that

$$o(\Lambda(\mathbf{h}_t)) = \frac{LLR(\mathbf{h}_t, Q)}{t} \qquad \mathbb{P}_\Pi\text{-a.s.}$$

then the decision rule is not $\frac{1}{10}$-consistent.

Let $\delta \in (0, 0.4)$ and suppose

$$A = \{1, -1, 0\} \qquad \text{and} \qquad Y = \{-1, 1\}.$$

The utility function is $u(a, y) = ay$. Again, each model p considered by the agent

is described by $p_a(1)$ for some arbitrary $a \in A$. With this, let $Q = \{0.6, 0.4\}$ and $p_a^*(1) = 0.6 + \delta$. Let $N(Q) = [0,1]$, i.e., the unstructured models include all the action-independent data-generating processes.

Let $\bar{\lambda}$ be such that $\{0\} = BR^{\lambda}(\mu)$ for all $\lambda \geq \bar{\lambda}$ and $\mu \in \Delta(Q)$. Such a $\bar{\lambda}$ exists because for $a \in \{-1, 1\}$ and $q \in Q$

$$\lim_{\lambda \to \infty} \min_{p_a \in \Delta(Y)} \mathbb{E}_{p_a}[u(a,y)] + \frac{1}{\lambda} R(p_a || q_a) = -1.$$

Let

$$C_t = \left\{ h_t \in \mathcal{H}_t : p^{h_t}(1) \geq 0.6, R\left(p^{h_t} || 0.6\right) \geq \frac{R(0.6 + \delta || 0.6)}{2} \right\}.$$

For every $h_t \in C_t$, by Lemmas 2, 6, and 7

$$\lim_{t \to \infty} \frac{LLR(h_t, Q)}{t} \geq \frac{R(0.6 + \delta || 0.6)}{2}$$

so that $\Lambda(h_t)$ is diverging to $+\infty$ and it is eventually larger than $\bar{\lambda}$. But by Sanov's Theorem (see, e.g., Theorem 2.2.1 in Dupuis and Ellis, 2011) the set C_t has a probability converging to 1, so the result follows. ■

Proof of Proposition 2. Suppose that a^* is a Λ-limit action. Thus, since for every policy $\Pi \in A^{\mathcal{H}}$

$$\mathbb{P}_{\Pi}[\sup\{\mathbf{t} \colon \mathbf{a_t} \neq a^*\} < \infty] \leq \sum_{t=0}^{\infty} \sum_{h_t \in \mathcal{H}_t} \mathbb{P}_{\Pi}\left[a^* \in BR^{\Lambda(\mathbf{h}_\tau)}(\mu(\cdot|\mathbf{h}_\tau)), \forall \tau \geq t | h_t\right] \mathbb{P}_{\Pi}[h_t],$$

there are a Λ-optimal policy $\tilde{\Pi} \in A^{\mathcal{H}}$, $t \in \mathbb{N}_0$, and $(a^t, y^t) \in \mathcal{H}_t$ with $\mathbb{P}_{\tilde{\Pi}}[(a^t, y^t)] > 0$ such that with positive probability $\tilde{\Pi}$ prescribes a^* after (a^t, y^t) in every future period. Define $\nu = \mu(\cdot|(a^t, y^t))$, and notice that by Assumption 1 (i)

$$\operatorname{supp} \nu = \operatorname{supp} \mu = Q.$$

As the evolution of beliefs and misspecification concern under Π^{a^*}, i.e., the policy that plays a^* in every period, is the same as under $\tilde{\Pi}$ for every history where the

agent continues to play a^*, we have that

$$\begin{aligned} & \mathbb{P}_{\tilde{\Pi}}[a^* = \tilde{\Pi}(\mathbf{h}_\tau) \text{ for all } \tau > t | (a^t, y^t)] > 0 \\ \implies \quad & \mathbb{P}_{\Pi^{a^*}}[a^* \in BR^{\Lambda\left((a^t,y^t),\mathbf{h}_\tau\right)}\left(\nu\left(\cdot|\mathbf{h}_\tau\right)\right) \text{ for all } \tau > 0] > 0. \end{aligned}$$

We now show that if a^* is not a selfconfirming equilibrium, the latter equals zero, which establishes that a^* cannot be a limit action under this way to adjust the misspecification concern. By the strong law of large numbers (see, e.g., Theorem 8.3.5 in Dudley, 2018),

$$\lim_{\tau\to\infty} p_{a^*}^{\mathbf{h}_\tau} = p_{a^*}^* \qquad \mathbb{P}_{\Pi_{a^*}}\text{-a.s.}$$

Therefore, by the assumptions of the proposition and Gibbs' inequality

$$\lim_{\tau\to\infty} \prod_{\tau'=1}^{\tau} p_{\mathbf{a}_{\tau'}}^{\mathbf{h}_\tau}(\mathbf{y}_{\tau'}) = \lim_{t\to\infty} \max_{q\in Q} \prod_{\tau'=1}^{\tau} q_{\mathbf{a}_{\tau'}}(\mathbf{y}_{\tau'}) = \lim_{t\to\infty} \max_{p\in N(Q)} \prod_{\tau'=1}^{\tau} p_{\mathbf{a}_{\tau'}}(\mathbf{y}_{\tau'}) \qquad \mathbb{P}_{\Pi_{a^*}}\text{-a.s.}$$

So, $\lim_{\tau\to\infty} \frac{LLR\left((a^t,y^t),\mathbf{h}_\tau,Q\right)}{\tau} = 0$, $\mathbb{P}_{\Pi_{a^*}}$ almost surely. With this, as by Assumption 1 (ii), the assumptions of Berk (1966), page 54, are satisfied, for every $\varepsilon \in \mathbb{R}_{++}$ we have

$$\lim_{\tau\to\infty} \nu\left(B_\varepsilon\left(\{q \in Q : q_{a^*} = p_{a^*}^*\}\right)|\mathbf{h}_\tau\right) = 1, \qquad \mathbb{P}_{\Pi_{a^*}}\text{-a.s.}$$

and the desired conclusion follows from Lemma 4. ■

Proof of Theorem 1. We start with the preliminary observation that by Lemma 6, $-\ln \tilde{q}_{\mathbf{a}_t}(\mathbf{y}_t) \leq K$ for all $t \in \mathbb{N}$ and $q \in Q$, $\mathbb{P}_{\Pi_{a^*}}$-almost surely. This will allow us to invoke Lemma 7 in all the various cases.

1) Suppose by contradiction that a^* is a Λ-limit action but is not a Berk-Nash equilibrium. Thus, since for every policy $\Pi \in A^{\mathcal{H}}$

$$\mathbb{P}_\Pi\left[\sup\{\mathbf{t} \colon \mathbf{a_t} \neq a^*\} < \infty\right] \leq \sum_{t=0}^{\infty} \sum_{h_t\in\mathcal{H}_t} \mathbb{P}_\Pi\left[a^* \in BR^{\Lambda(\mathbf{h}_\tau)}\left(\mu\left(\cdot|\mathbf{h}_\tau\right)\right), \forall \tau \geq t | h_t\right] \mathbb{P}_\Pi[h_t],$$

there are a Λ-optimal policy $\tilde{\Pi} \in A^{\mathcal{H}}$, $t \in \mathbb{N}_0$, and $(a^t, y^t) \in \mathcal{H}_t$ with $\mathbb{P}_{\tilde{\Pi}}[(a^t, y^t)] > 0$

such that with positive probability $\tilde{\Pi}$ prescribes a^* after (a^t, y^t) in every future period. Define $\nu = \mu(\cdot|(a^t, y^t))$, and notice that by Assumption 1 (i)

$$\operatorname{supp} \nu = \operatorname{supp} \mu = Q.$$

As the evolution of beliefs and misspecification concern under Π^{a^*}, i.e., the policy that plays a^* in every period, is the same as under $\tilde{\Pi}$ for every history where the agent continues to play a^*, we have that

$$\begin{aligned} & \mathbb{P}_{\tilde{\Pi}}[a^* = \tilde{\Pi}(\mathbf{h}_\tau) \text{ for all } \tau > t | (a^t, y^t)] > 0 \\ \implies \quad & \mathbb{P}_{\Pi^{a^*}}[a^* \in BR^{\Lambda\left((a^t,y^t),\mathbf{h}_\tau\right)}\left(\nu\left(\cdot|\mathbf{h}_\tau\right)\right) \text{ for all } \tau > 0] > 0. \end{aligned}$$

We now show that the latter equals zero, which establishes that a^* cannot be a limit action under this way to adjust the misspecification concern.

Since Y is a compact metric spave, it is separable (see, e.g., Proposition 9.24 in Royden and Fitzpatrick, 1988). Thus, by Theorems 1 and 3 in Varadarajan (1958), $\lim_{\tau\to\infty} p^{\mathbf{h}_\tau}_{a^*} = p^*_{a^*}$, $\mathbb{P}_{\Pi_{a^*}}$-a.s. Then, by Lemma 7 and equation (1.4) we have

$$\lim_{\tau\to\infty} \Lambda\left((a^t, y^t), \mathbf{h}_\tau\right) = 0 \quad \mathbb{P}_{\Pi_{a^*}}\text{-a.s.}$$

By Assumption 1 (ii), the assumptions of Berk (1966), page 54, are satisfied, and we have that for every $\varepsilon \in \mathbb{R}_{++}$,

$$\nu\left(Q^\varepsilon\left(a^*\right)|\mathbf{h}_\tau\right) \to 1, \qquad \mathbb{P}_{\Pi_{a^*}}\text{-a.s.}$$

Therefore, since Q is compact, $\left(\Lambda\left((a^t, y^t), \mathbf{h}_\tau\right), \nu\left(\cdot|\mathbf{h}_\tau\right)\right)_{\tau\in\mathbb{N}}$ admits $\mathbb{P}_{\Pi_{a^*}}$ almost surely a subsequence convergent to $(0, \nu^*)$ for some $\nu^* \in \Delta\left(Q\left(a^*\right)\right)$. With this, the result follows from Lemma 4.

2) Suppose by contradiction that

$$a^* \notin BR^{Meu}\left(\left\{p \in \Delta\left(Y\right)^A : \exists q \in Q, \forall a \in A, q_a \gg p_a\right\}\right).$$

and that a^* is a Λ-limit action. Thus, since for every policy $\Pi \in A^{\mathcal{H}}$

$$\mathbb{P}_{\Pi}\left[\sup\{t\colon \mathbf{a}_t \neq a^*\} < \infty\right] \leq \sum_{t=0}^{\infty} \sum_{h_t \in \mathcal{H}_t} \mathbb{P}_{\Pi}\left[a^* \in BR^{\Lambda(\mathbf{h}_\tau)}\left(\mu\left(\cdot|\mathbf{h}_\tau\right)\right), \forall \tau \geq t | h_t\right] \mathbb{P}_{\Pi}[h_t],$$

there are a policy $\tilde{\Pi} \in A^{\mathcal{H}}$ that is optimal given the adjustment of misspecification given by $\Lambda(\cdot)$, $t \in \mathbb{N}_0$, and $(a^t, y^t) \in \mathcal{H}_t$ with $\mathbb{P}_{\tilde{\Pi}}[(a^t, y^t)] > 0$ such that with positive probability $\tilde{\Pi}$ prescribes a^* after (a^t, y^t) in every future period. Define $\nu = \mu(\cdot|(a^t, y^t))$, and notice that by Assumption 1 (i)

$$\operatorname{supp} \nu = \operatorname{supp} \mu = Q.$$

As the evolution of beliefs and misspecification concern under Π^{a^*}, i.e., the policy that plays a^* in every period, is the same as under $\tilde{\Pi}$ for every history where the agent continues to play a^*, we have that

$$\begin{aligned} & \mathbb{P}_{\tilde{\Pi}}[a^* = \tilde{\Pi}(\mathbf{h}_\tau) \text{ for all } \tau > t | (a^t, y^t)] > 0 \\ \implies \quad & \mathbb{P}_{\Pi^{a^*}}[a^* \in BR^{\Lambda\left((a^t, y^t), \mathbf{h}_\tau\right)}\left(\nu\left(\cdot|\mathbf{h}_\tau\right)\right) \text{ for all } \tau > 0] > 0. \end{aligned}$$

We now show that the latter equals zero, which establishes that a^* cannot be a limit action under this way to adjust the misspecification concern.

By Theorems 1 and 3 in Varadarajan (1958),

$$\lim_{\tau \to \infty} p_{a^*}^{\mathbf{h}_\tau} = p_{a^*}^* \qquad \mathbb{P}_{\Pi_{a^*}}\text{-a.s.}$$

Then, by Lemmas 2 and 7, and equation (1.4), we have

$$\lim_{\tau \to \infty} \Lambda\left((a^t, y^t), \mathbf{h}_\tau\right) = \infty \qquad \mathbb{P}_{\Pi_{a^*}}\text{-a.s.}$$

By Assumption 1 (i) for all $q, q' \in Q$ and $a \in A$ we have

$$q_a \sim q'_a.$$

So we have

$$\left\{p \in \Delta(Y)^A : \exists q \in Q, \forall a \in A, q_a \gg p_a\right\} = \left\{p \in \Delta(Y)^A : \forall q \in Q, \forall a \in A, q_a \gg p_a\right\}.$$

Therefore, by Lemma 8 for all $a \in A$ we have that $\mathbb{P}_{\Pi_{a^*}}$ almost surely

$$\limsup_{\tau\to\infty} \min_{q\in Q} \min_{p_a\in\Delta(Y)} \int_Y u(a,y)\,\mathrm{d}p_a + \Lambda\left((a^t,y^t),\mathbf{h}_\tau\right) R(p_a||q_a) = \min_{y\in\cup_{q\in Q}\mathrm{supp}q_a} u(a,y).$$

But since by Assumption 1 (i) for all $\tau \in \mathbb{N}$, $\mu(\cdot|\mathbf{h}_\tau) \subseteq Q$, $\mathbb{P}_{\Pi_{a^*}}$ almost surely, we have

$$\lim_{\tau\to\infty} \mathbb{E}_{\mu(\cdot|\mathbf{h}_\tau)}\left[\min_{p_a\in\Delta(Y)} \int_Y u(a,y)\,\mathrm{d}p_a + \Lambda\left((a^t,y^t),\mathbf{h}_\tau\right) R(p_a||q_a)\right] = \min_{y\in\cup_{q\in Q}\mathrm{supp}q_a} u(a,y) \quad \mathbb{P}_{\Pi_{a^*}}\text{-a.s.}$$

With this, the result follows from the finiteness of the action space.

3) It follows from the more general Theorem 2. ■

Lemma 9. *For every $c \in \mathbb{R}_{++}$ the function $\alpha \mapsto \min_{q\in Q} \sum_{a\in A} \alpha(a) R(p_a^*||q_a)/c$ is continuous and the correspondence $Q(\cdot) : \Delta(A) \to 2^Q$ is upper hemicontinuous.*

Proof. We first show that the function

$$\begin{aligned} \Delta(A) \times Q &\to \mathbb{R} \\ (\alpha, q) &\mapsto \textstyle\sum_{a\in A} \alpha(a) R(p_a^*||q_a)/c \end{aligned} \tag{16}$$

is continuous. Fix an $a \in A$ and let $(q_n)_{n\in\mathbb{N}} \in Q^{\mathbb{N}}$ be a sequence that converges to $q \in Q$. By Assumption 1 (ii), $\tilde{q}_{a,n}$ is converging pointwise to $\tilde{q}_a$. Then

$$|R(p_a^*||q_{a,n}) - R(p_a^*||q_a)| = \left|\int_Y \log\left(\frac{\tilde{q}_a(y)}{\tilde{q}_{a,n}(y)}\right) \mathrm{d}p_a^*(y)\right|$$

and observe that the integrand on the right-hand side is dominated by a constant by Assumption 1 (i). Therefore, by the dominated convergence theorem $|R(p_a^*||q_{a,n}) - R(p_a^*||q_a)|$ converges to 0. Since A is finite and the function in equation (16) is linear in α, we have obtained the desired continuity. With this, the statement follows from

Theorem 17.31 in Aliprantis and Border (2013).

Proof of Proposition 3. Consider the following three-player game. The action sets are $A_1 = \Delta(A)$, $A_2 = \Delta(Q)$, $A_3 = \mathbb{R}_+$ with arbitrary elements denoted as α, ν, λ. The utility functions are

$$\begin{aligned}
U_1(\alpha,\nu,\lambda) &= \begin{cases} \sum_{a\in A}\alpha(a)\int_Q \min_{p_a\in\Delta(Y)}\left(\mathbb{E}_{p_a}[u(a,y)] + \frac{R(p_a||q_a)}{\lambda}\right)\mathrm{d}\nu(q) & \lambda\neq 0 \\ \sum_{a\in A}\alpha(a)\int_Q \mathbb{E}_{q_a}[u(a,y)]\,\mathrm{d}\nu(q) & \lambda = 0, \end{cases} \\
U_2(\alpha,\nu,\lambda) &= -\int_Q \sum_{a\in A}\alpha(a)\,R(p_a^*||q_a)\,\mathrm{d}\nu(q), \\
U_3(\alpha,\nu,\lambda) &= -\left(\lambda - \min_{q\in Q}\sum_{a\in A}\alpha(a)\,R(p_a^*||q_a)/c\right)^2.
\end{aligned}$$

Observe that for the purpose of finding the equilibria of this game, it is without loss of generality to limit the actions of player 3 to $\left[0,\bar{\lambda}\right]$ with

$$\begin{aligned}
\bar{\lambda}: &= \max_{\alpha\in\Delta(A)}\min_{q\in Q}\sum_{a\in A}\alpha(a)\,R(p_a^*||q_a)/c \\
&= \frac{\max_{\alpha\in\Delta(A)}\min_{q\in Q}\sum_{a\in A}\alpha(a)\int_Y -\log(\tilde{q}_a(y))\,\mathrm{d}p_a^*(y)}{c} < \infty,
\end{aligned}$$

where the inequality holds by Assumption 1 (i). Therefore, since the compactness of Q implies that also $\Delta(Q)$ is compact by Theorem 15.11 in Aliprantis and Border (2013) all the three action sets are compact. Moreover, they are clearly convex.

The utility function U_1 is jointly continuous in its second and third argument by Lemma 4. Moreover, U_2 is trivially continuous in its first and third argument and U_3 is continuous in its first and second argument by Lemma 9. Therefore the game is better-reply secure (see Reny, 1999, page 1033). Moreover, U_1 and U_2 are respectively linear in A_1 and A_2 while U_3 is concave in A_3.

Therefore, by Theorem 3.1 and Footnote 8 in Reny (1999) this game admits a pure-strategy equilibrium $(\alpha^*,\nu^*,\lambda^*)$. But observe that

$$\lambda^* \in \mathrm{argmax}_{\lambda\in\mathbb{R}_+} -\left(\lambda - \min_{q\in Q}\sum_{a\in A}\alpha^*(a)\,R(p_a^*||q_a)/c\right)^2 \implies \lambda^* = \frac{\min_{q\in Q}\sum_{a\in A}\alpha^*(a)\,R(p_a^*||q_a)}{c},$$

$$\alpha^* \in \text{argmax}_{\alpha\in\Delta(A)} U_1(\alpha, \nu^*, \lambda^*) \implies \alpha^* \in \Delta\left(BR^{\lambda^*}(\nu^*)\right),$$

and

$$\begin{aligned} \nu^* &\in \text{argmax}_{\nu\in\Delta(Q)} - \int_Q \sum_{a\in A} \alpha^*(a) R(p_a^* || q_a) \, \mathrm{d}\nu(q) \\ \implies \text{supp}\nu^* &\subseteq \text{argmin}_{q\in Q} \sum_{a\in A} \alpha^*(a) R(p_a^* || q_a) \implies \nu^* \in \Delta(Q(\alpha^*)). \end{aligned}$$

Therefore, α^* is a mixed c-robust equilibrium sustained by the belief ν^* and the concern for misspecification λ^*. ■

Proof of Theorem 2. We start with the preliminary observation that by Lemma 6, $-\ln \tilde{q}_{\mathbf{a}_t}(\mathbf{y}_t) \leq K$ for all $t \in \mathbb{N}$ and $q \in Q$, $\mathbb{P}_\Pi$-almost surely. This will allow us to invoke Lemma 7. Observe that $(\alpha_t)_{t\in\mathbb{N}}$ satisfies the following differential inclusion: for all $a \in A$, $t \in \mathbb{N}$, $h_t \in \mathcal{H}_t$, and $h_{t+1} \in \mathcal{H}_{t+1}$ such that $h_{t+1} \succ h_t$

$$\alpha_{t+1}(h_{t+1})(a) \in \left\{ \alpha_t(h_t)(a) + \frac{1}{t+1}\left(\mathbb{I}_{\{a'\}}(a) - \alpha_t(h_t)(a)\right) : a' \in BR^{\Lambda(h_t)}(\mu(\cdot|h_t)) \right\}.$$

Set $\tau_0 = 0$ and $\tau_t = \sum_{i=1}^t \frac{1}{i}$ for all $t \in \mathbb{N}$. The continuous-time interpolation of α_t is the function $w : \mathbb{R}_+ \to \Delta(A)$

$$w(\tau_t + l) = \begin{cases} \alpha_t + l\frac{\alpha_{t+1}-\alpha_t}{\tau_{t+1}-\tau_t}, & \forall t \in \mathbb{N}, \forall l \in \left[0, \frac{1}{t+1}\right] \\ \alpha_1 & t = 0, \forall l \in [0,1]. \end{cases} \tag{17}$$

For every $\alpha \in \Delta(A)$, let

$$\chi_\alpha = \left\{\alpha' \in \Delta\left(BR^{\min_{q\in Q}\sum_{a\in A}\alpha(a)R(p_a^*||q_a)/c}(\Delta(Q(\alpha)))\right)\right\} \subseteq \Delta(A).$$

We use the theory of stochastic approximation for differential inclusions (Benaim, Hofbauer, and Sorin, 2005 and Esponda, Pouzo, and Yamamoto, 2021a) to show that (17) can be approximated by a solution to

$$\dot{\underline{\alpha}}_t \in \chi_{\underline{\alpha}_t} - \underline{\alpha}_t. \tag{18}$$

A solution over $[0,T]$, $T \in \mathbb{R}_{++}$, to the differential inclusion (18) with initial point $\hat{\alpha} \in \Delta(A)$ is a mapping $\underline{\alpha}_{(\cdot)} : [0,T] \to \Delta(A)$ that is absolutely continuous over compact intervals such that $\underline{\alpha}_0 = \hat{\alpha}$ and (18) is satisfied for almost every t. Let $S^T_{\hat{\alpha}}$ be the set of the solutions to (18) over $[0,T]$, $T \in \mathbb{R}_{++}$, with initial conditions $\hat{\alpha} \in \Delta(A)$. Since by Lemmas 9 and 4 $\alpha \mapsto \min_{q \in Q} \sum_{a \in A} \alpha(a) R(p^*_a || q_a)/c$ is continuous and $Q(\cdot)$, $BR^{(\cdot)}(\cdot)$ are upper hemicontinuous,

$$\alpha \mapsto \Delta\left(BR^{\min_{q \in Q} \sum_{a \in A} \alpha(a) R(p^*_a || q_a)/c}\left(\Delta\left(Q\left(\alpha\right)\right)\right)\right) \tag{19}$$

is upper hemicontinuous. To see this, we show that it has a closed graph. Since $\Delta(A)$ is compact, this is enough by, e.g., Proposition E.3 in Ok (2011). Let $(\alpha_n, \alpha'_n)_{n \in \mathbb{N}} \in (\Delta(A) \times \Delta(A))^{\mathbb{N}}$ be such that

$$\alpha_n \in \Delta\left(BR^{\min_{q \in Q} \sum_{a \in A} \alpha'_n(a) R(p^*_a || q_a)/c}\left(\Delta\left(Q\left(\alpha'_n\right)\right)\right)\right) \qquad \forall n \in \mathbb{N}$$

and $\lim_{n \to \infty} (\alpha_n, \alpha'_n) = (\alpha, \alpha')$. By finiteness of A, we can take (possibly truncating some initial elements of the sequence) $(\alpha_n, \alpha'_n)_{n \in \mathbb{N}}$ to be such that $\text{supp}\alpha \subseteq \text{supp}\alpha_n$ for all $n \in \mathbb{N}$. Then for every $\bar{a} \in \text{supp}\alpha$ there is $(\nu^{\bar{a}}_n)_{n \in \mathbb{N}}$ such that $\nu^{\bar{a}}_n \in \Delta(Q(\alpha'_n))$ and

$$\bar{a} \in BR^{\min_{q \in Q} \sum_{a \in A} \alpha'_n(a) R(p^*_a || q_a)/c}\left(\nu^{\bar{a}}_n\right) \qquad \forall n \in \mathbb{N}.$$

Since Q is compact so is $\Delta(Q)$ by Theorem 15.11 in Aliprantis and Border (2013), and by restricting to a subsequence we can take $\nu^{\bar{a}}_n$ to be convergent to some $\nu^{\bar{a}} \in \Delta(Q)$. Since $Q(\cdot)$ is upper hemicontinuous, $\nu^{\bar{a}} \in Q(\alpha')$. Since $BR^{(\cdot)}(\cdot)$ is upper hemicontinuous,

$$\bar{a} \in BR^{\min_{q \in Q} \sum_{a \in A} \alpha'(a) R(p^*_a || q_a)/c}\left(\nu^{\bar{a}}\right) \subseteq BR^{\min_{q \in Q} \sum_{a \in A} \alpha'(a) R(p^*_a || q_a)/c}\left(Q\left(\alpha'\right)\right)$$

showing that (α, α') belongs to the graph of correspondence (19). Therefore, as χ_α is also convex- and closed-valued, a solution to (18) exists by Theorem 2.1.4 in Aubin and Cellina (2012), i.e., $S^T_{\hat{\alpha}}$ is nonempty for every $T \in \mathbb{R}_{++}$ and $\hat{\alpha} \in \Delta(A)$. Let $S^T = \cup_{\hat{\alpha} \in \Delta(A)} S^T_{\hat{\alpha}}$.

We next establish that the continuous-time interpolation of $(\alpha_t(\mathbf{h}_t))_{t\in\mathbb{N}}$ defined in (17) can in the long run be approximated arbitrarily well by a solution to (18). Observe that w is Lipschitz continuous of order 1 as for all history sequences $(h_t)_{t\in\mathbb{N}} \in \times_{t\in\mathbb{N}}\mathcal{H}_t$ with $h_{t+1} \succ h_t$ for all $t \in \mathbb{N}$,

$$\frac{||\alpha_{t+1}(h_{t+1}) - \alpha_t(h_t)||_\infty}{\tau_{t+1} - \tau_t} \leq \frac{1/(t+1)}{1/(t+1)} = 1 \qquad \forall t \in \mathbb{N}. \tag{20}$$

Therefore w is absolutely continuous (see, e.g., Proposition 7 in Royden and Fitzpatrick, 1988), and α_t is uniformly bounded because it takes values in $\Delta(A)$. Let $\Upsilon = \{\alpha - \alpha' : \alpha, \alpha' \in \Delta(A)\}$ and for all $\varepsilon \in \mathbb{R}_+$ and $\alpha' \in \Delta(A)$,

$$M_\varepsilon(\alpha') = \left\{\nu \in \Delta(Q) : \int_Q \sum_{a\in A} \alpha'(a) R(p_a^*||q_a)\, \mathrm{d}\nu(q) \leq \varepsilon + \min_{q\in Q} \sum_{a\in A} \alpha'(a) R(p_a^*||q_a)\right\}.$$

By Esponda, Pouzo, and Yamamoto, 2021a, Part 1a of the proof of Theorem 2 (observe that Assumption 1 (i-ii) implies their Assumption 2 (ii-iii), except for the fact that we do not require finite-dimensionality of Y and Q. It is readily checked that since these two sets are still assumed to be compact, this does not create any issue in the proof of their Theorem 2), $M_{(\cdot)}(\cdot)$ is upper hemicontinuous. We define

$$F : \mathbb{R}_+ \times \mathbb{R}_+ \times \Delta(A) \rightrightarrows \Upsilon$$

by

$$F(\varepsilon, \varepsilon', \alpha) = \left\{\iota \in \left[\Delta\left(\cup_{\lambda' \in B_{\varepsilon'}\left(\min_{q\in Q}\sum_{a\in A}\alpha(a)R(p_a^*||q_a)/c\right)\cap\left[0,\frac{2K}{c}\right]} BR^{\lambda'}(\Delta(M_\varepsilon(\alpha)))\right) - \alpha\right]\right\}.$$

Observe that $F(0,0,\alpha) = \chi_\alpha - \alpha$. Moreover, we now show that F has a closed graph, so it is upper hemicontinuous. Let

$$(\iota_n, \varepsilon_n, \varepsilon_n', \alpha_n)_{n\in\mathbb{N}} \in (\Upsilon \times \mathbb{R}_+ \times \mathbb{R}_+ \times \Delta(A))^{\mathbb{N}}$$

be such that $\iota_n \in F(\varepsilon_n, \varepsilon'_n, \alpha_n)$ for all $n \in \mathbb{N}$ and

$$\lim_{n\to\infty} (\iota_n, \varepsilon_n, \varepsilon'_n, \alpha_n) = (\iota, \varepsilon, \varepsilon', \alpha).$$

Since A is finite, it is without loss of generality to take $\iota_n(a) > -\alpha_n(a)$ for all $n \in \mathbb{N}$ and for all a for which $\iota(a) > -\alpha(a)$. Then for all $\hat{a}$ such that $\iota(\hat{a}) > -\alpha(\hat{a})$, there is a sequence $\left(\nu_n^{\hat{a}}, \lambda_n^{\hat{a}}\right)_{n\in\mathbb{N}} \in (\Delta(Q) \times [0, 2K/c])^{\mathbb{N}}$ such that

$$\nu_n^{\hat{a}} \in M_{\varepsilon_n}(\alpha'_n),\ \lambda_n^{\hat{a}} \in B_{\varepsilon'_n}\left(\min_{q\in Q} \sum_{a\in A} \alpha_n(a) R(p_a^* || q_a)/c\right), \text{ and } \hat{a} \in BR^{\lambda_n^{\hat{a}}}\left(\nu_n^{\hat{a}}\right).$$

Since $\Delta(Q)$ and $[0, 2K/c]$ are compact by restricting to a subsequence we can take $\left(\nu_n^{\hat{a}}, \lambda_n^{\hat{a}}\right)_{n\in\mathbb{N}}$ to be convergent to some $\left(\nu^{\hat{a}}, \lambda^{\hat{a}}\right) \in \Delta(Q) \times [0, 2K/c]$. Since $M_{(\cdot)}(\cdot)$ is upper hemicontinuous $\nu^{\hat{a}} \in M_\varepsilon(\alpha)$. Since

$$\hat{\alpha} \mapsto \min_{q\in Q} \sum_{a\in A} \hat{\alpha}(a) R(p_a^* || q_a)/c$$

is continuous by Lemma 9,

$$\lambda^{\hat{a}} \in B_{\varepsilon'}\left(\min_{q\in Q} \sum_{a\in A} \alpha(a) R(p_a^* || q_a)/c\right).$$

Since $BR^{(\cdot)}(\cdot)$ is upper hemicontinuous,

$$\hat{a} \in BR^{\lambda^{\hat{a}}}\left(\nu^{\hat{a}}\right) \subseteq \left\{BR^{\hat{\lambda}}(\hat{\nu}) : \hat{\nu} \in M_\varepsilon(\alpha), \hat{\lambda} \in B_{\varepsilon'}\left(\min_{q\in Q} \sum_{a\in A} \alpha(a) R(p_a^* || q_a)/c\right) \cap \left[0, \frac{2K}{c}\right]\right\}$$

showing that $(\iota, \varepsilon, \varepsilon', \alpha)$ belongs to the graph of the correspondence (19).

With this, $F(0, 0, \cdot) + (\cdot) : \Delta(A) \to \Delta(A)$ satisfies Hypothesis 1.1 in Benaim, Hofbauer, and Sorin (2005), as it is clearly compact- and convex-valued. Moreover, by Theorem 1 in Esponda, Pouzo, and Yamamoto (2021a) and Lemma 7 we have that $\mathbb{P}_\Pi$-almost surely, if $\lim_{t\to\infty} \alpha_t(\mathbf{h}_t) = \alpha^*$, we eventually have $\mu(\cdot|\mathbf{h}_t) \in M_\varepsilon(\alpha^*)$

and

$$\Lambda(\mathbf{h}_t) \in B_{\varepsilon'}\left(\min_{q\in Q}\sum_{a\in A}\alpha^*(a)\,R(p_a^*||q_a)/c\right)$$

for all $(\varepsilon,\varepsilon')\in\mathbb{R}^2_{++}$. Thus, there is a sequence $(\hat{\varepsilon}_t)_{t\in\mathbb{N}}\in\mathbb{R}^{\mathbb{N}}_{++}$ converging to 0 with $\chi_{\alpha_t}-\alpha_t\in F(\hat{\varepsilon}_t,\hat{\varepsilon}_t,\alpha_t)$.

Fix $T\in\mathbb{N}$ and define the flow operator $G: C(\mathbb{R},\Delta(A))\times\mathbb{R}\to C(\mathbb{R},\Delta(A))$ as

$$G^t(f)(s)=f(s+t)\qquad \forall f\in C(\mathbb{R},\Delta(A)),\forall s\in\mathbb{R},\forall t\in\mathbb{R}.$$

We now show that every limit point of $(G^t(w))_{t\in\mathbb{N}}$ is in S^T. This argument borrows extensively from the proofs Theorem 4.2 in Benaim, Hofbauer, and Sorin (2005) and Theorem 2 in Esponda, Pouzo, and Yamamoto (2021a). However they cannot be directly applied, because the interpolated process w we consider is not a perturbed solution in the sense of Benaim, Hofbauer, and Sorin (2005). Indeed, it may not be possible to find an α that *jointly* justifies a_t as a best reply to beliefs in $Q(\alpha)$ and the concern for misspecification $\min_{q\in Q}\sum_{a\in A}\alpha(a)\,R(p_a^*||q_a)/c$, as perturbation of the empirical frequency α_{t-1} in different directions may be needed for the concern and the belief. Nevertheless, the core of their arguments can be adapted leveraging the upper-hemicontinuity of F established above.

Since w is uniformly continuous by equation (20), the family $(G^t(w))_{t\in\mathbb{N}}$ is equicontinuous, and thus it is relatively compact in the topology of uniform convergence over compact sets by the Arzela-Ascoli theorem (see Willard, 2012 Theorem 43.15 for the version with a noncompact domain). The topology of uniform convergence over compact sets is metrizable since $\Delta(A)$ is metrizable and $\mathbb{R}$ is open (see Theorem 1.14b in Simon, 2020), and so there exists a limit point $z=\lim_{t_n}G^{t_n}(w)$. Define

$$m(t)=\max\{k\in\mathbb{N}:\tau_k\le t\}$$

and for all $s\in\mathbb{R}$, $v(s)=w'(s)=\alpha_{m(s)+1}-\alpha_{m(s)}\in F\left(\hat{\varepsilon}_{m(s)},\hat{\varepsilon}_{m(s)},\alpha_{m(s)}\right)$, and

$v_n(s) = v(t_n + s)$ so

$$\begin{aligned} z(T) - z(0) &= \lim_{t_n} \left(G^{t_n}(w)(T) - G^{t_n}(w)(0) \right) \\ &= \lim_{t_n} \left(w(T + t_n) - w(t_n) \right) = \lim_{n \to \infty} \int_0^T v_n(s) \, \mathrm{d}s. \end{aligned}$$

Since $(v_n)_{n \in \mathbb{N}}$ is uniformly bounded, it is bounded in $L^2\left([0,T], \mathbb{R}^A, Leb\right)$. By the Banach-Alaoglu theorem (see Theorem 6.21 in Aliprantis and Border, 2013), (by restricting to a subsequence) we can take $(v_n)_{n \in \mathbb{N}}$ to be a weakly-convergent subsequence with limit $v^* \in L^2\left([0,T], \mathbb{R}^A, Leb\right)$. By Mazur's lemma (see Corollary V.3.14 in Dunford and Schwartz (1988)), there exist a function $N : \mathbb{N} \to \mathbb{N}$ and a sequence of positive weights $\left(\rho_n(n), ..., \rho_{N(n)}(n)\right)_{n \in \mathbb{N}}$ with $\sum_{i=n}^{N(n)} \rho_i(n) = 1$ for all $n \in \mathbb{N}$ such that if we define

$$\bar{v}_n = \sum_{i=n}^{N(n)} \rho_i(n) \, v_i,$$

then $\bar{v}_n$ converges with respect to the $L^2\left([0,T], \mathbb{R}^A, Leb\right)$ norm, and thus almost surely, to v^*.

Let $\tau \in [0,T]$ be such that the convex combination of the elements of $(v_n)_{n \in \mathbb{N}}$ is converging to v^* at τ. For every $t \in [0,T]$ and $n \in \mathbb{N}$, define

$$\gamma_n(t) = \hat{\varepsilon}_{m(t_n + t)} + ||w(t_n + t) - \alpha_{m(t_n + t)}||$$

and

$$w_n(t) = w(t_n + t).$$

Observe that by definition of w, $(\hat{\varepsilon}_t)_{t \in \mathbb{N}}$, and z,

$$\lim_{n \to \infty} \gamma_n(t) = 0 \text{ and } \lim_{n \to \infty} w_n(t) = z(t).$$

But then, by the upperhemicontinuity of F, for every $\varepsilon \in \mathbb{R}_{++}$ there exists N_ε such that for $n \geq N_\varepsilon$, $F(\gamma_n(t), \gamma_n(t), w_n(t)) \subseteq B_\varepsilon(F(0,0,z(t)))$, where the latter set is

closed and convex. But since $v_n(t) \in F(\gamma_n(t), \gamma_n(t), w_n(t))$, also

$$\bar{v}_n(t) \in F(\gamma_n(t), \gamma_n(t), w_n(t)) \subseteq B_\varepsilon(F(0,0,z(t))).$$

Therefore, $v^* \in (F(0,0,z(t)))$. Since the fact that v_n is weakly convergent to v^* implies by definition that

$$\lim_{n\to\infty} \int_0^T v_n(s)\,\mathrm{d}s = \lim_{n\to\infty} \int_0^T v^*(s)\,\mathrm{d}s$$

we have that $z \in S^T$.

Therefore, by (ii) $\Longrightarrow$ (i) of Theorem 4.1 in Benaim, Hofbauer, and Sorin (2005) (see Esponda, Pouzo, and Yamamoto, 2021b for the slightly corrected version used here)

$$\lim_{t\to\infty} \inf_{\tilde{\boldsymbol{\alpha}}\in S^T} \sup_{0\le s\le T} ||w(t+s) - \tilde{\boldsymbol{\alpha}}_s|| = 0 \quad \mathbb{P}_\Pi\text{-a.s. for all } T \in \mathbb{N}. \tag{21}$$

With this, we can replicate an argument from Fudenberg, Lanzani, and Strack (2022b) to rule out convergence to non equilibria. If $\alpha^* \in \Delta(A)$ is not a mixed c-robust equilibrium, there is $a \in A$ with $\alpha^*(a) > 0$ and $\delta_a \notin \chi_{\alpha^*}$. Since $\chi_{(\cdot)} = F(0,0,\cdot) + (\cdot)$ and F has a closed graph and maps into the compact Υ, there exists $D \in \mathbb{R}_{++}$ such that for all $\alpha' \in B_D(\alpha^*)$, $\alpha'(a) - \max_{\hat{\alpha}\in\chi_{\alpha'}} \hat{\alpha}(a) > \alpha^*(a)/2$. Therefore, for every initial condition $\bar{\alpha} \in B_D(\alpha^*)$ and every solution of (18), $\underline{\alpha}(a)$ decreases at rate at least $\alpha^*(a)/4$ until it leaves $B_D(\alpha^*)$. So for every initial condition $\bar{\alpha} \in B_D(\alpha^*)$ and every solution, the differential inclusion leaves $B_D(\alpha^*)$ before time

$$T^* :\, = \frac{D + \alpha^*(a)}{\alpha^*(a)/4}.$$

With this, we can prove that $(\alpha_t(\mathbf{h}_t))_{t\in\mathbb{N}}$ does not converge to α^* on a sample path on which the convergence of equation (21) happens. Since the set of such sample paths has probability 1 under policy Π, this fact concludes the proof. Suppose by contradiction that on one of such paths $(\alpha_t(h_t))_{t\in\mathbb{N}}$ converges to α^*. Therefore, we

can choose $\hat{T} \in \mathbb{N}$ such that on that sample path $\alpha_t(h_t) \in B_{D/2}(\alpha^*)$ for all $t \geq \hat{T}$ and

$$\inf_{\underline{\alpha} \in S^{T^*}} \sup_{0 \leq s \leq T^*} ||w(\hat{T}+s) - \underline{\alpha}_s|| \leq D/4. \tag{22}$$

Take any $\underline{\alpha} \in S^{T^*}$ with $\sup_{0 \leq s \leq T^*} ||w(\hat{T}+s) - \tilde{\boldsymbol{\alpha}}_s|| \leq D/2$. Since $w\left(\hat{T}\right) \in B_{D/2}(\alpha^*)$, $\underline{\alpha} \in S^{T^*}_{\bar{\alpha}}$ for some initial condition $\bar{\alpha} \in B_D(\alpha^*)$. But then by definition of T^* the differential inclusion leaves $B_D(\alpha^*)$ by time $T^* + \hat{T}$, and by (22), $(\alpha_t(\mathbf{h}_t))_{t \in \mathbb{N}}$ does not stay in $B_{D/2}(\alpha^*)$, a contradiction. ■

Proof of Corollary 1. We first show that for a sufficiently low c there is no c-robust equilibrium. Observe that by Assumption 3 (i) and Proposition 8 in Battigalli, Cerreia-Vioglio, Maccheroni, Marinacci, and Sargent (2022) for every $\alpha \in \Delta(A)$, we have

$$Q(\alpha) = \left\{ q^{\left(\theta_0^*, \theta_{1\pi}^*, \theta_{1a}^*, \theta_2^*, \theta_3^*\right)} \right\}. \tag{23}$$

Moreover, since θ^* perfectly predicts the consequences under policy 0, we have

$$\min_{\theta \in \Theta} R\left(p_0^* || q_0^\theta\right) = 0.$$

By Assumption 3 (i) and Lemma 3 in Battigalli, Cerreia-Vioglio, Maccheroni, Marinacci, and Sargent (2022), $BR^{Seu}(\Delta(Q(0))) = \{1\}$, and therefore 0 is not a c-robust equilibrium for any $c \in \mathbb{R}_{++}$. Since f_1 is strictly concave on $\mathbb{R}_{++}$, by Assumption 3 (iii) it follows that $\min_{\theta \in \Theta} R\left(p_1^* || q_1^\theta\right) = R\left(p_1^* || q_1^{\theta^*}\right) > 0$. By Assumption 3 (ii) and Lemma 8 there exists a sufficiently small $\bar{c}$ such that for all $c \leq \bar{c}$,

$$BR^{\frac{\min_{\theta \in \Theta} R\left(p_1^* || q_1^\theta\right)}{c}}(\delta_{\theta^*}) = \{0\}$$

proving that there is no c-robust equilibrium if $c \leq \bar{c}$. That a mixed c-robust equilibrium exists follows from Proposition 3.[34]

[34] To formally invoke Proposition 3, that requires absolute continuity with respect to the true data generating process for all $\theta \in \Theta$, restrict the parameter space to $\{\theta^*\}$. Given equation (23) every mixed c-robust equilibrium with the reduced parameter space remains a mixed c-robust equilibrium with the original Θ.

In particular, the maximal (resp. the minimal) equilibrium is defined as the α such that $\sum_{a\in A}\alpha(a)R\left(p_a^*||q_a^{\theta^*}\right)/c$ is equal to the maximal (resp. minimal) misspecification concern λ such that $1\in BR^\lambda(\delta_{\theta^*})$ (resp. $0\in BR^\lambda(\delta_{\theta^*})$). Since a larger $\theta_{1\pi}^*+\theta_{1a}^*$ makes action 0 more favorable, the comparative statics follows. ■

.1.2 Representation

Preliminaries

Let $B_0(\Sigma)$ denote the set of all real-valued Σ-measurable simple functions endowed with the supnorm. The subset of functions in $B_0(\Sigma)$ that take values in $C\subseteq\mathbb{R}$ is denoted as $B_0(\Sigma,C)$. A functional $I:\Phi\to\mathbb{R}$ defined on a nonempty subset Φ of $B_0(\Sigma)$ is a *niveloid* if for every $\varphi,\psi\in\Phi$

$$I(\varphi)-I(\psi)\leq\sup(\varphi-\psi).$$

It is *translation invariant* if $I(\alpha\varphi+(1-\alpha)k\mathbb{I}_S)=I(\alpha\varphi)+(1-\alpha)k$ for all $\alpha\in[0,1]$, $\varphi\in\Phi$, and $k\in\mathbb{R}$ such that $\alpha\varphi+(1-\alpha)k\mathbb{I}_S$ and $\alpha\varphi$ are in Φ. It is *monotone continuous* if for every $(\varphi_n)_{n\in\mathbb{N}}\in\Phi^{\mathbb{N}}$ such that $\lim_{n\to\infty}\varphi_n=\varphi$ and $\varphi_n\leq\varphi_{n+1}$ for all $n\in\mathbb{N}$ we have $\lim_{n\to\infty}I(\varphi_n)=I(\varphi)$. A niveloid is *normalized* if $I(k\mathbb{I}_S)=k$ for all $k\in\mathbb{R}$ such that $k\mathbb{I}_S\in\Phi$. A function $c:\Delta(S)\to\mathbb{R}_+$ is *grounded* if $c^{-1}(0)\neq\emptyset$. An event is *strongly nonnull* if for every $x,x'\in X$ with $x\succ x'$, we have $x\succ x'Ex$.

Results

Our first lemma shows that the average robust control representation falls in the variational class.

Lemma 10. *Suppose that there exist a nonconstant affine function $u:X\to\mathbb{R}$, a nonempty and finite $Q\subseteq\Delta(S)$, $\mu\in\Delta(Q)$, and $(\lambda_q)_{q\in Q}\in\mathbb{R}_+^Q$ such that for all $f,g\in\mathcal{F}$*

$$f\succsim g\Longleftrightarrow\mathbb{E}_\mu\left[\min_{p\in\Delta(S)}\mathbb{E}_p[u(f)]+\frac{R(p||q)}{\lambda_q}\right]\geq\mathbb{E}_\mu\left[\min_{p\in\Delta(S)}\mathbb{E}_p[u(g)]+\frac{R(p||q)}{\lambda_q}\right]. \quad (24)$$

Then $\succsim$ satisfies Weak Order, Weak Certainty Independence, Continuity, Monotonicity, Uncertainty Aversion, Nondegeneracy, Weak Monotone Continuity, and admits the representation

$$f \succsim g \Longleftrightarrow \min_{p\in\Delta(S)} \int_S \hat{u}(f)\, dp + \hat{c}(p) \geq \min_{p\in\Delta(S)} \int_S \hat{u}(g)\, dp + \hat{c}(p) \tag{25}$$

for some nonconstant affine $\hat{u} : X \to \mathbb{R}$ and a grounded, convex, and lower semicontinuous function $\hat{c} : \Delta(S) \to [0,\infty]$. Moreover, we can choose $\hat{u} = u$ and $\hat{c}$ is such that $\hat{c}^{-1}(0) = \mathbb{E}_\mu[q]$.

Proof. We first observe that without loss of generality we can take u to be such that $0 \in \text{int}u(X)$ in the representation of equation (24). Indeed, since u is nonconstant and affine, there exists $x \in X$ with $u(x) \in \text{int}u(X)$. Define $u'(y) = u(y) - u(x)$ for all $y \in X$. Then, we have

$$\begin{array}{rcl} f & \succsim & g \\ \Longleftrightarrow \mathbb{E}_\mu\left[\min_{p\in\Delta(S)} \mathbb{E}_p[u(f)] + \frac{R(p||q)}{\lambda_q}\right] & \geq & \mathbb{E}_\mu\left[\min_{p\in\Delta(S)} \mathbb{E}_p[u(g)] + \frac{R(p||q)}{\lambda_q}\right] \\ \Longleftrightarrow \mathbb{E}_\mu\left[\min_{p\in\Delta(S)} \mathbb{E}_p[u'(f)] + \frac{R(p||q)}{\lambda_q}\right] & \geq & \mathbb{E}_\mu\left[\min_{p\in\Delta(S)} \mathbb{E}_p[u'(g)] + \frac{R(p||q)}{\lambda_q}\right] \end{array}$$

and $0 \in \text{int}u'(X)$.

Fix $q \in Q$. The functional $I_q : B_0(\Sigma, \mathbb{R}) \to \mathbb{R}$ defined as

$$I_q(\varphi) : = \min_{p\in\Delta(S)} \int_S \varphi(s)\, \mathrm{d}p + \frac{1}{\lambda_q} R(p||q) \qquad \forall \varphi \in B_0(\Sigma, \mathbb{R})$$

is easily seen to be monotone, translation invariant, and concave by Theorem 11.13 in Aliprantis and Border (2013) and the concavity of the minimum. Since Q is finite,

$$\hat{I}(\varphi) = \int_Q I_q(\varphi)\, \mathrm{d}\mu(q) \qquad \forall \varphi \in B_0(\Sigma, \mathbb{R})$$

is well-defined and $\hat{I}$ is monotone, concave, and represents $\succsim$. Let $\varphi \in B_0(\Sigma, u(X))$, $k \in u(X)$, and $\gamma \in (0,1)$. Since u is affine and $0 \in \text{int}u(X)$, we have $\gamma\varphi + (1-\gamma)k \in$

$B_0(\Sigma, u(X))$, $\gamma\varphi \in B_0(\Sigma, u(X))$, and

$$\begin{aligned}\hat{I}(\gamma\varphi + (1-\gamma)k) &= \int_Q I_q(\gamma\varphi + (1-\gamma)k)\,\mathrm{d}\mu(q) = \int_Q I_q(\gamma\varphi) + (1-\gamma)k\mathrm{d}\mu(q) \\ &= \int_Q I_q(\gamma\varphi)\,\mathrm{d}\mu(q) + (1-\gamma)k = \hat{I}(\gamma\varphi) + (1-\gamma)k.\end{aligned}$$

But then, notice that

$$\int_Q \left(\min_{p\in\Delta(S)} \int_S u(f)\,\mathrm{d}p + \frac{1}{\lambda_q} R(p||q)\right)\mathrm{d}\mu(q) = \int_Q I_q(u(f))\,\mathrm{d}\mu(q) = \hat{I}(u(f))$$

where $\hat{I}$ is monotone and translation invariant. Therefore, by Lemma 25 in Maccheroni, Marinacci, and Rustichini (2006a), $\hat{I}$ is a concave niveloid, and it is clearly normalized. With this, by Lemma 28 and Footnote 15 in Maccheroni, Marinacci, and Rustichini (2006a) $\succsim$ satisfies Weak Order, Weak Certainty Independence, Continuity, Monotonicity, Uncertainty Aversion, and Nondegeneracy.

Fix $f, g \in \mathcal{F}$, $x \in X$, and $(E_i)_{i\in\mathbb{N}} \in \Sigma^{\mathbb{N}}$ with $E_1 \supseteq E_2 \supseteq ...$, $\cap_{i\geq 1} E_i = \emptyset$, and $f \succ g$. Then, by Proposition 1.4.2 in Dupuis and Ellis (2011) for all $q \in Q$, $\lim_{i\to\infty} q(E_i) = 0$ and for all $i \in \mathbb{N}$

$$-\exp\left(-\lambda_q\left(I_q\left(x\mathbb{I}_{E_i} + u(f)\,\mathbb{I}_{S\setminus E_i}\right)\right)\right) = -\int_{S\setminus E_i} \exp\left(-\lambda_q u(f(s))\right)\mathrm{d}q(s) - \int_{E_i} \exp\left(-\lambda_q u(x)\right)\mathrm{d}q(s).$$

But then

$$\lim_{i\to\infty} -\exp\left(-\lambda_q\left(I_q\left(x\mathbb{I}_{E_i} + u(f)\,\mathbb{I}_{S\setminus E_i}\right)\right)\right) = \int_S -\exp\left(-\lambda_q u(f(s))\right)\mathrm{d}q(s) > \int_S -e^{-\lambda_q u(g(s))}\mathrm{d}q(s)$$

that is

$$\lim_{i\to\infty} I_q\left(x\mathbb{I}_{E_i} + u(f)\,\mathbb{I}_{S\setminus E_i}\right) > \frac{-\log\left(\int_S \exp\left(-\lambda_q u(g(s))\right)\mathrm{d}q(s)\right)}{\lambda_q}$$

proving that there exists $i \in \mathbb{N}$ such that $I_q\left(u(x)\,\mathbb{I}_{E_i} + u(f)\,\mathbb{I}_{S\setminus E_i}\right) > I_q(u(g))$. Since the statement holds for every $q \in Q$ and Q is finite, there exists $i \in \mathbb{N}$ such that $\hat{I}\left(u(x)\,\mathbb{I}_{E_i} + u(f)\,\mathbb{I}_{S\setminus E_i}\right) > \hat{I}(u(g))$ proving that $\succsim$ satisfies Weak Monotone

Continuity.

By Theorem 3 and Lemma 30 in Maccheroni, Marinacci, and Rustichini (2006a) it admits the representation in equation (25).

By the first part of the lemma we have

$$u(x) \geq u(x') \iff x \succsim x' \iff \hat{u}(x) \geq \hat{u}(x')$$

and therefore by the uniqueness up to a positive affine transformation of $\hat{u}$ guaranteed by Corollary 5 in Maccheroni, Marinacci, and Rustichini (2006a) and the fact that every two affine functions that represent $\succsim$ on X are positive affine transformations of each other (see, e.g., Theorem 5.11 in Kreps, 1988), we can choose $u = \hat{u}$. Finally, by (ii) $\Longrightarrow$ (iii) of Lemma 32 in Maccheroni, Marinacci, and Rustichini (2006a) for every $q \in Q$, and $k \in u(X)$, $\partial I_q(k) = \{q\}$. Let $\bar{k} \in \text{int}u(X) \neq \emptyset$ and observe that since Q is finite,

$$\begin{aligned}\lim_{\alpha \downarrow 0} \frac{\hat{I}\left(\bar{k}+\alpha\varphi\right) - \hat{I}\left(\bar{k}\right)}{\alpha} &= \lim_{\alpha \downarrow 0} \frac{\mathbb{E}_\mu\left[I_q\left(\bar{k}+\alpha\varphi\right)\right] - \mathbb{E}_\mu\left[I_q\left(\bar{k}\right)\right]}{\alpha} = \lim_{\alpha \downarrow 0} \mathbb{E}_\mu\left[\frac{I_q\left(\bar{k}+\alpha\varphi\right) - I_q\left(\bar{k}\right)}{\alpha}\right] \\ &= \mathbb{E}_\mu\left[\lim_{\alpha \downarrow 0} \frac{I_q\left(\bar{k}+\alpha\varphi\right) - I_q\left(\bar{k}\right)}{\alpha}\right] = \mathbb{E}_\mu\left[\int_S \varphi \mathrm{d}q\right].\end{aligned}$$

Now, applying (iii) $\Longrightarrow$ (ii) of Lemma 32 in Maccheroni, Marinacci, and Rustichini (2006a), we obtain that the unique $\hat{c}$ identified by the choice of $\hat{u}$ has

$$\hat{c}^{-1}(0) = \{\mathbb{E}_\mu[q]\}.$$

■

Lemma 11. *If $E \in \Sigma_{st}$ is nonnull and $\succsim$ satisfies Weak Order, Weak Certainty Independence, Continuity, Monotonicity, Uncertainty Aversion, and Weak Monotone Continuity, then $\succsim_E$ satisfies Weak Order, Weak Certainty Independence, Continuity, Monotonicity, Uncertainty Aversion, Nondegeneracy, and Weak Monotone Continuity.*

Proof. Let $f, g, h \in \mathcal{F}$. By Completeness of $\succsim$ at least one between

$$fEh \succsim gEh \text{ and } gEh \succsim fEh$$

holds. Therefore, by definition of $\succsim_E$ at least one between $f \succsim_E g$ and $g \succsim_E f$ holds.

Let $f, f', f'' \in \mathcal{F}$, with $f \succsim_E f'$ and $f' \succsim_E f''$. By definition of $\succsim_E$, there exist $h', h'' \in \mathcal{F}$ such that

$$fEh' \succsim f'Eh' \text{ and } f'Eh'' \succsim f''Eh''.$$

Since $E \in \Sigma_{st}$, we have

$$fEh'' \succsim f'Eh''.$$

By Transitivity of $\succsim$, $fEh'' \succsim f''Eh''$, and so by definition of $\succsim_E$, $f \succsim_E f''$.

Let $f, g \in \mathcal{F}$, $x, x' \in X$, and $\gamma \in (0,1)$, be such that

$$\gamma f + (1-\gamma)\, x \succsim_E \gamma g + (1-\gamma)\, x.$$

Since $E \in \Sigma_{st}$, we have

$$(\gamma f + (1-\gamma)\, x)\, Ex \succsim (\gamma g + (1-\gamma)\, x)\, Ex.$$

By Weak Certainty Independence of $\succsim$ we get

$$(\gamma f + (1-\gamma)\, x')\, E\, (\gamma x + (1-\gamma)\, x') \succsim (\gamma g + (1-\gamma)\, x')\, E\, (\gamma x + (1-\gamma)\, x').$$

But then by definition of $\succsim_E$, we have $\gamma f + (1-\gamma)\, x' \succsim_E \gamma g + (1-\gamma)\, x'$, proving that $\succsim_E$ satisfies Weak Certainty Independence.

Let $f, g, h, h' \in \mathcal{F}$. Since $E \in \Sigma_{st}$, we have that

$$\begin{aligned} \{\gamma \in [0,1] : \gamma f + (1-\gamma)\, g \succsim_E h\} &= \{\gamma \in [0,1] : (\gamma f + (1-\gamma)\, g)\, Eh' \succsim hEh'\} \\ &= \{\gamma \in [0,1] : \gamma fEh' + (1-\gamma)\, gEh' \succsim hEh'\} \end{aligned}$$

and

$$\begin{aligned}\{\gamma \in [0,1] : h \succsim_E \gamma f + (1-\gamma) g\} &= \{\gamma \in [0,1] : hEh' \succsim (\gamma f + (1-\gamma) g) Eh'\} \\ &= \{\gamma \in [0,1] : hEh' \succsim \gamma fEh' + (1-\gamma) gEh'\}\end{aligned}$$

where the sets on the bottom lines are closed by Continuity of $\succsim$, proving that $\succsim_E$ satisfies Continuity.

Let $f, g, h \in \mathcal{F}$ and $f(s) \succsim_E g(s)$ for all $s \in S$. Then, $fEh \succsim gEh$ by Monotonicity of $\succsim$. Therefore, by definition of $\succsim_E$, $f \succsim_E g$ and so $\succsim_E$ satisfies Monotonicity.

Let $f, g, h \in \mathcal{F}$, $\gamma \in (0,1)$ and $f \sim_E g$. Since $E \in \Sigma_{st}$, $fEh \sim gEh$ and by Uncertainty Aversion, $(\gamma f + (1-\gamma) g) Eh = \gamma fEh + (1-\gamma) gEh \succsim fEh$. By definition of $\succsim_E$, this implies that $\gamma f + (1-\gamma) g \succsim_E f$ and so $\succsim_E$ satisfies Uncertainty Aversion.

Since E is nonnull, there exist $f, g, h \in \mathcal{F}$ such that $fEh \succ gEh$. But then, since $E \in \Sigma_{st}$, there is no $h' \in \mathcal{F}$ with $gEh' \succsim fEh'$. Therefore, by definition of $\succsim_E$, $f \succ_E g$ and $\succsim_E$ satisfies Nondegeneracy.

Let $f, g, h \in \mathcal{F}$, $x \in X$, $(E_i)_{i \in \mathbb{N}} \in \Sigma^{\mathbb{N}}$ with $E_1 \supseteq E_2 \supseteq ...$ and $\cap_{n \geq 1} E_n = \emptyset$, and $f \succ_E g$. Since $E \in \Sigma_{st}$, $fEh \succ gEh$. Moreover, $(E'_i)_{i \in \mathbb{N}}$ where $E'_i = E_i \cap E$ is such that $E'_1 \supseteq E'_2 \supseteq ...$ and $\cap_{n \geq 1} E'_n \subseteq \cap_{n \geq 1} E_n = \emptyset$. Then $(xE'_i f) Eh = xE'_i (fEh)$ for all $i \in \mathbb{N}$ and by Weak Monotone Continuity and the fact that $fEh \succ gEh$ there exists $n_0 \in \mathbb{N}$ such that $\left(xE'_{n_0} f\right) Eh \succ gEh$. But notice that

$$(xE_{n_0} f) Eh = \left(xE'_{n_0} f\right) Eh \succ gEh$$

and therefore $xE_{n_0} f \succ_E g$, as $E \in \Sigma_{st}$. ■

Lemma 12. *Let $\Omega \times \{\rho\} \in \Sigma_{st}$ be nonnull and contain at least three disjoint nonnull events, and suppose $\succsim$ satisfies Weak Order, Weak Certainty Independence, Continuity, Monotonicity, Uncertainty Aversion, Nondegeneracy, Weak Monotone Continuity, the Intramodel Sure-Thing Principle, and Structured Savage. For every $f, g \in \mathcal{F}$,*

we have

$$f \succsim_\rho g \iff \min_{q\in\Delta(S)} \mathbb{E}_q[u_\rho(f)] + \frac{R(q||p_\rho)}{\lambda_\rho} \geq \min_{q\in\Delta(S)} \mathbb{E}_q[u_\rho(g)] + \frac{R(q||p_\rho)}{\lambda_\rho}$$

where u_ρ is a nonconstant affine function, $\lambda_\rho \in [0,\infty)$, and $p_\rho \in \Delta(S)$. Moreover, if $\Omega \times \{\rho\} \in \Sigma_{st}$ is strongly nonnull, u_ρ can be chosen to be the same for all such ρ and $suppp_\rho \subseteq \Omega \times \{\rho\}$.

Proof. By Lemma 11 $\succsim_\rho$ satisfies Weak Order, Weak Certainty Independence, Continuity, Monotonicity, Uncertainty Aversion, Nondegeneracy, and Weak Monotone Continuity. We now show that for every $f, g, h, \bar{h} \in \mathcal{F}$ and $E \in \Sigma$, we have

$$fEh \succsim_\rho gEh \implies fE\bar{h} \succsim_\rho gE\bar{h}.$$

Observe that by definition of $\succsim_\rho$, $fEh \succsim_\rho gEh$ implies that there exists $\hat{h} \in \mathcal{F}$ such that

$$(fEh)\,\rho\hat{h} \succsim (gEh)\,\rho\hat{h}.$$

But then, there exists $h' \in \mathcal{F}$ such that

$$\begin{aligned}
& (fEh)\,\rho\hat{h} \succsim (gEh)\,\rho\hat{h} \\
\implies\ & (f\{(\omega,\rho) : (\omega,\rho) \in E\}\,h)\,\rho\hat{h} \succsim (g\{(\omega,\rho) : (\omega,\rho) \in E\}\,h)\,\rho\hat{h} \\
\implies\ & (f\{(\omega,\rho') : \rho' \in \Delta(\Omega), (\omega,\rho) \in E\}\,h)\,\rho\hat{h} \succsim (g\{(\omega,\rho') : \rho' \in \Delta(\Omega), (\omega,\rho) \in E\}\,h)\,\rho\hat{h} \\
\implies\ & (f\{(\omega,\rho') : \rho' \in \Delta(\Omega), (\omega,\rho) \in E\}\,h) \succsim_\rho (g\{(\omega,\rho') : \rho' \in \Delta(\Omega), (\omega,\rho) \in E\}\,h) \\
\implies\ & \left(f\{(\omega,\rho') : \rho' \in \Delta(\Omega), (\omega,\rho) \in E\}\,\bar{h}\right) \succsim_\rho \left(g\{(\omega,\rho') : \rho' \in \Delta(\Omega), (\omega,\rho) \in E\}\,\bar{h}\right) \\
\implies\ & \left(f\{(\omega,\rho') : \rho' \in \Delta(\Omega), (\omega,\rho) \in E\}\,\bar{h}\right)\rho h' \succsim \left(g\{(\omega,\rho') : \rho' \in \Delta(\Omega), (\omega,\rho) \in E\}\,\bar{h}\right)\rho h' \\
\implies\ & \left(f\{(\omega,\rho) : (\omega,\rho) \in E\}\,\bar{h}\right)\rho h' \succsim \left(g\{(\omega,\rho) : (\omega,\rho) \in E\}\,\bar{h}\right)\rho h' \\
\implies\ & \left(fE\bar{h}\right)\rho h' \succsim \left(gE\bar{h}\right)\rho h' \implies fE\bar{h} \succsim_\rho gE\bar{h}
\end{aligned}$$

where the third, fifth, and eighth implications follow from the definition of $\succsim_\rho$, the fourth implication follows from the Intramodel Sure-Thing Principle, and the other

implications only rewrite the acts involved.

Next, observe that if $E \subseteq \Omega \times \{\rho\}$ is nonnull, then there exist $f, g, h \in \mathcal{F}$ with $(fEh)\,\rho h = fEh \succ gEh = (gEh)\,\rho h$. By Structured Savage P2, this implies that $fEh \succ_\rho gEh$, so that E is nonnull for the preference $\succsim_\rho$. With this, the first part follows from Theorem 1 in Strzalecki (2011). For the second part, notice that by Theorem 3 and Lemma 30 in Maccheroni, Marinacci, and Rustichini (2006a), $\succsim$ admits a variational representation:

$$f \succsim g \Longleftrightarrow \min_{p \in \Delta(S)} \left(\int u(f)\, \mathrm{d}p + c(p) \right) \geq \min_{p \in \Delta(S)} \left(\int u(g)\, \mathrm{d}p + c(p) \right) \tag{26}$$

for some nonconstant affine $u : X \to \mathbb{R}$ and a lower semicontinuous and grounded function $c : \Delta(S) \to [0, \infty]$.

Next, assume $\Omega \times \{\rho\}$ is strongly nonnull. Notice that $\succsim$ and $\succsim_\rho$ coincide on X. Indeed, let $x \succ x'$. Since $\Omega \times \{\rho\}$ is strongly nonnull $x \succ x'\rho x$ and given that $\Omega \times \{\rho\} \in \Sigma_{st}$ it follows that $x \succ_\rho x'$. Conversely, let $x \succsim x'$, then by equation (26) $u(x) \geq u(x')$. Since c is grounded, there exists $q^* \in \Delta(S)$ with $c(q^*) = 0$. But then

$$\begin{aligned} u(x) &\geq u(x')\, q^*(\Omega \times \{\rho\}) + (1 - q^*(\Omega \times \{\rho\}))\, u(x) \\ &\geq \min_{q \in \Delta(S)} \left(u(x')\, q(\Omega \times \{\rho\}) + (1 - q(\Omega \times \{\rho\}))\, u(x) + c(q) \right) \end{aligned}$$

that is, $x\,(\Omega \times \{\rho\})\, x \succsim x'\,(\Omega \times \{\rho\})\, x$, and $x \succsim_\rho x'$. Therefore, by the uniqueness up to a positive affine transformation of u guaranteed by Corollary 5 in Maccheroni, Marinacci, and Rustichini (2006a) and the fact that every two affine functions that represent $\succsim$ on X are positive affine transformations of each other (see, e.g., Theorem 5.11 in Kreps, 1988), we can choose $u = u_\rho$. Suppose by way of contradiction that there exists $E \in \Sigma$ such that $E \cap (\Omega \times \{\rho\}) = \emptyset$ and $p_\rho(E) > 0$. Let $x, y \in X$ with $x \succ y$. Then,

$$u(x) > u(y)\, p_\rho(E) + u(x)\,(1 - p_\rho(E)) \geq \min_{q \in \Delta(S)} \int u(yEx)\, \mathrm{d}q + \frac{1}{\lambda_\rho} R(q || p_\rho)$$

and so by equation (**??**), $x \succ_\rho yEx$. But since $x = x(\Omega \times \{\rho\})x$, $x = (yEx)(\Omega \times \{\rho\})x$ and $\Omega \times \{\rho\} \in \Sigma_{st}$ this would imply $x \succ x$, a contradiction to the Weak Order of $\succsim$.
■

Lemma 13. *Suppose that the assumptions of Theorem 3 hold. Let $\succsim$ be such that for all $f, g \in \mathcal{F}$*

$$f \succsim g \Longleftrightarrow \mathbb{E}_\mu \left[\min_{p \in \Delta(S)} \mathbb{E}_p \left[u(f) \right] + \frac{R(p||q)}{\lambda_q} \right] \geq \mathbb{E}_\mu \left[\min_{p \in \Delta(S)} \mathbb{E}_p \left[u(g) \right] + \frac{R(p||q)}{\lambda_q} \right]$$

where $u : X \to \mathbb{R}$ is a nonconstant affine function, $Q \subseteq \Delta(S)$ is a finite and nonempty set such that

$$q(\{\omega, \rho_q\}) = \rho_q(\omega) \qquad \forall q \in Q, \forall \omega \in \Omega, \tag{27}$$

for some $\rho_q \in \Delta(\Omega)$, $\mu \in \Delta(Q)$, and $(\lambda_q)_{q \in Q} \in \mathbb{R}_+^Q$. Then:

1. *For every $\Omega \times B \in \Sigma_s$ and $f, h \in \mathcal{F}$*

$$\begin{aligned} & \int_Q \min_{p \in \Delta(S)} \int_S u(f_{\Omega \times B} h)\, dp + \frac{R(p||q)}{\lambda_q} d\mu(q) \\ = & \int_{\{q \in Q : \rho_q \in B\}} \min_{p \in \Delta(S)} \mathbb{E}_p \left[u(f) \right] + \frac{R(p||q)}{\lambda_q} d\mu(q) \\ & + \int_{Q \setminus \{q \in Q : \rho_q \in B\}} \min_{p \in \Delta(S)} \mathbb{E}_p \left[u(h) \right] + \frac{R(p||q)}{\lambda_q} d\mu(q). \end{aligned}$$

2. *For every $\Omega \times B \in \Sigma_s$, if $\mu(\{q \in Q : \rho_q \in B\}) = 0$, then $\Omega \times B$ is null.*

Proof. 1) Let $\Omega \times B \in \Sigma_s$ and $f, h \in \mathcal{F}$. We have

$$\begin{aligned}
&\int_Q \min_{p\in\Delta(S)} \int_S u\left(f_{\Omega\times B}h\right) \mathrm{d}p + \frac{R\left(p||q\right)}{\lambda_q} \mathrm{d}\mu\left(q\right) \\
= &\int_{\{q\in Q:\rho_q\in B\}} \min_{p\in\Delta(S)} \int_S u\left(f_{\Omega\times B}h\right) \mathrm{d}p + \frac{R\left(p||q\right)}{\lambda_q} \mathrm{d}\mu\left(q\right) \\
&+ \int_{Q\setminus\{q\in Q:\rho_q\in B\}} \min_{p\in\Delta(S)} \int_S u\left(f_{\Omega\times B}h\right) \mathrm{d}p + \frac{R\left(p||q\right)}{\lambda_q} \mathrm{d}\mu\left(q\right) \\
= &\int_{\{q\in Q:\rho_q\in B\}} \min_{p\in\Delta(S):q\gg p} \int_S u\left(f_{\Omega\times B}h\right) \mathrm{d}p + \frac{R\left(p||q\right)}{\lambda_q} \mathrm{d}\mu\left(q\right) \\
&+ \int_{Q\setminus\{q\in Q:\rho_q\in B\}} \min_{p\in\Delta(S):q\gg p} \int_S u\left(f_{\Omega\times B}h\right) \mathrm{d}p + \frac{R\left(p||q\right)}{\lambda_q} \mathrm{d}\mu\left(q\right) \\
= &\int_{\{q\in Q:\rho_q\in B\}} \min_{p\in\Delta(S):q\gg p} \int_S u\left(f\right) \mathrm{d}p + \frac{R\left(p||q\right)}{\lambda_q} \mathrm{d}\mu\left(q\right) \\
&+ \int_{Q\setminus\{q\in Q:\rho_q\in B\}} \min_{p\in\Delta(S):q\gg p} \int_S u\left(h\right) \mathrm{d}p + \frac{R\left(p||q\right)}{\lambda_q} \mathrm{d}\mu\left(q\right) \\
= &\int_{\{q\in Q:\rho_q\in B\}} \min_{p\in\Delta(S)} \int_S u\left(f\right) \mathrm{d}p + \frac{R\left(p||q\right)}{\lambda_q} \mathrm{d}\mu\left(q\right) + \int_{Q\setminus\{q\in Q:\rho_q\in B\}} \min_{p\in\Delta(S)} \int_S u\left(h\right) \mathrm{d}p + \frac{R\left(p||q\right)}{\lambda_q} \mathrm{d}\mu\left(q\right)
\end{aligned}$$

where the third equality follows from the fact that by equation (27) $q \gg p$ and $\rho_q \in B$ imply $\mathrm{supp}p \subseteq \mathrm{supp}q \subseteq \Omega \times B$ (and conversely $q \gg p$ and $\rho_q \notin B$ imply $\mathrm{supp}p \subseteq \mathrm{supp}q \subseteq S \setminus (\Omega \times B)$).

2) It follows from 1), since in this case for every $f, g, h \in \mathcal{F}$

$$\begin{aligned}
&f_{\Omega\times B}h \succsim g_{\Omega\times B}h \\
\iff &\int_{\{q\in Q:\rho_q\notin B\}} \min_{p\in\Delta(S)} \mathbb{E}_p\left[u\left(h\right)\right] + \frac{R\left(p||q\right)}{\lambda_q} \mathrm{d}\mu\left(q\right) \geq \int_{\{q\in Q:\rho_q\notin B\}} \min_{p\in\Delta(S)} \mathbb{E}_p\left[u\left(h\right)\right] + \frac{R\left(p||q\right)}{\lambda_q} \mathrm{d}\mu\left(q\right)
\end{aligned}$$

and the RHS is always trivially satisfied as an equality. ■

Lemma 14. *Suppose that the assumptions of Theorem 3 hold. Let $\succsim$ be such that for all $f, g \in \mathcal{F}$*

$$f \succsim g \iff \mathbb{E}_\mu\left[\min_{p\in\Delta(S)} \mathbb{E}_p\left[u\left(f\right)\right] + \frac{R\left(p||q\right)}{\lambda_q}\right] \geq \mathbb{E}_\mu\left[\min_{p\in\Delta(S)} \mathbb{E}_p\left[u\left(g\right)\right] + \frac{R\left(p||q\right)}{\lambda_q}\right]$$

where $u : X \to \mathbb{R}$ is a nonconstant affine function, $Q \subseteq \Delta\left(S\right)$ is finite, nonempty, and

such that

$$q\left(\{\omega, \rho_q\}\right) = \rho_q\left(\omega\right) \qquad \forall q \in Q, \forall \omega \in \Omega,$$

for some $\rho_q \in \Delta\left(\Omega\right)$, $\mu \in \Delta\left(Q\right)$, *and* $\left(\lambda_q\right)_{q \in Q} \in \mathbb{R}_+^Q$. *Then* $\succsim$ *satisfies Uniform Misspecification Concern if and only if there exists* λ^* *with* $\lambda_q = \lambda^*$ *for all* $q \in \text{supp}\mu$.

Proof. (If) Let $\rho, \rho' \in \Delta\left(\Omega\right)$, $f, g \in \mathcal{F}$, and $x \in X$ be such that $\Omega \times \{\rho\}$ and $\Omega \times \{\rho'\}$ are nonnull,

$$\rho\left(\{\omega : f\left(\omega, \rho\right) = y\}\right) = \rho'\left(\{\omega : g\left(\omega, \rho'\right) = y\}\right) \quad \forall y \in X, \tag{28}$$

and $f \succsim_{\Omega \times \{\rho\}} x$. Since $\Omega \times \{\rho\}$ and $\Omega \times \{\rho'\}$ are nonnull, by part 2 of Lemma 13 there exist $q, q' \in Q$ with $\mu\left(\{q\}\right) > 0$, $\mu\left(\{q'\}\right) > 0$, $\rho_q = \rho$, and $\rho_{q'} = \rho'$. Let

$$\phi\left(c\right) = -\exp\left(-\lambda^* c\right), \qquad \forall c \in u\left(X\right)$$

and let $\xi \in \Delta\left(X\right)$ be the finite support probability measure such that for all $y \in X$, $\xi\left(y\right) = q\left(\{(\omega, \rho_q) : f\left(\omega, \rho_q\right) = y\}\right)$, then

$$\int_\Omega \phi\left(u\left(f\right)\right) \mathrm{d}q = \int_X \phi\left(u(y)\right) \mathrm{d}\xi\left(y\right).$$

Moreover, equation (28) implies

$$\int_\Omega \phi\left(u\left(g\right)\right) \mathrm{d}q' = \int_X \phi\left(u(y)\right) \mathrm{d}\xi\left(y\right).$$

Therefore, by Lemma 13 both $f \succsim_{\Omega \times \{\rho\}} x$ and $g \succsim_{\Omega \times \{\rho'\}} x$ mean that

$$\int_X \phi\left(u(y)\right) \mathrm{d}\xi\left(y\right) \geq \phi\left(u\left(x\right)\right)$$

proving that $\succsim$ satisfies Uniform Misspecification Concern.

(Only if) Suppose by way of contradiction that there exist $q, q' \subseteq Q$ and $k \in \mathbb{R}_{++}$

with $\mu(\{q\}) > 0$ and $\mu(\{q'\}) > 0$ and

$$\lambda_q > k > \lambda_{q'}. \tag{29}$$

Since the state space is adequate there exist two events $W_q \subseteq \Delta(\Omega), W_{q'} \subseteq \Delta(\Omega)$ and $c \in (0,1)$ with

$$\rho_q(W_q) = \rho_{q'}(W_{q'}) = c.$$

Moreover, $q(W_q \times \{\rho_q\}) = c = q'(W_{q'} \times \{\rho_{q'}\})$ and

$$q(W_{q'} \times \{\rho_{q'}\}) = 0 = q'(W_q \times \{\rho_q\}).$$

Pick $z, y \in X$ with $z \succ y$. We have that

$$\begin{aligned} & \rho_q(\{\omega : z((W_q \times \{\rho_q\}) \cup (W_{q'} \times \{\rho_{q'}\}))\, y(\omega, \rho_q) = x\}) \\ = \;& \rho_{q'}(\{\omega : z((W_q \times \{\rho_q\}) \cup (W_{q'} \times \{\rho_{q'}\}))\, y(\omega, \rho_{q'}) = x\}) \end{aligned}$$

for all $x \in X$. By the convexity of X and Lemma 13 there exists $\hat{x} \in X$ with $z \succ \hat{x} \succ y$ and

$$z((W_q \times \{\rho_q\}) \cup (W_{q'} \times \{\rho_{q'}\}))\, y \sim_{\rho_{q'}} \hat{x}.$$

But by equation (29) and Lemma 13 we have

$$\hat{x} \succ_{\rho_q} z((W_q \times \{\rho_q\}) \cup (W_{q'} \times \{\rho_{q'}\}))\, y$$

a violation of Uniform Misspecification Concern. ■

Proof of Theorem 3. (Only if) That $\succsim$ satisfies Weak Order, Weak Certainty Independence, Continuity, Monotonicity, Uncertainty Aversion, Nondegeneracy, and Weak Monotone Continuity follows from Lemma 10.

Let $\rho \in \Delta(\Omega)$, $W \subseteq \Omega$, $f, g, h, h' \in \mathcal{F}$, and $fWh \succsim_\rho gWh$. If $\Omega \times \{\rho\}$ is null then we trivially have $fWh' \succsim_\rho gWh'$. Therefore, suppose $\Omega \times \{\rho\}$ is nonnull. By Lemma

13, and since $q \mapsto \rho_q$ is injective, there exists $\bar{q} \in \Delta(S)$ with $\rho_{\bar{q}} = \rho$, $\mu(\{\bar{q}\}) > 0$, and

$$\min_{p \in \Delta(S)} \int_S u(fWh)\,\mathrm{d}p + \frac{1}{\lambda} R(p||\bar{q}) \geq \min_{p \in \Delta(S)} \int_S u(gWh)\,\mathrm{d}p + \frac{1}{\lambda} R(p||\bar{q}).$$

By Proposition 1.4.2 in Dupuis and Ellis (2011) this is equivalent to

$$\int_S \phi(u(fWh))\,\mathrm{d}\bar{q} \geq \int_S \phi(u(gWh))\,\mathrm{d}\bar{q}$$

with $\phi(\cdot) = -\exp(-\lambda(\cdot))$. This is also equivalent to

$$\begin{aligned} & \int_{W \times \Delta(\Omega)} \phi(u(f))\,\mathrm{d}\bar{q} + \int_{(\Omega \setminus W) \times \Delta(\Omega)} \phi(u(h))\,\mathrm{d}\bar{q} \\ \geq & \int_{W \times \Delta(\Omega)} \phi(u(g))\,\mathrm{d}\bar{q} + \int_{(\Omega \setminus W) \times \Delta(\Omega)} \phi(u(h))\,\mathrm{d}\bar{q} \end{aligned} \tag{30}$$

or

$$\int_{W \times \Delta(\Omega)} \phi(u(f))\,\mathrm{d}\bar{q} \geq \int_{W \times \Delta(\Omega)} \phi(u(g))\,\mathrm{d}\bar{q}.$$

But then, by reversing all the steps with h' in place of h we get

$$fWh' \succsim_\rho gWh'$$

and therefore $\succsim$ satisfies Intramodel Sure-Thing Principle.

Moreover, $\succsim$ satisfies Uniform Misspecification Concern by Lemma 14. That there is a finite set $B \subseteq \Delta(\Omega)$ such that $\Omega \times (\Delta(\Omega) \setminus B)$ is null immediately follows from the representation and part 2 of Lemma 13. Let $\Omega \times B \in \Sigma_s$ and $f, g, h, h' \in \mathcal{F}$. If $\Omega \times B$ is null, we clearly have that $\Omega \times B \in \Sigma_{st}$. Suppose $\Omega \times B$ is nonnull, then

$$\begin{array}{ccc} f(\Omega \times B)\,h & \succsim & g(\Omega \times B)\,h \\ & \Longleftrightarrow & \\ \displaystyle\int_{\{q \in Q : \rho_q \in B\}} \min_{p \in \Delta(S)} \mathbb{E}_p[u(f)] + \frac{R(p||q)}{\lambda_q} \mathrm{d}\mu(q) & \geq & \displaystyle\int_{\{q \in Q : \rho_q \in B\}} \min_{p \in \Delta(S)} \mathbb{E}_p[u(g)] + \frac{R(p||q)}{\lambda_q} \mathrm{d}\mu(q) \\ & \Longleftrightarrow & \\ f(\Omega \times B)\,h' & \succsim & g(\Omega \times B)\,h' \end{array}$$

where the two equivalences follow by Lemma 13. This proves that $\Omega \times B \in \Sigma_{st}$. Since B was chosen to be an arbitrary measurable subset of $\Delta(\Omega)$, $\Sigma_s \subseteq \Sigma_{st}$, and Structured Savage P2 holds.

That $\succsim$ satisfies Structured Savage P4 and Uncertainty Neutrality over Models immediately follows from Lemma 13 and the representation.

(If) By Structured Savage's P2, $\Sigma_s \subseteq \Sigma_{st}$. Suppose $E \in \Sigma_s$ is nonnull, and let $x, x' \in X$ with $x \succ x'$. Then there exist $f, g, h \in \mathcal{F}$ such that $fEh \succ gEh$. Since f and g are simple acts, they assume finitely many values, and by Weak Order, there exist $\bar{x}, \underline{x} \in X$ with

$$\bar{x} \succsim f(s), \qquad g(s) \succsim \underline{x}, \qquad \forall s \in E.$$

Since $E \in \Sigma_s \subseteq \Sigma_{st}$, $fE\bar{x} \succ gE\bar{x}$. By the Monotonicity and Weak Order parts of the Variational Axiom, $\underline{x}\emptyset\bar{x} = \bar{x}E\bar{x} \succsim fE\bar{x} \succ gE\bar{x} \succsim \underline{x}E\bar{x}$. Therefore, by Structured Savage P4, $x = x'\emptyset x \succ x'Ex$. Since $E \in \Sigma_s$ and $x, x' \in X$ were arbitrarily chosen, each nonnull $E \in \Sigma_s$ is also strongly nonnull.

Next, fix a finite B, such that for each $\rho \in B$, $\Omega \times \{\rho\}$ is nonnull, and such that $S \setminus \{\Omega \times B\}$ is null. Such a set exists by the Structured Savage axiom, and the cardinality of B is at least 3 by assumption of the theorem. For every $\rho \in B$, by the previous part of the proof $\Omega \times \{\rho\}$ is strongly nonnull and so by Lemma 12 we have

$$f \succsim_\rho g \iff \min_{p \in \Delta(S)} \int_S \hat{u}(f) \, \mathrm{d}p + \frac{1}{\lambda_\rho} R(p || q_\rho) \tag{31}$$

for some $q_\rho \in \Delta(S)$ with support contained in $\Omega \times \{\rho\}$ and a nonconstant affine $\hat{u}$.

Claim 5. *We have $q_\rho = \rho \times \delta_\rho$.*

Proof of the Claim. Since the space is adequate, there exists $v_\rho \in (0, 1)$ such that $\rho(\omega) \in \{0, v_\rho\}$. In particular, by applying Uniform Misspecification Concern with $\rho = \rho'$, we obtain that $q_\rho(\omega, \rho) = v_\rho \iff \rho(\omega) = v_\rho$, and the desired conclusion follows. □

Let

$$Q = \{q_\rho \in \Delta(S) : \rho \in B\}. \tag{32}$$

Identify each act $f \in \mathcal{F}$ with the real-valued function $\hat{f} : Q \to \hat{u}(X)$ with

$$\hat{f}(q_\rho) = \min_{p \in \Delta(S)} \int_S \hat{u}(f)\, \mathrm{d}p + \frac{1}{\lambda_\rho} R(p||q_\rho) \qquad \forall \rho \in B$$

where λ_ρ is given by equation (31).

We now show that

$$\hat{f} = \hat{g} \implies f \sim g \qquad \forall f, g \in \mathcal{F}.$$

We partition S in $\left\{ \{\Omega \times \rho\}_{\rho \in B}, S \setminus \{\Omega \times B\} \right\}$ and establish the claim by induction on the number of cells of the partition on which f and g are not identical. Let f and g be such that $\hat{f} = \hat{g}$ and they differ on one element of the partition, say E. Then $f = fEg \sim g$ by definition of $\sim_E$ and Structured Savage P2, so $f \sim g$. For the inductive step, suppose that whenever f and g are such that $\hat{f} = \hat{g}$ and they differ at most on $n \in \mathbb{N}$ elements of the partition, we have $f \sim g$. Let f and g be such that $\hat{f} = \hat{g}$ and they differ on $n+1 \in \mathbb{N}$ elements of the partition. Let E be an element of the partition on which they differ. Then, fEg and g differ on one element of the partition, and fEg and f differ on n elements of the partition. Therefore, by the inductive hypothesis, we have $g \sim fEg \sim f$.

Moreover, it is immediate to see that $\hat{u}(X)^Q \subseteq \left\{ \hat{f} : f \in \mathcal{F} \right\}$. Therefore, with a slight abuse of notation we let $\succsim$ denote also the binary relation on $\hat{u}(X)^Q$ defined by $\hat{f} \succsim \hat{g}$ if and only if $f \succsim g$.

Claim 6. *For every $v, v', w, z \in \hat{u}(X)$, $\rho \in B$, and $\gamma \in (0,1)$*

$$v_\rho w \succsim (\gamma v + (1-\gamma) v')_\rho z \iff ((1-\gamma) v + \gamma v')_\rho w \succsim v'_\rho z.$$

Proof of the Claim. If $v = v'$ the equivalence is obvious. Suppose without loss of generality that $v' > v$.

1. Let $v_\rho w \succsim (\gamma v + (1-\gamma) v')_\rho z$. This implies that $w > z$. Then, by Continuity, Structured Savage, and the fact that $\Omega \times \{\rho\}$ is strongly nonnull there exists $\alpha \in [0,1]$ with

$$v_\rho (\alpha w + (1-\alpha) z) \sim (\gamma v + (1-\gamma) v')_\rho z.$$

By Uncertainty Neutrality over Models, this implies that $((1-\gamma) v + \gamma v')_\rho (\alpha w + (1-\alpha) z) \sim v'_\rho z$. By Monotonicity, this implies that $((1-\gamma) v + \gamma v')_\rho w \succsim v'_\rho z$.

2. Let $((1-\gamma) v + \gamma v')_\rho w \succsim v'_\rho z$. This implies that $w > z$. Then, by Continuity, Structured Savage, and the fact that $\Omega \times \{\rho\}$ is strongly nonnull there exists $\alpha \in [0,1]$ with

$$((1-\gamma) v + \gamma v')_\rho (\alpha w + (1-\alpha) z) \sim v'_\rho z.$$

By Uncertainty Neutrality over Models, this implies that $v_\rho (\alpha w + (1-\alpha) z) \sim (\gamma v + (1-\gamma) v')_\rho z$. By Monotonicity, this implies that $v_\rho w \succsim (\gamma v + (1-\gamma) v')_\rho z$. □

By the previous claim, Continuity, Structured Savage, and Theorem VII.3.5 in Wakker (2013) there exists $\mu \in \Delta(Q)$ such that for all $\psi, \psi' \in \hat{u}(X)^Q$[35]

$$\psi \succsim \psi' \iff \sum_{q \in Q} \psi(q) \mu(q) \geq \sum_{q \in Q} \psi'(q) \mu(q).$$

Moreover, by Observation VII.3.5 in Wakker (2013), μ is uniquely identified.

[35]Formally, one needs to apply Theorem VII.3.5 in Wakker (2013) twice. The first application gives

$$\psi \succsim \psi' \iff \sum_{q \in Q} U_q(\psi(q)) \geq \sum_{q \in Q} U_q(\psi'(q))$$

for some concave and increasing functions $(U_q : \hat{u}(X) \to \mathbb{R})_{q \in Q}$. The second application is to the preference $\succsim^-$ defined over $(\hat{u}(X))^Q$ by $\psi \succsim^- \psi' \iff \psi' \succsim \psi$ for all $\psi, \psi' \in (\hat{u}(X))^Q$. It gives that for all $\psi, \psi' \in \hat{u}(X)^Q$

$$\psi \succsim \psi' \iff \psi' \succsim^- \psi \iff \sum_{q \in Q} U_q^-(\psi'(q)) \geq \sum_{q \in Q} U_q^-(\psi(q)) \iff \sum_{q \in Q} -U_q^-(\psi(q)) \geq \sum_{q \in Q} -U_q^-(\psi'(q))$$

for some decreasing and concave functions $\left(U_q^- : -\hat{u}(X) \to \mathbb{R}\right)_{q \in Q}$. But since $-U_q^-(\cdot)$ is an increasing and convex function, then by Observation VII.3.5 in Wakker (2013) $\psi \succsim \psi' \iff \sum_{q \in Q} U_q^L(\psi(q)) \geq \sum_{q \in Q} U_q^L(\psi'(q))$ for some increasing and linear functions $\left(U_q^L : \hat{u}(X) \to \mathbb{R}\right)_{q \in Q}$, and the result follows by the Riesz Representation theorem.

But then, by definition of $\succsim$, we obtain that for all $f, g \in \mathcal{F}$

$$f \succsim g \Longleftrightarrow \mathbb{E}_{\mu}\left[\min_{p\in\Delta(S)} \mathbb{E}_p\left[u\left(f\right)\right] + \frac{R\left(p||q_\rho\right)}{\lambda_\rho}\right] \geq \mathbb{E}_{\mu}\left[\min_{p\in\Delta(S)} \mathbb{E}_p\left[u\left(g\right)\right] + \frac{R\left(p||q_\rho\right)}{\lambda_\rho}\right].$$

Moreover, by Lemma 14 Uniform Misspecification Concern implies that $\lambda = \lambda_\rho$ for all $\rho \in B$, proving the result. ■

Proof of Corollary 2. It immediately follows from Lemma 10 and Proposition 8 in Maccheroni, Marinacci, and Rustichini (2006a). ■

Proposition 7. *Let $\left(\succsim^h\right)_{h\in\mathcal{H}}$ be such that:*

1. *For every $h \in \mathcal{H}$, $\succsim^h$ satisfies the axioms of Theorem 3,*
2. *$\left(\succsim^h\right)_{h\in\mathcal{H}}$ satisfies Constant Preference Invariance and Dynamic Consistency over Models.*

Then for every $h \in \mathcal{H}$, $\succsim^h$ admits an average robust control representation $(u, Q, \mu\left(\cdot|h\right), \lambda_h)$.

Proof. That each $\succsim^h$ admits an average robust control representation $(u_h, Q_h, \mu_h, \lambda_h)$ where

$$q\left(\{\omega, \rho_q\}\right) = \rho_q\left(\omega\right) \qquad \forall q \in Q_h, \forall \omega \in \Omega,$$

for some $\rho_q \in \Delta\left(\Omega\right)$ follows from (the proof of) Theorem 3. That $u_h = u$ for some constant affine u follows from Constant Preference Invariance.

We now prove that Dynamic Consistency over Models implies $\mu\left(\cdot|h_t\right) = \mu_{h_t}$ for all $h_t = \left(\omega_i\right)_{i=1}^t \in \mathcal{H}_t$ such that $\prod_{i=1}^t \rho_q\left(\omega_i\right) > 0$ for some $q \in Q$. By definition, we have $\mu_{h_t} = \mu$ for the empty history. Let f and g be measurable with respect to Σ_s. Then we can suppress the dependence on ω in $f\left(\omega, \rho\right)$ and $g\left(\omega, \rho\right)$ and we have that

$$f \succsim^{h_t} g \iff f^0 \succsim^{\emptyset} g^0.$$

But by construction, the latter is equivalent to

$$\mathbb{E}_{\mu}\left[\gamma_{f(\rho_q)} \prod_{i=1}^t \rho_q\left(\omega_i\right)\left(u\left(\bar{z}\right) - u\left(\underline{z}\right)\right)\right] \geq \mathbb{E}_{\mu}\left[\gamma_{g(\rho_q)} \prod_{i=1}^t \rho_q\left(\omega_i\right)\left(u\left(\bar{z}\right) - u\left(\underline{z}\right)\right)\right].$$

Dividing both sides by the strictly positive ex-ante probability of history h_t, we obtain

$$\begin{aligned}&\frac{\int_{\Delta(\Delta(S))}\gamma_{f(\rho_q)}\prod_{i=1}^{t}\rho_q(\omega_i)\left(u(\bar{z})-u(\underline{z})\right)\mathrm{d}\mu(q)}{\int_{\Delta(\Delta(S))}\prod_{i=1}^{t}\rho_q(\omega_i)\,\mathrm{d}\mu(q)}\\ \geq\ &\frac{\int_{\Delta(\Delta(S))}\gamma_{g(\rho_q)}\prod_{i=1}^{t}\rho_q(\omega_i)\left(u(\bar{z})-u(\underline{z})\right)\mathrm{d}\mu(q)}{\int_{\Delta(\Delta(S))}\prod_{i=1}^{t}\rho_q(\omega_i)\,\mathrm{d}\mu(q)}.\end{aligned}$$

But then, by the formula for Bayesian updating, this is equivalent to

$$\int_{\Delta(\Delta(S))}\gamma_{f(\rho_q)}\left(u(\bar{z})-u(\underline{z})\right)\mathrm{d}\mu(q|h_t)\geq\int_{\Delta(\Delta(S))}\gamma_{g(\rho_q)}\left(u(\bar{z})-u(\underline{z})\right)\mathrm{d}\mu(q|h_t)$$

that is

$$\int_{\Delta(\Delta(S))}u(f(\rho_q))\,\mathrm{d}\mu(q|h_t)\geq\int_{\Delta(\Delta(S))}u(g(\rho_q))\,\mathrm{d}\mu(q|h_t).$$

That is, $\succsim^{h_t}$ admits an SEU representation of the acts measurable with respect to Σ_s with Bernoulli utility u and probability measure $\mu(\cdot|h_t)$. Since for the histories $h_t=(\omega_i)_{i=1}^{t}\in\mathcal{H}_t$ where $\prod_{i=1}^{t}\rho_q(\omega_i)=0$ for all $q\in Q$ Bayesian updating does not impose any restriction, the result follows. ■

Proof of Proposition 4. By Proposition 7, $\succsim^{h}$ admits an average robust control representation $(u,Q,\mu(\cdot|h),\lambda_h)$ for every $h\in\mathcal{H}$. Observe that since the outcome frequency is constant along the sequence $(h_{t_n})_{n\in\mathbb{N}}$, by Lemma 2, $\frac{LLR(h_{t_n},Q)}{t_n}=1/c$ for some $c\in\mathbb{R}_{++}$ and for all $n\in\mathbb{N}$. Suppose by way of contradiction that

$$l:=\liminf_{n\to\infty}c\lambda^{h_{t_n}}=\liminf_{n\to\infty}\frac{\lambda_{h_{t_n}}}{\frac{LLR(h_{t_n},Q)}{t_n}}<\limsup_{n\to\infty}\frac{\lambda_{h_{t_n}}}{\frac{LLR(h_{t_n},Q)}{t_n}}=\limsup_{n\to\infty}c\lambda_{h_{t_n}}=:L.$$

Let $\bar{q}\in Q$ be such that $\Omega\times\{\rho_{\bar{q}}\}$ is nonnull and so that $\bar{q}\in\min_{q\in Q}R\left(p^{h_{t_1}}||q\right)$. Since $\Omega\times\{\rho_{\bar{q}}\}$ contains at least three nonnull events, by Lemma 13, there is $E\subseteq W$ and $r\in(0,1)$ with $\rho_{\bar{q}}(E)=r$. Let $x,z\in X$, $\gamma^*,\gamma_*\in(0,1)$, and $\lambda^*,\lambda_*\in\left(\frac{l}{c},\frac{L}{c}\right)$ be such that $x\succ^{\emptyset}z$, $\lambda^*>\lambda_*$,

$$\begin{aligned}&\frac{-\mu(\bar{q})\log\left(r\exp\left(-\lambda^*\left(u(z)\right)\right)+(1-r)\exp\left(-\lambda^*\left(u(x)\right)\right)\right)}{\mu\left(\min_{q\in Q}R\left(p^{h_{t_1}}||q\right)\right)\lambda^*}+\left(1-\frac{\mu(\bar{q})}{\mu\left(\min_{q\in Q}R\left(p^{h_{t_1}}||q\right)\right)}\right)u(z)\\ =\ &u\left(\gamma^*x+(1-\gamma^*)z\right),\end{aligned}$$

and

$$\begin{aligned} & \frac{-\mu\left(\bar{q}\right)\log\left(r\exp\left(-\lambda_{*}\left(u\left(z\right)\right)\right)+\left(1-r\right)\exp\left(-\lambda_{*}\left(u\left(x\right)\right)\right)\right)}{\mu\left(\min_{q\in Q}R\left(p^{h_{t_1}}||q\right)\right)\lambda_{*}}+\left(1-\frac{\mu\left(\bar{q}\right)}{\mu\left(\min_{q\in Q}R\left(p^{h_{t_1}}||q\right)\right)}\right)u\left(z\right) \\ = \quad & u\left(\gamma_{*}x+\left(1-\gamma_{*}\right)z\right), \end{aligned}$$

where the existence of such γ_*, γ^* is guaranteed by u being affine. Moreover, it is easy to see that $\gamma_* > \gamma^*$. Consider a subsequence $(n_m)_{m\in\mathbb{N}}$ such that

$$\lim_{m\to\infty} c\lambda^{h_{t_{n_m}}} = l.$$

Moreover, let $M \in \mathbb{N}$ be such that for all $m \geq M$

$$\lambda^{h_{t_{n_m}}} < \frac{\lambda_* + \frac{l}{c}}{2}.$$

Similarly, let $(n_{\tilde{m}})_{\tilde{m}\in\mathbb{N}}$ such that

$$\lim_{\tilde{m}\to\infty} c\lambda^{h_{t_{n_{\tilde{m}}}}} = L.$$

Moreover, let $\tilde{M} \in \mathbb{N}$ be such that for all $\tilde{m} \geq \tilde{M}$

$$\lambda^{h_{t_{n_{\tilde{m}}}}} > \frac{\lambda^* + \frac{L}{c}}{2}.$$

With this, by Proposition 7 and Proposition 1.4.2 in Dupuis and Ellis (2011) we have that for all $m \geq M$ and $\tilde{m} \geq \tilde{M}$

$$\gamma^{x(E\times\{\rho_{\bar{q}}\})z}_{\succsim^{h_{t_{n_m}}}} > \gamma_* \text{ and } \gamma^{x(E\times\{\rho_{\bar{q}}\})z}_{\succsim^{h_{t_{n_{\tilde{m}}}}}} < \gamma^*.$$

But this in turn implies that $\succsim^{h_{t_{n_m}}}$ is never $(x, y, (E \times \{\rho_{\bar{q}}\}), (\gamma_* - \gamma^*))$-similar to $\succsim_{h_{t_{n_{\tilde{m}}}}}$ for

$$\min\left\{m, \tilde{m}\right\} \geq \max\left\{M, \tilde{M}\right\},$$

a contradiction. This shows that either $\lambda_{h_{t_n}}$ converges or it diverges to plus infinity.

The last part of the statement immediately by taking E in the first part of the proof to be equal to the one whose existence is asserted in the statement, and by the construction of γ_* and γ^* above. ■

Proof of Proposition 5. By Proposition 7 we know that each $\succsim^h$ admits an average robust control representation $(u, Q, \mu(\cdot|h), \lambda_h)$. Without loss of generality suppose that $\mu(\{q\}|\emptyset) > 0$ for all $q \in Q$. Let $(h_{t_n})_{n\in\mathbb{N}} \in \times_{n\in\mathbb{N}}\Omega^{t_n}$ be a equence of histories with empirical frequency $\hat{\rho} \notin \{\rho_q : q \in Q\}$. Observe that since the outcome frequency is constant along the sequence $(h_{t_n})_{n\in\mathbb{N}}$, by Lemma 2, $\frac{LLR(h_{t_n},Q)}{t_n} = c$ for some $c \in \mathbb{R}_{++}$ and for all $n \in \mathbb{N}$. Suppose by way of contradiction that

$$L : = \limsup_{n\to\infty} \frac{LLR(h_{t_n}, Q)}{\lambda_{h_{t_n}} t_n} > 0. \tag{33}$$

Since the state space is adequate there exist $k \in (0,1)$ and $(W_q)_{q\in Q} \in \left(2^\Omega\right)^Q$ such that $\rho_q(W_q) = k$ for all $q \in Q$. Define $E = \cup_{q\in Q}(W_q \times \{\rho_q\})$. Let $x, y \in X$ with $x \succ^\emptyset y$ and choose also $z \in X$ such that $x \succ^\emptyset z \succ^\emptyset y$ and

$$-\exp\left(-2\frac{c}{L}u(z)\right) = -k\exp\left(-2\frac{c}{L}u(x)\right) - (1-k)\exp\left(-2\frac{c}{L}u(y)\right)$$

where the existence of such z is guaranteed by u being affine and X being convex. Let $f \in \mathcal{F}$ be defined as $f = xEy$. But then equation (33) implies that for infinitely many $n \in \mathbb{N}$

$$f \succ_\rho^{h_{t_n}} z$$

a contradiction with Asymptotic Concern for every $\rho \in \{\rho_q : q \in Q\}$. ■

Proof of Proposition 6. By Proposition 7 we know that each $\succsim^h$ admits an average robust control representation $(u, Q, \mu(\cdot|h), \lambda_h)$ where

$$q(\{\omega, \rho_q\}) = \rho_q(\omega) \qquad \forall q \in Q, \forall \omega \in \Omega,$$

for some $\rho_q \in \Delta(\Omega)$. Without loss of generality suppose that $\mu(\{q\}|\emptyset) > 0$ for all $q \in Q$. Consider a sequence of histories $(h_{t_n})_{n\in\mathbb{N}}$ with outcome frequency constant

and not in $\{\rho_q : q \in Q\}$. Observe that since the outcome frequency is constant along the sequence $(h_{t_n})_{n\in\mathbb{N}}$, by Lemma 2, $\frac{LLR(h_{t_n},Q)}{t_n} = c$ for some $c \in \mathbb{R}_{++}$ and for all $n \in \mathbb{N}$. Suppose by way of contradiction that

$$L\colon = \liminf_{n\to\infty} \frac{LLR\left(h_{t_n}, Q\right)}{\lambda_{h_{t_n}} t_n} \in \mathbb{R}. \tag{34}$$

As the state space is adequate there exist $k \in (0,1)$ and $(W_q)_{q\in Q} \in \left(2^{\Omega}\right)^{Q}$ such that for every $q \in Q$, $\rho_q\left(W_q\right) = k$. Let $x, z \in X$ and $\gamma \in (0,1)$ be such that $x \succ^{\emptyset} z$ and

$$-\ln\left(k\exp\left(\frac{-cu\left(x\right)}{2\max\left\{1,L\right\}}\right) + (1-k)\exp\left(\frac{-cu\left(z\right)}{2\max\left\{1,L\right\}}\right)\right) = \frac{c\left(u\left(\gamma x + (1-\gamma)\, z\right)\right)}{2\max\left\{1,L\right\}},$$

where the existence of such γ is guaranteed by u being affine, and $\gamma < k$. Define $E = \cup_{q\in Q}\left(W_q \times \{\rho_q\}\right)$. Consider a subsequence $(n_m)_{m\in\mathbb{N}}$ such that

$$\lim_{m\to\infty} \frac{LLR\left(h_{t_{n_m}}, Q\right)}{\lambda_{h_{t_{n_m}}} t_{n_m}} = L.$$

Moreover, let M be such that for all $m \geq M$

$$\lambda^{h_{t_{n_m}}}/c > \frac{1}{2\max\left\{1,L\right\}}.$$

With this, by Proposition 7 if we let $\geq$ be the subjective utility preference with utility index u and belief $\int_Q p\mathrm{d}\mu\left(q\right)$, we have that for all $m \geq M$

$$\gamma^{xEz}_{\succsim^{h_{t_{n_m}}}} < \gamma \text{ and } \gamma^{xEz}_{\geq^{h_{t_{n_m}}}} = k.$$

By Corollary 2, this contradicts Asymptotic Leniency as then $\succsim^{h_{t_{n_m}}}$ and $\geq^{h_{t_{n_m}}}$ are not $(x, y, E, k-\gamma)$-similar for $m \geq M$. ■

.1.3 General Statistical Distances

The results of the chapter that involve the statistically sophisticated type extend easily to the case of an average of general divergence preferences (Cerreia-Vioglio,

Hansen, Maccheroni, and Marinacci, 2022), i.e., to decision criteria of the form

$$\int_Q \min_{p\in\Delta(S)} \mathbb{E}_{p_a}\left[u\left(a,y\right)\right] + \frac{1}{\lambda} D_\phi\left(p_a||q_a\right) \mathrm{d}\mu\left(q\right)$$

where

$$D_\phi\left(p_a||q_a\right) = \begin{cases} \int \phi\left(\frac{\mathrm{d}p_a}{\mathrm{d}q_a}\right) \mathrm{d}q_a & q_a \gg p_a \\ \infty & \text{otherwise} \end{cases}$$

for some continuous strictly convex function $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ with

$$\phi\left(1\right) = 0 \text{ and } \lim_{t\to\infty} \frac{\phi\left(t\right)}{t} = \infty.$$

From this expression, it is clear that the main case studied in the chapter is the one where $\phi\left(c\right) = c\log c - c + 1$. The only caveat is that the best reply function BR^λ must now be defined with respect to the relevant divergence.

.1.4 Computations supporting

Example 2

Observe that, compared Esponda and Pouzo (2016), we are adding (arbitrarily small) noise $l\varepsilon_1$ to the true tax schedule, fixing a problem in their original example. Indeed, without this modification, the relative entropy between the true and conjectured distribution is infinity for every model. We have

$$\begin{aligned} & R\left(p_a^*||q_a^\theta\right) = \text{const.} \\ + & \int_{\mathbb{R}}\int_{\mathbb{R}} \frac{-\exp\left(-\frac{(t-\tau(a+\omega_a))^2}{2l^2}\right)}{\sqrt{2\pi}} \log\left(\frac{\exp\left(-\left(\frac{t}{a+\omega_a}-\theta\right)^2/2\right)}{\sqrt{2\pi}}\right) \mathrm{d}t\mathrm{d}p_a^*\left(\omega_a\right) \\ = & \text{const.} + \int_{\mathbb{R}}\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\varepsilon_1^2}{2}\right)\left(\frac{\tau\left(a+\omega_a\right)+\varepsilon_1}{a+\omega_a}-\theta\right)^2/2\mathrm{d}\varepsilon_1\mathrm{d}p_a^*\left(\omega_a\right) \\ = & \text{const.} + \int_{\mathbb{R}}\int_{\mathbb{R}} \left(\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\varepsilon_1^2}{2}\right)\left(-2\frac{\tau\left(a+\omega_a\right)+\left(l\varepsilon_1\right)^2}{a+\omega_a}\theta+\theta^2\right)\right)/2\mathrm{d}\varepsilon_1\mathrm{d}p_a^*\left(\omega_a\right) \end{aligned}$$

taking the FOC, we get

$$\theta = \mathbb{E}_{p_a^*}\left[\frac{\tau\left(a+\omega_a\right)+l^2}{a+\omega_a}\right]$$

and so $Q\left(a\right) \sim \left\{q^{\mathbb{E}_{p_a^*}\left[\frac{\tau(a+\omega_a)}{a+\omega_a}\right]}\right\}$ for small l. The condition for not switching from an action a to an action a' with $a \geq a'$ in a Berk-Nash equilibrium in which the belief is concentrated on θ is

$$\begin{aligned}\mathbb{E}_{\omega_a,\varepsilon_2}\left[\left(a-a'\right)\left(1-\theta-\varepsilon_2\right)\right] &= \mathbb{E}_{\omega_a,\varepsilon_2}\left[\left(a+\omega_a\right)\left(1-\theta-\varepsilon_2\right)\right] - \mathbb{E}_{\omega_{a'},\varepsilon_2}\left[\left(a'+\omega_{a'}\right)\left(1-\theta-\varepsilon_2\right)\right] \\ &\geq c\left(a\right)-c\left(a'\right).\end{aligned}$$

By Proposition 1.4.2 in Dupuis and Ellis (2011), the condition for not switching from an action a to an action a' with $a \geq a'$ in a k-robust equilibrium in which the belief is concentrated on θ is

$$\begin{aligned}&\frac{-k}{R\left(p_a^*||q_a^\theta\right)}\log\mathbb{E}_{\omega_a,\omega_{a'},\varepsilon_2}\left[\exp\left(\frac{R\left(p_a^*||q_a^\theta\right)\left[\left(a'+\omega_{a'}\right)\left(1-\theta-\varepsilon_2\right)-\left(a+\omega_a\right)\left(1-\theta-\varepsilon_2\right)\right]}{k}\right)\right] \\ = &\frac{\mathbb{E}_{\omega_a,\varepsilon_2}\left[\exp\left(-\frac{R\left(p_a^*||q_a^\theta\right)\left(a+\omega_a\right)\left(1-\theta-\varepsilon_2\right)}{k}\right)\right]}{\mathbb{E}_{\omega_{a'},\varepsilon_2}\left[\exp\left(-\frac{R\left(p_a^*||q_a^\theta\right)\left(a'+\omega_a\right)\left(1-\theta-\varepsilon_2\right)}{k}\right)\right]} \\ = &\frac{-k}{R\left(p_a^*||q_a^\theta\right)}\log\mathbb{E}_{\omega_a,\omega_{a'},\varepsilon_2}\left[\exp\left(-\frac{R\left(p_a^*||q_a^\theta\right)\left(a+\omega_a\right)\left(1-\theta-\varepsilon_2\right)}{k}\right)\right] \\ &+\frac{k}{R\left(p_a^*||q_a^\theta\right)}\log\mathbb{E}_{\omega_a,\omega_{a'},\varepsilon_2}\left[\exp\left(-\frac{R\left(p_a^*||q_a^\theta\right)\left(a'+\omega_{a'}\right)\left(1-\theta-\varepsilon_2\right)}{k}\right)\right] \\ \geq &\ c\left(a\right)-c\left(a'\right).\end{aligned}$$

Since $\mathbb{E}_{\omega_a,\varepsilon_2}\left[\left|\left(a+\omega_a\right)\left(1-\theta-\varepsilon_2\right)\right|\right]$ and $\mathbb{E}_{\omega_{a'},\varepsilon_2}\left[\left|\left(a'+\omega_{a'}\right)\left(1-\theta-\varepsilon_2\right)\right|\right]$ are finite, by Jensen inequality (see 10.2.6 in Dudley, 2018 for the version that applies here) the LHS is lower in the second case, and we obtain the desired conclusion.

Example 3

The condition for not switching from action 0 to an action a with $a \geq 0$ in a Berk-Nash equilibrium is

$$p_a^*(s \leq a)\left(\mathbb{E}_{p_a^*}(v) - a\right) \leq 0.$$

By Proposition 1.4.2 in Dupuis and Ellis (2011), the condition for not switching from action 0 to an action a with $a \geq 0$ in a k-robust equilibrium is

$$-\frac{k}{R\left(p_a^* || q_a\right)} \log \int_{\mathbb{R}} \int_{\mathbb{R}} \exp\left(-\frac{R\left(p_a^* || q_a\right)}{k}\left[v - a\right] \mathbb{I}_{[0,a]}(s)\right) \mathrm{d}p_a^*(s)\, \mathrm{d}p_a^*(v) \leq 0.$$

Since $\mathbb{E}_{p_a^*}(|v|) < \infty$, by Jensen inequality (see 10.2.6 in Dudley, 2018 for the version that applies here) the LHS is lower in the second case, and we obtain the desired conclusion.

Example of Dynamic Incosistency

Example 4. *To provide a simple illustration of dynamic inconsistency, we consider the two-period truncated problem. An urn contains black (b) or green (g) balls. At each time t, the DM is asked to bet 1 dollar on the color of the ball drawn from the urn or to opt-out (o) and observe the drawn with a payoff of 0.5. That is, $u(a, y) = I_{\{a=y\}}$ if $a \in \{b, g\}$ and $u(o, y) = 0.6$. Suppose that at period 0, the level of concern for misspecification is $\Lambda(h_0) = 0$ and that the agent considers two models, q, q', that assign respectively probability 0.7 and 0.3 to the black ball, independently of the agent action. The prior μ assigns equal probability to these two models.*

To illustrate the possibility of dynamic inconsistencies of a forward-looking agent, we introduce a discount factor equal to $\delta = 0.9$ and suppose that $\Lambda((0, b)) = 2$. In this case, at time 0, the decision maker would like to commit to the following plan: opt-out in the first period and then, in the second period, bet on the color of the ball drawn in the first period. However, the increase in concern for misspecification created by the observation of the black drawn makes this plan not feasible: at history $(0, b)$, the

agent will opt out again.

Bibliography

Aliprantis, C. and K. Border (2013). *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer-Verlag.

Anderson, E. W., L. P. Hansen, and T. J. Sargent (2003). "A quartet of semigroups for model specification, robustness, prices of risk, and model detection". In: *Journal of the European Economic Association* 1.1, pp. 68–123.

Anderson, R. M., H. Duanmu, A. Ghosh, and M. A. Khan (2022). "On Existence of Berk-Nash Equilibria in Misspecified Markov Decision Processes with Infinite Spaces". In: *arXiv preprint arXiv:2206.08437*.

Angeletos, G.-M., Z. Huo, and K. A. Sastry (2021). "Imperfect macroeconomic expectations: Evidence and theory". In: *NBER Macroeconomics Annual* 35.1, pp. 1–86.

Arrow, K. and J. Green (1973). "Notes on Expectations Equilibria in Bayesian Settings". Working Paper No. 33, Stanford University.

Aubin, J.-P. and A. Cellina (2012). *Differential inclusions: set-valued maps and viability theory*. Vol. 264. Springer Science & Business Media.

Ba, C. (2022). "Robust Model Misspecification and Paradigm Shifts". In: *arXiv preprint arXiv:2106.12727*.

Banerjee, A. V., S. Chassang, S. Montero, and E. Snowberg (2020). "A theory of experimenters: Robustness, randomization, and balance". In: *American Economic Review* 110.4, pp. 1206–30.

Barillas, F., L. P. Hansen, and T. J. Sargent (2009). "Doubts or variability?" In: *journal of economic theory* 144.6, pp. 2388–2418.

Battigalli, P. (1987). "Comportamento razionale ed equilibrio nei giochi e nelle situazioni sociali". In: *Unpublished undergraduate dissertation, Bocconi University, Milano*.

Battigalli, P., S. Cerreia-Vioglio, F. Maccheroni, M. Marinacci, and T. Sargent (2022). "A framework for the analysis of self-confirming policies". In: *Theory and Decision*, pp. 1–58.

Battigalli, P., A. Francetich, G. Lanzani, and M. Marinacci (2019). "Learning and self-confirming long-run biases". In: *Journal of Economic Theory* 183, pp. 740–785.

Benaim, M., J. Hofbauer, and S. Sorin (2005). "Stochastic approximations and differential inclusions". In: *SIAM Journal on Control and Optimization* 44.1, pp. 328–348.

Berk, R. H. (1966). "Limiting behavior of posterior distributions when the model is incorrect". In: *The Annals of Mathematical Statistics*, pp. 51–58.

Bohren, J. A. and D. N. Hauser (2021). "Learning with heterogeneous misspecified models: Characterization and robustness". In: *Econometrica* 89.6, pp. 3025–3077.

Cabrales, A., O. Gossner, and R. Serrano (2013). "Entropy and the value of information for investors". In: *American Economic Review* 103.1, pp. 360–77.

Casella, G. and R. L. Berger (2021). *Statistical inference*. Cengage Learning.

Cerreia-Vioglio, S., L. P. Hansen, F. Maccheroni, and M. Marinacci (2022). "Making decisions under model misspecification". In.

Cerreia-Vioglio, S., F. Maccheroni, M. Marinacci, and L. Montrucchio (2013a). "Ambiguity and robust statistics". In: *Journal of Economic Theory* 148.3, pp. 974–1049.

— (2013b). "Classical subjective expected utility". In: *Proceedings of the National Academy of Sciences* 110.17, pp. 6754–6759.

Chew, H. S. and J. S. Sagi (2006). "Event exchangeability: Probabilistic sophistication without continuity or monotonicity". In: *Econometrica* 74.3, pp. 771–786.

Cho, I.-K. and K. Kasa (2015). "Learning and model validation". In: *The Review of Economic Studies* 82.1, pp. 45–82.

Clarida, R., J. Gali, and M. Gertler (2000). "Monetary policy rules and macroeconomic stability: evidence and some theory". In: *The Quarterly journal of economics* 115.1, pp. 147–180.

Cogley, T. and T. J. Sargent (2005). "The conquest of US inflation: Learning and robustness to model uncertainty". In: *Review of Economic dynamics* 8.2, pp. 528–563.

De Filippis, R., A. Guarino, P. Jehiel, and T. Kitagawa (2022). "Non-Bayesian updating in a social learning experiment". In: *Journal of Economic Theory* 199, p. 105188.

Dean, M. and P. Ortoleva (2017). "Allais, Ellsberg, and preferences for hedging". In: *Theoretical Economics* 12.1, pp. 377–424.

— (2019). "The empirical relationship between nonstandard economic behaviors". In: *Proceedings of the National Academy of Sciences* 116.33, pp. 16262–16267.

Dekel, E., J. C. Ely, and O. Yilankaya (2007). "Evolution of preferences". In: *The Review of Economic Studies* 74.3, pp. 685–704.

Denti, T. and L. Pomatto (2022). "Model and predictive uncertainty: A foundation for smooth ambiguity preferences". In: *Econometrica* 90.2, pp. 551–584.

Diamond, P. A. (1982). "Aggregate demand management in search equilibrium". In: *Journal of political Economy* 90.5, pp. 881–894.

Dillenberger, D., J. S. Lleras, P. Sadowski, and N. Takeoka (2014). "A theory of subjective learning". In: *Journal of Economic Theory* 153, pp. 287–312.

Dudley, R. M. (2018). *Real Analysis and Probability*. Chapman and Hall/CRC.

Dunford, N. and J. T. Schwartz (1988). *Linear operators, part 1: general theory*. Vol. 10. John Wiley & Sons.

Dupuis, P. and R. S. Ellis (2011). *A weak convergence approach to the theory of large deviations*. Vol. 902. John Wiley & Sons.

Enke, B. and F. Zimmermann (2019). "Correlation neglect in belief formation". In: *The Review of Economic Studies* 86.1, pp. 313–332.

Epstein, L. G. and S. Ji (2022). "Optimal learning under robustness and time-consistency". In: *Operations Research* 70.3, pp. 1317–1329.

Esponda, I. (2008). "Behavioral equilibrium in economies with adverse selection". In: *American Economic Review* 98.4, pp. 1269–91.

Esponda, I. and D. Pouzo (2016). "Berk–Nash equilibrium: A Framework for Modeling Agents with Misspecified Models". In: *Econometrica* 84.3, pp. 1093–1130.

Esponda, I., D. Pouzo, and Y. Yamamoto (2021a). "Asymptotic behavior of Bayesian learners with misspecified models". In: *Journal of Economic Theory* 195, p. 105260.

— (2021b). "Corrigendum: Asymptotic behavior of Bayesian learners with misspecified models". In: *Journal of Economic Theory* 195, p. 105260.

Esponda, I., E. Vespa, and S. Yuksel (2022). *Mental models and learning: The case of base-rate neglect*. Tech. rep.

Fishburn, P. C. (1970). *Utility theory for decision making*. Wiley.

Foster, D. P. and H. P. Young (2003). "Learning, hypothesis testing, and Nash equilibrium". In: *Games and Economic Behavior* 45.1, pp. 73–96.

Foutz, R. V. and R. Srivastava (1977). "The performance of the likelihood ratio test when the model is incorrect". In: *The annals of Statistics*, pp. 1183–1194.

Frick, M., R. Iijima, and Y. Ishii (2023). "Belief Convergence under Misspecified Learning: A Martingale Approach". In: *Review of the Economic Studies* 90, pp. 781–814.

Fudenberg, D. and D. M. Kreps (1993). "Learning mixed equilibria". In: *Games and economic behavior* 5.3, pp. 320–367.

Fudenberg, D. and G. Lanzani (2022). "Which misperceptions persist?" In: *Theoretical Economics, Forthcoming*.

Fudenberg, D., G. Lanzani, and P. Strack (2021). "Limit points of endogenous misspecified learning". In: *Econometrica* 89.3, pp. 1065–1098.

— (2022a). "Pathwise Concentration Bounds for Misspecified Bayesian Beliefs". In: *Available at SSRN 3805083*.

— (2022b). "Selective Memory Equilibrium". In: *Available at SSRN 4015313*.

Fudenberg, D. and D. K. Levine (1993). "Self-confirming equilibrium". In: *Econometrica*, pp. 523–545.

— (1995). “Consistency and cautious fictitious play”. In: *Journal of Economic Dynamics and Control* 19.5-7, pp. 1065–1089.

Gagnon-Bartsch, T., M. Rabin, and J. Schwartzstein (2022). “Channeled Attention and Stable Errors”. In.

Ghirardato, P. and M. Marinacci (2002). “Ambiguity made precise: A comparative foundation”. In: *Journal of Economic Theory* 102.2, pp. 251–289.

Giacomini, R., V. Skreta, and J. Turén (2015). “Models, inattention and expectation updates”. In.

Gilboa, I., S. Minardi, and L. Samuelson (2020). “Theories and cases in decisions under uncertainty”. In: *Games and Economic Behavior* 123, pp. 22–40.

Gilboa, I. and D. Schmeidler (1989). “Maxmin expected utility with non-unique prior”. In: *Journal of Mathematical Economics* 18.2, pp. 141–153.

Gul, F. and W. Pesendorfer (2014). “Expected uncertain utility theory”. In: *Econometrica* 82.1, pp. 1–39.

Hall, P. and C. C. Heyde (2014). *Martingale limit theory and its application*. Academic press.

Hansen, L. P. and T. J. Sargent (2001). “Robust control and model uncertainty”. In: *American Economic Review* 91.2, pp. 60–66.

— (2007). “Recursive robust estimation and control without commitment”. In: *Journal of Economic Theory* 136.1, pp. 1–27.

— (2011). “Robustness”. In: *Robustness*. Princeton university press.

— (2022). “Risk, Ambiguity, and Misspecification: Decision Theory, Robust Control, and Statistics”. In.

Hansen, L. P., T. J. Sargent, G. Turmuhambetova, and N. Williams (2006). “Robust control and model misspecification”. In: *Journal of Economic Theory* 128, pp. 45–90.

Hausman, J. A. (1978). “Specification tests in econometrics”. In: *Econometrica: Journal of the econometric society*, pp. 1251–1271.

He, K. and J. Libgober (2022). “Evolutionarily stable (mis) specifications: Theory and applications”. In.

Kallenberg, O. (1973). “Random measures”. In: *Random Measures*. De Gruyter.

Karni, E. and M.-L. Vierø (2013). ““Reverse Bayesianism": A choice-based theory of growing awareness”. In: *American Economic Review* 103.7, pp. 2790–2810.

Klibanoff, P., M. Marinacci, and S. Mukerji (2005). “A smooth model of decision making under ambiguity”. In: *Econometrica* 73.6, pp. 1849–1892.

— (2009). “Recursive smooth ambiguity preferences”. In: *Journal of Economic Theory* 144.3, pp. 930–976.

Kreps, D. (1988). *Notes On The Theory Of Choice*. Westview Press.

Liebman, J. B. and R. J. Zeckhauser (2004). “Schmeduling”. In.

Liese, F. and I. Vajda (1987). *Convex statistical distances*. Vol. 95. Teubner.

Maccheroni, F., M. Marinacci, and A. Rustichini (2006a). “Ambiguity aversion, robustness, and the variational representation of preferences”. In: *Econometrica* 74.6, pp. 1447–1498.

— (2006b). “Dynamic variational preferences”. In: *Journal of Economic Theory* 128.1, pp. 4–44.

Maenhout, P. J. (2004). “Robust portfolio rules and asset pricing”. In: *Review of financial studies* 17.4, pp. 951–983.

Mu, X., L. Pomatto, P. Strack, and O. Tamuz (2021). “Monotone additive statistics”. In: *arXiv preprint arXiv:2102.00618*.

Nyarko, Y. (1991). “Learning in Mis-specified Models and the Possibility of Cycles”. In: *Journal of Economic Theory* 55.2, pp. 416–427.

Ok, E. A. (2011). “Real analysis with economic applications”. In: *Real Analysis with Economic Applications*. Princeton University Press.

Ortoleva, P. (2012). “Modeling the change of paradigm: Non-Bayesian reactions to unexpected news”. In: *American Economic Review* 102.6, pp. 2410–36.

Parthasarathy, K. R. (2005). *Probability Measures on Metric Spaces*. American Mathematical Soc.

Pathak, P. et al. (2002). “Notes on robust portfolio choice”. In: *unpublished paper, Harvard University* 3, p. 14.

Primiceri, G. E. (2005). "Time varying structural vector autoregressions and monetary policy". In: *The Review of Economic Studies* 72.3, pp. 821–852.

Rabin, M. (2002). "Inference by believers in the law of small numbers". In: *The Quarterly Journal of Economics* 117.3, pp. 775–816.

Rees-Jones, A. and D. Taubinsky (2020). "Measuring "schmeduling"". In: *The Review of Economic Studies* 87.5, pp. 2399–2438.

Reny, P. J. (1999). "On the existence of pure and mixed strategy Nash equilibria in discontinuous games". In: *Econometrica* 67.5, pp. 1029–1056.

Robatto, R. and B. Szentes (2017). "On the biological foundation of risk preferences". In: *Journal of Economic Theory* 172, pp. 410–422.

Royden, H. L. and P. Fitzpatrick (1988). *Real analysis*. Vol. 32. Macmillan New York.

Rudin, W. (1970). *Real and Complex Analysis*. McGraw-Hill.

Sargent, T. J. (1999). "The conquest of American inflation". In: *The Conquest of American Inflation*. Princeton University Press.

— (2008). "Evolution and intelligent design". In: *American Economic Review* 98.1, pp. 5–37.

Savage, L. J. (1954). *The foundations of statistics*. Courier Corporation.

Schwartzstein, J. and A. Sunderam (2021). "Using models to persuade". In: *American Economic Review* 111.1, pp. 276–323.

Serfozo, R. (1982). "Convergence of Lebesgue integrals with varying measures". In: *Sankhyā: The Indian Journal of Statistics, Series A*, pp. 380–402.

Simon, J. (2020). *Continuous Functions*. John Wiley & Sons.

Sims, C. A. and T. Zha (2006). "Were there regime switches in US monetary policy?" In: *American Economic Review* 96.1, pp. 54–81.

Siniscalchi, M. (2011). "Dynamic choice under ambiguity". In: *Theoretical Economics* 6.3, pp. 379–421.

Sobel, J. (1984). "Non-linear prices and price-taking behavior". In: *Journal of Economic Behavior & Organization* 5.3-4, pp. 387–396.

Spiegler, R. (2020). "Can agents with causal misperceptions be systematically fooled?" In: *Journal of the European Economic Association* 18.2, pp. 583–617.

Strzalecki, T. (2011). "Axiomatic foundations of multiplier preferences". In: *Econometrica* 79.1, pp. 47–73.

Sweeting, T. (1986). "On a converse to Scheffé's theorem". In: *The Annals of Statistics* 14.3, pp. 1252–1256.

Tversky, A. and D. Kahneman (1971). "Belief in the law of small numbers." In: *Psychological bulletin* 76.2, p. 105.

Varadarajan, V. S. (1958). "On the convergence of sample probability distributions". In: *Sankhyā: The Indian Journal of Statistics (1933-1960)* 19.1/2, pp. 23–26.

Vuong, Q. H. (1989). "Likelihood ratio tests for model selection and non-nested hypotheses". In: *Econometrica*, pp. 307–333.

Wakker, P. P. (2013). *Additive representations of preferences: A new foundation of decision analysis*. Vol. 4. Springer Science & Business Media.

Willard, S. (2012). *General topology*. Courier Corporation.

Appendix A

Correlation Made Simple

A.1 Introduction

Correlation between risky alternatives can play a significant role in decisions. First, it may be relevant because the Decision Maker (henceforth DM) cares about what she would have received had she chosen differently, a channel emphasized by regret theory. For example, an agent who has decided not to invest part of her resources in stocks the day before a press release by the Fed may be better off if the market's effect is negative since she does not suffer for the foregone opportunity.

Second, the correlation structure can determine the attention and weight that the DM allocates to the various contingencies, as emphasized by salience theory. For example, when deciding whether to purchase comprehensive car insurance, the (unlikely) event in which the car is destroyed in a crash may disproportionately attract the DM's attention due to the vast difference between the two alternatives' consequences. In this chapter, we study these possibilities from an axiomatic perspective.

We provide a simple axiomatization for a general class of correlation-sensitive preferences. The motivation is two-fold. First, we show that our general framework nests the recent models that highlight the role of correlation (see, e.g., Bordalo, Gennaioli, and Shleifer, 2012, henceforth BGS, and Koszegi and Szeidl, 2013) as particular cases so that we can characterize them in terms of additional testable axioms. The study of these axioms lets us understand better where they depart from

the preexisting theories. Second, we use this axiomatization to provide new insights into the difference between classical models for which correlation is relevant (see Bell, 1982, Loomes and Sugden, 1982, Fishburn, 1989) and the benchmark model for choice under risk, expected utility (henceforth EU).

We accomplish these goals by taking a different route than those followed in the usual axiomatizations of correlation-sensitive preferences.[1] We represent the preferences of the DM in the space of lotteries. In doing so, we face a complication: when the correlation between alternatives matters, binary relations over lotteries are not sufficiently rich as modeling tools. To see why, suppose that we have the two lotteries $p = \left(10, \frac{1}{3}; 5, \frac{1}{3}; 0, \frac{1}{3}\right)$ and $q = \left(10, \frac{1}{3}; 4, \frac{1}{3}; 1, \frac{1}{3}\right)$, and consider the following two possible correlation structures:

π	1	4	10
0	0	1/3	0
5	0	0	1/3
10	1/3	0	0

π'	1	4	10
0	0	0	1/3
5	1/3	0	0
10	0	1/3	0

Both the joint distributions π and π' have marginal distributions p and q. However, we will see that a salience-sensitive DM may strictly prefer p under the first correlation structure (driven by the salient realization $(10, 1)$) and q under the second correlation structure (driven by the salient realization $(0, 10)$). Therefore, the classical approach of describing the DM's tastes using a binary relation over lotteries is not viable since the DM cannot rank p and q without additional information about their joint distribution. Indeed, applied researchers (e.g., Smith 1996, Braun and Muermann, 2004, Filiz-Ozbay and Ozbay, 2007) have shown that in various economically significant situations as auctions, insurance decisions, and health interventions, the correlation between lotteries impacts choices.

Instead of using a binary relation, we use the preference set concept introduced by Fishburn (1990a):[2] Given a fixed set of possible outcomes X, tastes are represented

[1] See, e.g., Fishburn (1989), Sugden (1993), and Diecidue and Somaundaram (2017). All these papers represent the preferences as a binary relation over acts à la Savage.

[2] Fishburn (1990a) introduces the concept of preference sets for intransitive preferences over multi-

by a preference set $\Pi \subseteq \Delta(X \times X)$, with the following interpretation. The DM contemplates a joint distribution π over $X \times X$. Facing this joint lottery, the DM decides if, *given the marginals and the correlation structure,* she prefers to be paid according to the realized row or column outcome.[3] Then, we say that π belongs to the preference set Π if and only if the DM prefers to be paid according to the row outcome. In our previous example, we have $\pi \in \Pi$, and $\pi' \notin \Pi$.

There are several motivations for this modeling choice. On a theoretical side, it avoids introducing an ancillary state-space and provides a clear comparison with expected utility. If we want to test the theory, having an axiomatization for the case of choice under risk, instead of one for acts defined over a state space in which probabilities are not specified, allows us to disentangle violation of the axioms at the cornerstone of our correlation sensitive theory from the ubiquitous failures in formulating a unique, coherent probability measure over the states of the world.

The second motivation comes from our salience theory application. Indeed, BGS define their preferences on the joint distributions of two alternative random variables, and the correlation is part of the data exactly as under our proposed approach. Moreover, the subsequent experimental papers consider choices between lotteries, where the only state space is the one *defined* as the space of all the possible joint realizations of the two lotteries under scrutiny.[4] Therefore, axioms stated in terms of joint lotteries are more natural to map into the BGS model, and they can be directly challenged by the existing experimental evidence on the model. Finally, under the alternative state-space formulation, the characterization of the salience properties postulated by BGS is much more demanding in terms of the underlying state space's structural properties.

We first identify three axioms on the preference set Π necessary and sufficient to

attribute products and applies it to choices between acts in Fishburn (1990b). To the best of our knowledge, this is the first work in which preference sets are used to axiomatize preferences under risk.

[3]As we can always represent a joint distribution in the tabular form used above, we will refer to the first and second marginal respectively as the row and column marginals.

[4]For the experimental tests of salience theory, see Dertwinkel-Kalt and Köster (2020), Frydman and Mormann, (2017), Konigsheim, Lukas, and Noth (2019), Dertwinkel-Kalt, Frey, and Koster (2021) Nielsen, Sebald, and Sorensen (2021).

obtain a representation for correlation-sensitive preferences. This representation includes regret and salience theory as particular cases. These axioms are Completeness, Strong Independence, and Archimedean Continuity, and they are equivalent to the correlation-sensitive representation

$$\pi \in \Pi \Leftrightarrow \sum_{x,y} \phi(x,y)\,\pi(x,y) \geq 0$$

where ϕ is skew-symmetric. Here, $\phi(x,y)$ corresponds to how much the joint realization (x,y) contributes in favor of the row marginal. That is, we have $\phi(x,y) \geq 0$ if and only x is preferred to y, and larger values imply a comparison more favorable to x. Under expected utility, $\phi(x,y)$ reduces to the separable form $u(x) - u(y)$, but more generally (e.g., in the salience model), the two components are entangled. Indeed, how much attention an outcome x attracts may depend on how much it contrasts with the counterfactual realization y. The skew symmetry of ϕ means that the "row" and "column" labeling are irrelevant: $\phi(x,y) = -\phi(y,x)$, so that the contribution of the joint realization (x,y) in favor of the row component is equal to the contribution of (y,x) to the column component. We show that these axioms are mild relaxations of their more familiar counterparts for binary relations and that if Transitivity is added, the representation reduces to EU.

After weakening the EU axioms to allow for this more general correlation-sensitive representation, we look for the additional axioms needed to characterize the particular case of salience theory. BGS's salience model provides a theory of choice under risk based on few psychological properties of salience detection: Ordering, Diminishing Sensitivity, and Weak Reflexivity. Most importantly, Ordering prescribes that joint realizations in which the two components are farther apart are overweighted. In addition, there is Diminishing Sensitivity to the differences between the components as their absolute values increase. At the same time, Weak Reflexivity can be loosely paraphrased as the requirement that the salience ranking between two joint realizations that only involve gains remains the same if all the gains are transformed into losses of the same size.

A payoff of our preference sets setup is that it allows us to state and characterize the testable versions of these properties in a straightforward manner. We also find that Ordering is the property that brings salience theory outside the prospect theory realm—instead, Diminishing Sensitivity and Weak Reflexivity combined amount to the usual risk-aversion in gains, risk-loving in loss property featured by prospect theory. We then characterize the salience model as the result of the Ordering, Diminishing Sensitivity, and Weak Reflexivity axioms combined with continuity and monotonicity requirements.

We also provide a partial solution to the problem of choice between multiple alternatives. A DM with correlation-sensitive preferences may not have an alternative that is weakly preferred to all the others when facing a set of at least three options. However, we prove that an optimal stochastic choice rule always exists.

Related Literature This chapter belongs to the literature studying the axiomatization of correlation-sensitive models of choice. This literature starts with the classical works of Fishburn (1989), Sugden (1993), and Quiggin (1994). Recently, Diecidue and Somasundaram (2017) significantly improve the regret model's previous representation, providing an axiomatization that delivers a continuous regret function on an arbitrary finite state space. Their main conceptual contribution is to single out the axioms for the more restrictive version of regret theory initially formulated by Loomes and Sugden (1982) and separate the edonic utility from the regret function. In this sense, their work is complementary to ours. In the first part of the chapter, we want to axiomatize the more general form of correlation-sensitive preferences to characterize later regret theory and salience theory as particular cases of this model.

Fishburn (1990b) uses preference sets to provide an axiomatization of the Skew-Symmetric Additive (SSA) model. On a technical side, the object on which the preferences are defined is different: Fishburn defines the preference sets as subsets of the space of acts with two outcomes, whereas we focus on joint distribution over outcomes. Notice that by letting the preference sets being a subset of the multivariate acts, Fishburn (1990b) faces the general disadvantages discussed above: potential

confusion with ambiguity aversion, axioms that are sufficient for the representation but not necessary, more difficult comparison with EU, and a more relevant departure from the version of the model that has been experimentally tested.[5] These disadvantages become even more relevant in our salience theory application: first, the additional properties characterizing salience theory as a particular case become much more involved under the act formulation. Second, generalizations that build on our axiomatization to combine salience theory for consumption and risk (see Köster, 2021) cannot be conciliated with the "structure axiom" needed in Fishburn (1990b), therefore limiting the scope of his axiomatization.

Fishburn (1982) axiomatizes the class of SSB preferences over the space of lotteries. With these preferences, each alternative's realization has a value that depends on *all* the possible realizations of the other alternative. When restricted to the comparison between *independent* lotteries, the two models coincide.[6] In this sense, Theorem 4 provides an alternative set of axioms for the SSB model. More importantly, the two models are highly different in their predictions about correlation. Fishburn (1982) explicitly rules out any correlation effect, and so it excludes the salience model, where significantly different joint realizations attract the attention of the DM and imposes an awkward structure on the regret model.[7] Farther afield, the quadratic preferences of Chew, Epstein, and Segal (1991) and the reference-dependent model of Koszegi and Rabin (2007) also use of a joint evaluation of outcomes, although they do not allow for correlation sensitivity and satisfy Transitivity.

[5] Among other things, the state space has to be atomless, a property at odds with the small finite set of joint realizations used as the state space in BGS. Moreover, the use of atomless state space in the classical axiomatizations of correlation-sensitive preferences is particularly unsatisfactory since it is a direct consequence of what Fishburn (1990b) calls axiom P6*. This axiom is made for technical convenience but is not necessary for the representation. Therefore, such a richness of the space is not an intrinsic feature of the model but more the result of a technically convenient assumption.

[6] However, when Transitivity is imposed in the two models, the implications are different. Since by definition of the domain of preferences in the SSB model Transitivity can only be imposed on independent distributions, SSB reduces to the weighted utility model of Chew (1983), see Theorem 3 in Fishburn (1983). Instead, when I impose Transitivity of the marginal regardless of the correlation structure, I obtain the Expected Utility model, as in Bikhchandani and Segal (2011). I thank Chew Soo Hong for pushing me to explore this additional difference.

[7] For example, the form of regret compatible with the SSB model requires that when choosing not to bet on a horse in a race, the DM must feel regret for the foregone possibility of a significant payoff, regardless of whether the horse wins the race.

This work is the first to axiomatize the salience theory of choice under risk. Ellis and Masatioglu (2021) provide an axiomatization of the salience theory of consumption (Bordalo, Gennaioli, and Shleifer, 2013). They focus on the rank-dependent version of the salience model, while we focus on the continuous version. Herweg and Muller (2021) provide a comparison between salience and regret theory, arguing that the former can be interpreted as a particular case of the latter, but they do not identify the axioms underlying the representation.

Outline The rest of the chapter is structured as follows. Section A.2 introduces preference sets. Then, in Section A.3 we describe the *weakening* of EU that is necessary to capture sensitivity to correlation, while in Section A.4 we provide the *additional* axioms characterizing salience theory. All the proofs of the results in the main text are in Supplementary Appendix A.6. Supplementary Appendix A.7 establishes the formal connection between axioms stated for preference sets and their counterparts in terms of binary relations. Supplementary Appendix A.8 extends the model to choice from nonbinary subsets. Finally, Supplementary Appendix A.9 studies the rank-based version of salience theory.

A.2 Preference Sets

Let X be an arbitrary nonempty set of outcomes (or prizes), and denote as $\Delta(X \times X)$ the set of (joint) probability measures over $X \times X$ with finite support. We model the DM preferences by a subset Π (called preference set) of $\Delta(X \times X)$. The interpretation is that the DM faces a $\pi \in \Delta(X \times X)$, and she has to decide whether to be paid according to the row or column outcome. Then, we say that $\pi \in \Pi$ if and only if she (weakly) prefers to be paid according to the row outcome. The fact that the knowledge of the marginal π_1 and π_2 may be insufficient to determine whether $\pi \in \Pi$ is the deviation from the standard paradigm of rational choice.

A.2.1 Eliciting Preference Sets

Here is a roadmap of how to test axioms imposed on the preference set. The DM faces a finite-support joint distribution π over prizes that a table can summarize:

π	y_1	...	y_m
x_1	π_{11}	...	π_{1m}
...	...	...	...
x_n	π_{n1}	...	π_{nm}

That is, the DM knows that every pair of outcomes (x_i, y_j) realizes with probability π_{ij}. Then, given the correlation structure between the two alternatives, the subject chooses between being paid according to the row prizes (the x's) or the column prizes (the y's). If she chooses to be paid according to the rows (resp. the columns), if outcome (x_i, y_j) realizes she gets x_i (resp. y_j) regardless of the value of y_j (resp. x_i).[8]

A joint distribution belongs to the preference set if, when faced, the DM chooses to be paid according to the row prizes. The typical axioms we impose on preference sets have the form "if $\pi \in \Pi$ then π' belongs to Π," where π' has some particular relation with π.

A.2.2 Preference Sets and Binary Relations

For every joint distribution $\pi \in \Delta(X \times X)$, we denote as $\pi_1 \in \Delta(X)$ and $\pi_2 \in \Delta(X)$, respectively, the row and column marginals of π. Formally:

$$\pi_1(x) = \sum_{y \in X} \pi(x, y) \text{ and } \pi_2(y) = \sum_{x \in X} \pi(x, y).$$

Notice that a binary relation $\succsim$ over marginal distributions induces a *unique* preference set $\Pi_{\succsim}$ that contains a joint distribution if and only if the row marginal is preferred to the column according to $\succsim$.

[8] Our theory is silent about the information revealed to the subject after a joint outcome (x, y) is drawn. One may expect that the behavior may differ, whether only the component paid out to the DM or the joint realization is revealed.

Definition 12. The preference set $\Pi_{\succsim}$ induced by a binary relation $\succsim$ is defined as

$$\pi \in \Pi_{\succsim} \Leftrightarrow \pi_1 \succsim \pi_2.$$

It is easy to see that two different binary relations induce different preference sets, so no information is lost by describing the DM's tastes using preference sets rather than binary relations. Also, every preference set induces a (possibly incomplete) binary relation over marginal distributions.

Definition 13. The binary relation $\succsim^{\Pi}$ induced by a preference set Π is defined as

$$p \succsim^{\Pi} q \Leftrightarrow \left(\forall \pi \in \Delta\left(X \times X\right) : \left(\pi_1, \pi_2\right) = \left(p, q\right),\ \pi \in \Pi\right).$$

Requiring $p \succsim^{\Pi} q$ ensures that all the joint distributions with those marginals are in the preference set (i.e., p has to be preferred to q regardless of their correlation structure). Of course, when $\succsim^{\Pi}$ is complete, it describes the DM's tastes fully. However, $\succsim^{\Pi}$ may not be complete for a correlation-sensitive DM. In this case, the patterns of behavior that can be described using preference sets are much richer than those for binary relations.[9]

A.3 General Representation Theorem

We first define the general form of risk preferences we are interested in. Recall that a function $\phi : X \times X \to \mathbb{R}$ is skew symmetric if $\phi\left(x, y\right) = -\phi\left(y, x\right)$ for all $x, y \in X$.

Definition 14. A preference set Π admits a correlation-sensitive representation if there exists a skew-symmetric $\phi : X \times X \to \mathbb{R}$ such that for all $\pi \in \Delta(X \times X)$

$$\pi \in \Pi \Leftrightarrow \sum_{x,y} \phi\left(x, y\right) \pi\left(x, y\right) \geq 0. \tag{A.1}$$

[9]Lemma 15 in the Supplementary Appendix shows that for every binary relation $\succeq$, the binary relation $\succsim^{\Pi_{\succeq}}$ coincides with $\succeq$. A weaker notion of $\succsim^{\Pi}$ would have replaced "for all π" with "for some π" in its definition. Proposition 8 shows that our definition is more fruitful.

To better understand this representation, it is helpful to compare it with expected utility. Let $\pi \in \Delta(X \times X)$. Under EU, there exists a utility function u such that

$$\pi_1 \succsim \pi_2 \quad \Leftrightarrow \quad \sum_x u(x)\,\pi_1(x) \geq \sum_y u(y)\,\pi_2(y) \tag{A.2}$$

$$\Leftrightarrow \quad \sum_{x,y} (u(x) - u(y))\,\pi(x,y) \geq 0. \tag{A.3}$$

Given these equivalences, the difference between EU and the correlation-sensitive representation can be described in the following way. In principle, when contemplating a joint lottery π, two algorithmic procedures can determine according to which component to be paid. The first algorithm is the following: (i) Take marginal π_1. Consider the utility obtained under each realization. Aggregate these utilities according to the probability measure π_1 to get a "score" $U(\pi_1) = \sum_x u(x)\,\pi_1(x)$. Note that this score is *independent of* π_2. (ii) Follow the same procedure for marginal π_2. (iii) Compare these scores obtained for the two alternatives, and choose to be paid according to the row outcome if and only if $U(\pi_1) \geq U(\pi_2)$. There is no role for correlation between the two marginal distributions under this procedure. This procedure consists of a *comparison of aggregations*, and in the case of EU is given by (A.2).

Alternatively, one may consider the following procedure: (i) Take a possible joint realization (x, y). Compare the two prizes and give a score $\phi(x, y)$, representing a combination of how much x is preferred to y and the attention diverted to that realization, with 0 meaning indifference or zero attention. (ii) Do the same for every joint realization. (iii) Aggregate all these comparisons according to the probability measure π obtaining $\Phi(\pi) = \sum_{x,y} \pi(x,y)\,\phi(x,y)$. (iv) Choose to be paid according to the row outcome if and only if $\Phi(\pi) \geq 0$.

This *aggregation of comparisons* allows for correlation to matter. It is the kind of reasoning that characterizes both regret and salience-sensitive DMs**,** and for EU it corresponds to line (A.3). The pioneering works by Bell (1982) and Loomes and Sugden (1982) already recognize the descriptive and normative value of this procedure. However, under expected utility, aggregation of comparisons reduces to $\phi(x, y) =$

$u(x) - u(y)$, which makes correlation irrelevant because, for an EU agent, the value of receiving x is $u(x)$ independent of the realization of the counterfactual. Therefore, in this case, the two algorithms reach the same conclusion. We formalize this reasoning in the following definition.

Definition 15. A preference set Π admits an expected utility representation if there exists $u : X \to \mathbb{R}$ such that

$$\pi \in \Pi \iff \sum_{(x,y)\in X\times X} (u(x) - u(y))\, \pi(x,y) \geq 0. \tag{A.4}$$

Instead, our first step is to provide a set of axioms that characterize the general correlation-sensitive representation for a (possibly) nonseparable ϕ. We will call these axioms Completeness, Strong Independence, and Archimedean Continuity after the names of the standard axioms for binary relations they resemble. In Supplementary Appendix A.7, we show formally how each of the axioms for preference sets is a weakening of the original one that only applies to joint distributions and that they coincide when Transitivity is satisfied.

Before going further, a piece of notation is needed. Given $\pi \in \Delta(X \times X)$, we define its conjugate distribution $\bar{\pi}$ as

$$\forall (x,y) \in X \times X \quad \bar{\pi}(x,y) = \pi(y,x).$$

Therefore, the conjugate distribution is just a relabeling of the row and column outcomes into each other.

Axiom 11 (Completeness)**.** *For all* $\pi \in \Delta(X \times X)$

$$\pi \notin \Pi \Rightarrow \bar{\pi} \in \Pi.$$

Completeness is a minimal requirement about the rationality of the DM. If she prefers to be paid according to the column marginal when the joint distribution is π, she (weakly) prefers to be paid according to the row marginal after relabeling row

outcomes into column ones and vice-versa.

Given a preference set $\Pi \subseteq \Delta(X \times X)$, the strict preference set is defined as

$$\hat{\Pi} = \{\pi \in \Pi : \bar{\pi} \notin \Pi\}.$$

In words, a joint distribution π is in the strict preference set if the DM weakly prefers to be paid according to the row outcome (i.e., $\pi \in \Pi$), and she does not prefer to be paid according to the column outcome (i.e., $\bar{\pi} \notin \Pi$). It is the counterpart of the asymmetric part of a binary relation in the language of preference sets. We use the strict preference set in our second axiom. This axiom is a generalization to intransitive preferences of the standard principle of reduction for compound lotteries. If there are two joint distributions π and π' such that under each of them the DM prefers to be paid according to the row outcome, it then seems reasonable she prefers to be paid according to the row outcome even if the joint distribution that is going to be used is π with probability α and π' with probability $(1-\alpha)$. The preference is strict whenever one of the initial preferences is.

Axiom 12 (Strong Independence)**.** *For all $\pi, \pi' \in \Pi$, and all $\alpha \in (0,1)$*

$$\alpha\pi + (1-\alpha)\pi' \in \Pi.$$

Moreover, if $\pi' \in \hat{\Pi}$, then

$$\alpha\pi + (1-\alpha)\pi' \in \hat{\Pi}.$$

The difference between the previous axiom and the standard Strong Independence for binary relations can be understood in the setting of the Allais Paradox.

Example 5. *Recall that in the Allais paradox, the marginal distributions faced by the DM are*

$$\begin{aligned} p &= (2500, 0.33; 0, 0.01; z, 0.66) \\ q &= (2400, 0.34; z, 0.66) \end{aligned}$$

for $z \in \{0, 2400\}$. It is immediate to see that the Strong Independence axiom for binary relations implies that the choice of the DM does not depend on the particular value of z. The conclusion is more nuanced for our Strong Independence axiom. Indeed, the version of the Allais paradox in which the alternatives are independent corresponds to the joint distribution

$\pi_{ind,z}$	2400	z
2500	0.1122	0.2178
0	0.0034	0.0066
z	0.2244	0.4356

Here, Strong Independence formulated as above does not impose cross-restrictions for the behavior with different values of z. Therefore it accommodates the widely documented pattern that for most of the DMs, $\pi_{ind,0} \in \Pi$ and $\pi_{ind,2400} \notin \Pi$. Instead, the correlated version of the Allais paradox corresponds to the joint distribution

$\pi_{cor,z}$	2400	z
2500	0.33	0
0	0.01	0
z	0	0.66

Here, Strong Independence formulated as above has bite: it requires that $\pi_{cor,0} \in \Pi$ if and only if $\pi_{cor,2400} \in \Pi$. This is consistent with the empirical evidence in BGS, which shows how almost all the subjects do not change behavior when z changes in the correlated version of the problem. ▲

The example above highlights how preference sets allow us to disentangle two components of Strong Independence for binary relations: the sure-thing principle and probabilistic sophistication. The sure-thing principle is the part that is maintained by Strong Independence for preference sets, as realizations where the two alternatives pay the same are irrelevant for the evaluation. Instead, probabilistic sophistication requires that the marginal distributions are sufficient for the comparison, and there-

fore identical realizations can be canceled out even if they do not realize jointly. This probabilistic sophistication is not imposed by Strong Independence for preference sets.

Finally, we impose a weak continuity axiom guaranteeing the nonexistence of a joint distribution such that one marginal is "infinitely preferred" to the other.

Axiom 13 (Archimedean Continuity)**.** *For all* $\pi \in \hat{\Pi}$, $\pi' \notin \Pi$, *there exist* $\alpha, \beta \in (0,1)$ *such that*

$$\alpha\pi + (1-\alpha)\,\pi' \in \hat{\Pi} \text{ and } \beta\pi + (1-\beta)\,\pi' \notin \Pi.$$

The following theorem provides a representation of the preference sets satisfying these three axioms.

Theorem 4. *A preference set* Π *satisfies Completeness, Strong Independence, and Archimedean Continuity if and only if* Π *admits a correlation-sensitive representation. Moreover, the representing* ϕ *is unique up to a positive linear transformation.*

The theorem's proof combines the standard techniques used to prove the vN-M theorem with those used to deal with preference sets (see Fishburn 1990a) and intransitive preferences over acts (see Fishburn 1989). The theorem's importance stems from the fact that it connects a subset of the EU axioms to a general representation sensitive to the alternatives' correlation. Moreover, the value $\phi(x,y)$ has a cardinal interpretation as the contribution of the joint outcome (x,y) in favor of the row distribution. This cardinal role is the reason why the representing ϕ is unique up to a positive linear transformation.[10]

The representation still meaningfully restricts the pattern of behavior of the DM. To begin, if the joint distribution π is such that the row distribution dominates *realization by realization* the column distribution, then the joint distribution must be in the preference set, that is, if for all $(x,y) \in \operatorname{supp}\pi$, $\delta_{(x,y)} \in \Pi$, then $\pi \in \Pi$.[11] Moreover, Section A.3.1 shows that the conclusion can be strengthened from

[10] It may be interesting to explore a decision criterion that treats the two distributions asymmetrically because the row one is the status quo. We illustrate this possibility in Section A.4.5 where we compare the correlation sensitive representation to the reference-dependent model of Koszegi and Rabin (2007). Note that if the preference set admits a correlation-sensitive representation, by Theorem 1 of Fishburn (1982) the function ϕ is determined by the independent joint distributions.

[11] We denote as $\delta_{(x,y)}$ the joint lottery such that with probability one, the row outcome is x, and the column outcome is y.

realization by realization dominance to first-order stochastic dominance if the two lotteries under consideration are independent.

As the names of the previous axioms suggest, when the Transitivity axiom is added, the correlation-sensitive representation reduces to EU. Proposition 8 shows that this interpretation is correct. To do so, we need to translate Transitivity into the language of preference sets.

Axiom 14 (Transitivity)**.** *For all* $\pi, \chi, \rho \in \Delta(X \times X)$*, if* $\pi_2 = \chi_1$*,* $\rho_1 = \pi_1$*, and* $\rho_2 = \chi_2$*, then*

$$(\pi \in \Pi, \chi \in \Pi) \Rightarrow \rho \in \Pi.$$

The axiom has the following interpretation: Since $\pi \in \Pi$, $\pi_1 = \rho_1$ is preferred to $\pi_2 = \chi_1$ (given the correlation structure described by π). Since $\chi \in \Pi$, $\chi_1 = \pi_2$ is preferred to $\chi_2 = \rho_2$ (given the correlation structure described by χ). For Transitivity to hold, we then need that ρ_1 is preferred to ρ_2, i.e., $\rho \in \Pi$.

Example 6. *The following three joint distributions illustrate a typical failure of Transitivity due to salience sensitivity. By changing the correlation structure between alternatives, the realization with the most striking difference between outcomes changes, reversing the comparison between a fixed marginal and two similar alternatives. Let*

π	7	2
10	0	$\frac{1}{4}$
5	$\frac{1}{2}$	0
0	0	$\frac{1}{4}$

χ	8	1
7	1/2	0
2	0	1/2

ρ	8	1
10	1/4	0
5	0	1/2
0	1/4	0

we will see that for a salience sensitive DM, it is reasonable to have $\pi \in \Pi$*,* $\chi \in \Pi$*, and* $\rho \notin \Pi$*. Indeed, in* π *the large difference in the realization* $(10, 2)$ *tilts the evaluation in favor of the row marginal, and in* ρ *the large difference in the realization* $(0, 8)$ *tilts the evaluation in favor of the column marginal. Moreover, the property of Diminishing Sensitivity implies that* $\chi \in \Pi$*.*

The following result proves that when Transitivity is added to the previous axioms,

the decision criterion reduces to expected utility maximization, confirming a similar conclusion obtained by Bikhchandani and Segal (2011) in a slightly different setting.

Proposition 8. *If* Π *admits a correlation-sensitive representation then the following are equivalent:*

1. Π *satisfies Transitivity;*

2. $\succsim^{\Pi}$ *admits an expected utility representation.*

The intuition behind the additional strengthening imposed by Transitivity on the correlation-sensitive representation is the following. Since the preference set is complete, the representing ϕ must be skew symmetric, and imposing Archimedean Continuity and Strong Independence ensures that probabilities are correctly taken into account. However, only when Transitivity is added the alternatives are valued independently.

A.3.1 Monotonicity and Continuity

Since salience theory is defined for lotteries with monetary outcomes, from now on, we will focus on the case where $X = \mathbb{R}$ endowed with the usual topology. In this setting, we discuss using preference sets to axiomatically describe standard regularity conditions for the representing function, such as monotonicity and continuity.

Axiom 15 (Monotonicity)**.** *For all* $x, y, z \in X$ *and* $\pi \in \Delta(X \times X)$*, if* $x > y$ *and* $\alpha \in (0, 1)$*, then*

$$\alpha\delta_{(y,z)} + (1-\alpha)\pi \in \Pi \Rightarrow \alpha\delta_{(x,z)} + (1-\alpha)\pi \in \hat{\Pi}.$$

Since we do not, in general, impose Transitivity, our monotonicity axiom slightly departs from the usual one: it requires that whenever x is strictly larger than y, x is more favorably compared than y to every alternative z. Given a correlation-sensitive representation, Monotonicity is easily characterized in terms of ϕ.

Remark 1. If Π admits a correlation-sensitive representation, Π satisfies Monotonicity if and only if ϕ is strictly increasing in the first argument and strictly decreasing in the second argument.

Before proceeding with salience theory, a few observations about the connection between first-order stochastic dominance (FOSD) and Monotonicity in the general correlation sensitive representation are in order. It is worth noting that Monotonicity is not enough to guarantee that the preference set Π satisfies first-order stochastic dominance, where the latter is defined as the requirement that

$$\pi_1 \geq_{FOSD} \pi_2 \Rightarrow \pi \in \Pi \tag{A.5}$$

with $\pi \in \hat{\Pi}$ if $\pi_1 \neq \pi_2$. However, the decision criterion axiomatized in Theorem 4 has a few stochastic monotonicity implications. Indeed, the preference set Π satisfies (A.5) when π is an *independent* joint distribution, i.e., $\pi = \pi_1 \times \pi_2$.[12]

Finally, this setup also allows for a simple characterization of the continuity properties of ϕ.

Axiom 16 (Continuity in Outcomes)**.** *Let* $(x_n)_{n\in\mathbb{N}} \to x$*. Then, for every* $\alpha \in [0,1]$*,* $y \in X$*,* $\pi \in \Delta(X \times X)$

$$\alpha\delta_{(x_n,y)} + (1-\alpha)\pi \in \Pi \quad \forall n \in \mathbb{N} \implies \alpha\delta_{(x,y)} + (1-\alpha)\pi \in \Pi$$

and

$$\alpha\delta_{(y,x_n)} + (1-\alpha)\pi \in \Pi \quad \forall n \in \mathbb{N} \implies \alpha\delta_{(y,x)} + (1-\alpha)\pi \in \Pi.$$

Given Completeness, Strong Independence, and Archimedean Continuity, Continuity in Outcomes is one to one with a continuous ϕ.

Remark 2. If Π admits a correlation-sensitive representation, Π satisfies Continuity in Outcomes if and only if ϕ is continuous in both arguments.

[12]Indeed, Remark 1 guarantees that a preference set that satisfies Completeness, Strong Independence, Archimedean Continuity, and Monotonicity, admits a representing ϕ that satisfies the OPT and I properties of Loomes and Sugden (1987).

A.4 Salience Characterization

This section describes salience theory as introduced by BGS and shows why it is a particular case of our correlation sensitive representation in which $\phi(x,y) = (x-y)\,\sigma(x,y)$ and σ is a function that captures the salience of the joint realization (x,y), and satisfies some psychologically motivated conditions. We then propose an equivalent but testable formulation of salience theory's critical properties of Ordering, Diminishing Sensitivity, and Weak Reflexivity. Finally, we characterize the salience model entirely as the result of these Ordering, Diminishing Sensitivity and Weak Reflexivity axioms combined with continuity and monotonicity requirements.

As formulated in BGS, salience theory explains the behavior of a DM that is facing a joint lottery $\pi \in \Delta(X \times X)$. Salience's main departure from EU theory is that expectations are calculated with a distorted probability measure that overweights salient pairs of outcomes. To formalize this idea, BGS introduced the concept of *salience function.*

Definition 16. A function $\sigma : \mathbb{R}^2 \to \mathbb{R}$ satisfies:

1. *Symmetry* if $\sigma(x,y) = \sigma(y,x)$;

2. *BGS-Ordering* if $x' < y'$, $x < y$ and $[x',y'] \subset [x,y]$ imply $\sigma(x',y') < \sigma(x,y)$;

3. *BGS-Diminishing Sensitivity* if $x, y, k \in \mathbb{R}_{++}$ and $x > y$ imply $\sigma(x+k, y+k) < \sigma(x,y)$;

4. *BGS-Weak Reflexivity* if for all $x, y, x', y' \in \mathbb{R}_+$ with $|x-y| = |x'-y'|$,

$$\sigma(x,y) \geq \sigma(x',y') \Longleftrightarrow \sigma(-x,-y) \geq \sigma(-x',-y').$$

A *salience function* is a function $\sigma : \mathbb{R}^2 \to \mathbb{R}_+$ satisfying Symmetry, BGS-Ordering, BGS-Diminishing Sensitivity, BGS-Weak Reflexivity and such that $\sigma(x,x) = 0$ for all $x \in \mathbb{R}$.

We will interpret the properties momentarily when we introduce their testable counterparts. A fundamental feature is that a joint realization's salience depends only on its value, not its probability, a key difference with prospect theory. Indeed, relative to the original vN-M set of axioms, prospect theory relaxes even the weaker version of Strong Independence for joint distributions introduced by this chapter, while salience theory relaxes Transitivity.

Definition 17. A preference set Π admits a σ-distorted representation if there exists a continuous function $\sigma : \mathbb{R}^2 \to \mathbb{R}_+$ that satisfies symmetry such that

$$\pi \in \Pi \Leftrightarrow \sum_{(x,y)\in X\times X} (x-y)\,\sigma(x,y)\,\pi(x,y) \geq 0. \tag{A.6}$$

It admits a (smooth) salience representation if σ is also a salience function.

It is easy to see that a σ-distorted representation is a particular case of our correlation-sensitive model. The latter is much more general, and allows for behaviors that are at odds with salience theory's key idea that states where the alternatives differ more are overweighted. Therefore, we next characterize BGS-Ordering, BGS-Diminishing Sensitivity, and BGS-Weak Reflexivity in terms of testable axioms.

Notice that BGS mainly used the rank-based version of their model, but they recognized that its discontinuity causes some problems, and they suggest using the smooth version of Definition 17.[13] In what follows, we stick with the smooth version, which has been the most used in empirical studies of the salience model.[14] Supplementary Appendix A.9 analyzes the weaknesses of the rank-based version.

A.4.1 The Ordering Axiom

The idea behind the BGS-Ordering property is straightforward. Fix the outcomes $x > y$. Then, we can take some $\alpha, \beta \in (0,1)$, $\beta > \alpha$ and consider the two outcomes

[13] In their words: *"A smooth specification would also address a concern with the current model that states with similar salience may obtain very different weights. This implies that (1) splitting states and slightly altering payoffs could have a large impact on choice, and (2) in choice problems with many states the (slightly) less salient states are effectively ignored."*

[14] See Dertwinkel-Kalt and Koster (2020), Dertwinkel-Kalt, Frey, and Koster (2021), Nielsen, Sebald, and Sorensen (2021) and the references therein.

obtained by mixing x and y

$$x > \beta x + (1-\beta)\, y > \alpha x + (1-\alpha)\, y > y.$$

If we consider the two realizations (x, y) and $(\alpha x + (1-\alpha)\, y, \beta x + (1-\beta)\, y)$ the first pair of outcomes has more widespread values, and therefore BGS-Ordering implies that its contribution in favor of the row outcome will be relatively overweighted. However, distortions of probabilities are not observable, and therefore, we cannot directly test BGS-Ordering. Nevertheless, we can propose a testable version of the property.

Now, if we look at the joint distribution

$$\left((x, y)\,, \frac{\beta-\alpha}{1+\beta-\alpha}; (\alpha x + (1-\alpha)\, y, \beta x + (1-\beta)\, y)\,, \frac{1}{1+\beta-\alpha}\right)$$

the row and column marginals have the same expected value, and they should be indifferent to an expected value maximizer. However, a salience-sensitive DM's attention is disproportionately drawn to the outcome with the most significant difference between payoff (in the inclusion sense). Since this outcome is (x, y), and favors the row component, a salience-sensitive DM prefers (at least weakly) to be paid according to the row component. This reasoning is formalized in the Ordering axiom.

Axiom 17 (Ordering)**.** *For every $x, y \in \mathbb{R}$, $\alpha, \beta \in [0,1]$ if $x > y$, $\beta > \alpha$, and at least one between β and α is in $(0,1)$, we have that*

$$\left((x, y)\,, \frac{\beta-\alpha}{1+\beta-\alpha}; (\alpha x + (1-\alpha)\, y, \beta x + (1-\beta)\, y)\,, \frac{1}{1+\beta-\alpha}\right) \in \hat{\Pi}.$$

The following proposition shows that the axiom corresponds to the original property of BGS.

Proposition 9. *Let Π admit a σ-distorted representation. Then Π satisfies Ordering if and only if σ satisfies BGS-Ordering. In that case, the representing ϕ satisfies*

$$\phi\,(x, y) > \frac{\phi\,(\beta x + (1-\beta)\, y, \alpha x + (1-\alpha)\, y)}{(\beta - \alpha)} \tag{A.7}$$

for all $x, y \in \mathbb{R}$ *and* $\alpha, \beta \in [0,1]$ *with* $x > y$, $\beta > \alpha$ *and at least one between* β *and* α *in* $(0,1)$.

Equation (A.7) confirms the intuition behind Ordering: under this axiom, the positive contribution of the realization $\phi(x,y)$ decreases sufficiently fast as the two components are mixed, because of the combined effect of a smaller difference and a decreased salience.

A.4.2 The Diminishing Sensitivity Axiom

The BGS-Diminishing Sensitivity property requires that when two pairs of outcomes have the same absolute difference, the one with the highest relative difference is overweighted. The interpretation is easier for two-outcome lotteries. Suppose that the DM is envisioning the joint probability distribution π that assigns probability $\frac{1}{2}$ both to (x,y) and $(y+k, x+k)$, with $x, y, k \in \mathbb{R}_+$ and $x > y$. The two pairs of outcomes have the same absolute difference, but (x,y) has a higher relative difference. Therefore, (x,y) is overweighted to $(y+k, x+k)$. Since (x,y) favors the row marginal, the DM chooses to be paid according to the row outcome. This reasoning is formalized in the Diminishing Sensitivity axiom.

Axiom 18 (Diminishing Sensitivity)**.** *For every* $x > y > 0$*, and* $k \in \mathbb{R}_+$

$$\pi = \left((x,y), \frac{1}{2}; (y+k, x+k), \frac{1}{2} \right) \in \Pi.$$

If moreover $\pi \in \hat{\Pi}$ *whenever* $k \in \mathbb{R}_{++}$, $\hat{\Pi}$ *satisfies strict Diminishing Sensitivity.*

The following proposition shows that our testable definition of Diminishing Sensitivity corresponds to the original property of BGS.

Proposition 10. *If* Π *admits a* σ*-distorted representation, it satisfies strict Diminishing Sensitivity if and only if* σ *satisfies BGS-Diminishing Sensitivity.*

In particular, it turns out that Diminishing Sensitivity alone is *not* in contrast with the conventional notion of prospect theory. It is a generalization of the property

of risk aversion over positive outcomes and risk loving over negative outcomes (cf. also Proposition 13) to decision criteria that are not necessarily transitive. Denote as $\mathbb{E}(p) = \sum_{x \in X} p(x)\, x$ the expected value of the marginal distribution $p \in \Delta(X)$.

Definition 18. Π satisfies risk aversion (risk loving) for outcomes in (a, b) if $\pi \in \Pi$ (resp. $\bar{\pi} \in \Pi$) for all $\pi \in \Delta(X)$ with $\operatorname{supp} \pi \subseteq (a, b)$ and such that π_2 is a mean preserving spread of π_1.

The previous definition is a translation of the usual risk aversion notion in the language of preference sets: a DM is risk averse over the outcome range (a, b) if she prefers the expected value of a lottery supported over (a, b) to the lottery itself.

Proposition 11. *Let Π admit an expected utility representation with a strictly increasing utility function. Then Π satisfies Diminishing Sensitivity if and only if Π satisfies risk aversion for positive outcomes.*

This result confirms that the BGS-Diminishing Sensitivity of the function σ allows for risk-aversion of the agents in the main specification of the BGS model (Equation (A.6)) without relying on the more general form[15]

$$\sum_{(x,y) \in X \times X} (u(x) - u(y))\, \sigma(x, y)\, \pi(x, y).$$

Remark 3. Under the correlation-sensitive representation, risk aversion for positive outcomes always implies Diminishing Sensitivity. However, the following example shows that risk aversion for positive outcomes is a strictly more demanding property. Let the salience function be equal to the leading example in BGS, that is

$$\sigma(x, y) = \frac{x - y}{x + y + 1}. \tag{A.8}$$

Then σ satisfies BGS-Ordering and BGS-Diminishing Sensitivity, and by Proposition 11, Π satisfies Diminishing Sensitivity. The joint distribution π given in the follow-

[15] Moreover, a representation where σ satisfies Diminishing Sensitivity and u is concave and differentiable can always be reformulated as $\sum_{(x,y) \in X \times X} (x - y)\, \hat{\sigma}(x, y)\, \pi(x, y) \geq 0$, where $\hat{\sigma}(x, y) = \begin{cases} \frac{\sigma(x,y)[u(x)-u(y)]}{x-y} & x \neq y \\ 0 & x = y \end{cases}$ is a continuous function that satisfies BGS-Diminishing Sensitivity.

ing table is such that the row marginal is a mean preserving spread of the column marginal:

π	0	1	2
0	0	1/4	0
1	0	1/2	0
2	1/8	0	1/8

Therefore risk aversion in the positive domain would prescribe that $\pi \notin \Pi$. However, $\pi \in \Pi$ for a DM with salience function given by (A.8) because of the high salience of the realization $(2, 0)$. Therefore, the preference set of such a DM satisfies Diminishing Sensitivity but not risk aversion for positive outcomes.

A.4.3 The Weak Reflexivity Axiom

The last property introduced by BGS is Weak Reflexivity, which captures the symmetry around 0 of the distortions. Again, we provide a testable counterpart of their axiom.

Axiom 19 (Weak Reflexivity)**.** *For every $x, y, w, z \in \mathbb{R}_+$, with $x - y = z - w$*

$$\left((x, y), \frac{1}{2}; (w, z), \frac{1}{2}\right) \in \hat{\Pi} \Leftrightarrow \left((-y, -x), \frac{1}{2}; (-z, -w), \frac{1}{2}\right) \in \hat{\Pi}.$$

The axiom is easily seen to be one to one with the corresponding property of the distortion function σ.

Proposition 12. *If Π admits a σ-distorted representation, Π satisfies Weak Reflexivity if and only if σ satisfies BGS-Weak Reflexivity.*

So far, we have not attached any specific interpretation to the lotteries' realizations, except that they are expressed in monetary units. In particular, they can represent either the total wealth or gains and losses obtained after realizing some uncertainty. However, the Weak Reflexivity axiom, with the implied role for outcome 0, better suits the latter interpretation. We notice that Weak Reflexivity implies the preference reversal of risk attitudes featured by prospect theory.

Proposition 13. *Suppose that* Π *has an EU representation and satisfies Monotonicity and Weak Reflexivity. Then* Π *is risk-averse (resp. risk-loving) for lotteries with values in* $(a, b) \subseteq \mathbb{R}_+$ *if and only if* $\succsim$ *is risk loving (resp. risk-averse) for lotteries with values in* $(-b, -a)$*.*

This result sheds light on the observation made in BGS that salience theory can explain the experimental evidence in favor of the fourfold pattern (see, e.g., Bruhin, Fehr-Duda, and Epper 2010). Diminishing Sensitivity would only induce risk aversion in the gain domain. Its combination with Ordering creates the risk aversion for small gains vs. risk loving for large gains, and Weak Reflexivity gives the opposite patterns for losses.

A.4.4 Complete Characterization of Salience Theory

We now put the pieces together and provide a complete characterization of the salience model. To do so, we need a final continuity axiom.

Axiom 20 (Continuity at Identity)**.** *Let* $x \in X$*. Then, for every* $(x_n)_{n\in\mathbb{N}}$ *such that* $x_n \downarrow x$*, and for every* $k \in X$ *and* $\varepsilon \in \mathbb{R}_{++}$ *there exists an* $m \in \mathbb{N}$ *such that for all* $n \geq m$

$$((x, x_n), (1 - (x_n - x)); (k + \varepsilon, k), (x_n - x)) \in \Pi.$$

Moreover, for every $(x_n)_{n\in\mathbb{N}}$ *such that* $x_n \uparrow x$*, and for every* $k \in X$ *and* $\varepsilon \in \mathbb{R}_{++}$ *there exists an* $m \in \mathbb{N}$ *such that for all* $n \geq m$

$$((x_n, x), (1 - (x - x_n)); (k + \varepsilon, k), (x - x_n)) \in \Pi.$$

The axiom requires that joint realizations with two components that are arbitrarily closed can be almost neglected. More precisely, the weight to these realizations declines more than linearly in their differences when these become sufficiently small, capturing a form of indistinguishability. With this, we have a complete characterization of the salience model.[16]

[16] As shown by the proof of Theorem 5, adding Monotonicity, Continuity in Outcomes and Continuity

Theorem 5. *A preference set* Π *admits a salience representation if and only if* Π *satisfies Completeness, Strong Independence, Archimedean Continuity, Monotonicity, Continuity in Outcomes, Continuity at Identity, Ordering, Diminishing Sensitivity, and Weak Reflexivity.*

A.4.5 Comparison with Other Models

Relation with Regret Theory Theorem 5 allows us to compare salience theory with regret theory readily. Indeed, recall that the most general version of regret theory, proposed by Loomes and Sugden (1987), requires that the preference set Π of the DM admits a correlation-sensitive representation, it satisfies Monotonicity, and the representing ϕ satisfies Regret Aversion:

$$\phi(x,y) > \phi(x,z) + \phi(z,y) \text{ for all } x > z > y.$$

Corollary 3. *If a preference set* Π *satisfies Completeness, Strong Independence, Archimedean Continuity, and Ordering, the representing* ϕ *satisfies Regret Aversion.*

The two models remain inherently different despite Ordering being a stronger property than Regret Aversion in binary decision problems. First, they have different psychological foundations that imply different behaviors when the DM is given additional information. The behavior of a salience-sensitive DM is the same when only the realization of the chosen marginal is shown and when the counterfactual is announced. Instead, regret theory prescribes an EU consistent behavior in the first scenario but is highly sensitive to correlation in the second.

Second, by making additional assumptions such as Ordering and Diminishing Sensitivity, salience theory delivers a novel set of predictions. This is particularly evident for problems with more than two alternatives, where the salience model predicts the decoy effect, background contrast effects, and other context effects, a phenomenon

at Identity to a correlation sensitive representation implies that $\phi(x,y) = (x-y)\,\sigma(x,y)$ for some continuous and symmetric $\sigma : \mathbb{R}^2 \to \mathbb{R}_+$ such that $\sigma(x,x) = 0$ for all $x \in \mathbb{R}$. Then, by adding Ordering, Diminishing Sensitivity, and Weak Reflexivity, σ is forced to be a salience function.

that we illustrate in the extension of Supplementary Appendix A.8. As the empirical literature has highlighted the widespread presence of these effects, salience theory seems a better correlation-sensitive model in terms of the performance across different decision environments.[17]

Relation with Reference-Dependent Preferences Another model under for which the correlation between alternatives play a key role is the one of Koszegi and Rabin (2007, 2009). Here, we show how to translate the model in the language of preference set, highlight that it does not fall in the general class of correlation-sensitive preferences we have proposed, and shows that this comes from violations of the Strong Independence axiom.

Koszegi and Rabin (2007, 2009) model endogenous reference-dependent preferences as a (personal) choice-unacclimating equilibrium. More precisely, under their criterion, a lottery p can be chosen if, when alternatives are evaluated taking p as the reference point, p has the highest evaluation for the DM. The reference dependence is endogenous because the reference point is the candidate choice. It is called choice-unacclimating, as when looking from deviations from a candidate, the agent still evaluates them with the candidate as the reference point (as opposed to using the deviation itself as the reference point). When generalized to allow for correlated alternatives (similarly to Sugden 2003) and rephrased in the language of preference sets, their decision criterion says that:

$$\pi \in \Pi \Leftrightarrow \sum_{(x,y):x \geq y} \lambda (x-y) \pi (x,y) + \sum_{(x,y):x<y} (x-y) \pi (x,y) \geq 0$$

for some $\lambda > 1$ that measures how much worse losses are than gains.[18] The interpretation is that the agent prefers the row marginal (i.e., $\pi \in \Pi$) when taking the row

[17] Of course, salience theory also makes some predictions about non-choice behavior that separate it from preexisting models, such as the attention dedicated to each dimension of the alternatives. Thus, the use of additional instruments such as eye-tracking to further investigate is critical. Moreover, Herweg and Muller (2021) argue that the more restrictive form of regret theory presented in Loomes and Sugden (1982) is itself a special case of salience theory.

[18] We focus on their main specification and thank a referee for suggesting this link.

marginal as the status quo. Here, the fact that the row marginal is the status quo is captured by the fact that the gains induced by the column marginal are evaluated with weight 1, while the losses are evaluated with weight λ.

Interestingly, by letting $\hat{\phi}(x,y) = \lambda(x-y)$, for $x \geq y$ and $\hat{\phi}(x,y) = x-y$ for $x < y$, this criterion admits a representation

$$\pi \in \Pi \Leftrightarrow \sum_{(x,y)\in X\times X} \hat{\phi}(x,y)\,\pi(x,y) \geq 0.$$

However, the loss aversion coefficient makes this $\hat{\phi}$ not skew symmetric. Moreover, by Theorem 4 the reference-dependent model does not admit an alternative correlation-sensitive representation. Indeed, a simple example shows that this criterion violates Strong Independence.

Example 7. *Let* $\pi = \frac{\delta_{(10,1)}}{2} + \frac{\delta_{(1,10)}}{2}$ *and* $\pi' = \delta_{(1.1,1)}$. *Then both the row and column marginals can be chosen under* π, *and only the row marginal can be chosen under* π', *(i.e.,* $\pi \in \Pi\backslash\hat{\Pi}, \pi' \in \hat{\Pi}$*), and Strong Independence would prescribe that* $\alpha\pi + (1-\alpha)\pi' \in \hat{\Pi}$ *for all* $\alpha \in (0,1)$. *However, for* λ *and* α *sufficiently high* $\alpha\pi + (1-\alpha)\pi' \in \Pi\backslash\hat{\Pi}$. *The intuition is simple: a highly loss-averse agent that takes the column marginal as the reference point can stick to it because of the high loss associated with the realization* $(1,10)$.

Therefore, beyond establishing the formal distinction between this model and the class of correlation-sensitive preferences that contain salience and regret, the preference set approach hints at violations of the "strict" part of the Strong Independence axiom as the essential relaxation to allow for status quo biases. Loosely speaking, the reference-dependent model have "too much Completeness" due to loss aversion. Recall that the row marginal is the reference point, so there will be several instances in which both marginals can be chosen if they were the original reference point. This thickness of the indifference curves can lead to violations of Strong Independence, as even if one of the original joint distributions is in the strict preference set, the resulting convex combination may fall in the thick indifference curve part, i.e., it may

be in the preference set but not in the strict preference set. Instead, it is easy to see that Completeness, Archimedean Continuity, and the "weak" part of the Strong Independence are still satisfied.

A.4.6 Identification of the Salience Function

Another advantage of our use of preference sets is that in light of Theorem 4, we can directly test salience theory by first constructing a candidate salience function σ and then checking whether it satisfies the properties imposed by BGS. As a preliminary observation, it is immediate from (A.6) that if the preferences set Π admits a salience representation with salience function σ, they also admit a salience representation with salience function $k\sigma$ whenever $k \in \mathbb{R}_{++}$. Therefore, to eliminate this degree of freedom, we set $\sigma(1,0) = 1$.

Now, for every $(x, y) \in \mathbb{R}^2$ with $y > x$, if the preference set Π admits a smooth salience representation, by Theorem 5 Π satisfies Completeness, Archimedean Continuity, Strong Independence and Monotonicity, and therefore by Theorem 4 there exists a unique $\alpha_{x,y} \in (0, 1)$ such that

$$\alpha_{x,y}\delta_{(x,y)} + (1 - \alpha_{x,y})\,\delta_{(1,0)} \in \Pi \backslash \hat{\Pi}.$$

Therefore, we can define

$$\sigma(x, y) = \frac{(1 - \alpha_{x,y})}{\alpha_{x,y}(y - x)}.$$

It is immediate to check that this is the only possible value for σ. We can use this procedure and the fact that by symmetry $\sigma(x, y) = \sigma(y, x)$ for those $(x, y) \in \mathbb{R}^2$ with $x > y$ to construct the candidate salience function. At this point, checking salience theory boils down to verifying that σ satisfies BGS-Ordering, BGS-Diminishing Sensitivity, and BGS-Weak Reflexivity.

A.5 Conclusion

This work provides a simple axiomatic characterization of preferences over risky choices when the agent cares about the correlation between the alternatives considered. We proved that when the joint distribution is included in the decision environment, we can pinpoint Transitivity as the EU's relaxation needed for correlation sensitivity. This setting, moreover, allows a cleaner axiomatic comparison of theories such as regret and salience with EU.

As the second payoff of our approach, we obtain a simple axiomatization of the salience model of Bordalo, Gennaioli, and Shleifer (2012) within the realm of these correlation-sensitive preferences. This provides a one-to-one map from the BGS assumptions of Ordering, Diminishing Sensitivity, and Weak Reflexivity to testable counterparts. Our characterization reveals that Ordering is the property that cannot be reconciled with prospect theory, whereas Diminishing Sensitivity paired with Weak Reflexivity corresponds to the usual risk-averse in gains, risk-loving in losses. Moreover, the axiomatization allows for direct comparisons of the different EU axioms relaxed by salience theory, prospect theory, regret theory, and the reference-dependent preferences of Koszegi and Rabin (2007).

Bibliography

[1] Bell, David (1982), "Regret in decision making under uncertainty," *Operations research*, 30, 961-981.

[2] Bikhchandani, Sushil, and Uzi Segal, (2011): "Transitive regret," *Theoretical Economics,* 6, 95-108.

[3] Bleichrodt, Han, Alessandra Cillo, and Enrico Diecidue, (2010): "A quantitative measurement of regret theory," *Management Science*, 56, 161-175.

[4] Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer (2012): "Salience theory of choice under risk," *Quarterly Journal of Economics,* 127, 1243-1295.

[5] Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer (2013): "Salience and Consumer Choice," *Journal of Political Economy,* 121, 803-843.

[6] Braun, Michael, and Muermann, Alexander (2004): "The impact of regret on the demand for insurance", *Journal of Risk and Insurance,* vol. 71, pp. 737–67.

[7] Bruhin, Adrian, Helga Fehr-Duda, and Thomas Epper, (2010): "Risk and rationality: Uncovering heterogeneity in probability distortion," *Econometrica*, 78, 1375-1412.

[8] Cerreia-Vioglio, Simone, Alfio Giarlotta, Salvatore Greco, Fabio Maccheroni, and Massimo Marinacci (2020). "Rational Preference and Rationalizable Choice," *Economic Theory*, 69, 61–105.

[9] Cerreia-Vioglio, Simone, and Efe Ok (2020): "The rational core of preference relations," Università Bocconi, mimeo.

[10] Chew, Soo Hong: "A generalization of the quasilinear mean with applications to the measurement of income inequality and decision theory resolving the Allais paradox," *Econometrica*, 1065-1092.

[11] Chew, Soo Hong, Larry, Epstein, and Uzi Segal, (1991): "Mixture symmetry and quadratic utility," *Econometrica*, 139-163.

[12] Dertwinkel-Kalt, Markus, Jonas Fray, and Mats Koster (2021): "Optimal Stopping in a Dynamic Salience Model," *mimeo.*

[13] Dertwinkel-Kalt, Markus, and Mats Koster (2020): "Salience and Skewness Preferences," *Journal of the European Economic Association,* 18, 2057-2107.

[14] Diecidue Enrico, and Jeeva Somasundaram, (2017): "Regret Theory: a new foundation," *Journal of Economic Theory,* 172, 88-119.

[15] Ellis, Andrew, and Yusufcan Masatioglu (2021): "Choice with Endogenous Categorization," *Review of Economic Studies*, forthcoming.

[16] Filiz-Ozbay, Emel, and Erkut Ozbay, (2007). "Auctions with anticipated regret: theory and experiment", *American Economic Review*, vol. 97, pp. 1407–18

[17] Fishburn, Peter (1982): "Nontransitive measurable utility," *Journal of Mathematical Psychology*, 26, 31-67.

[18] Fishburn, Peter (1983): "Transitive measurable utility," *Journal of Economic Theory, 31,* 293-317.

[19] Fishburn, Peter (1989): "Non-transitive measurable utility for decison under uncertainty," *Journal of Mathematical Economics,* 18, 187-207.

[20] Fishburn, Peter (1990a): "Continuous nontransitive additive conjoint measurement," *Mathematical Social Sciences* 20: 165-193.

[21] Fishburn, Peter (1990b): "Skew symmetric additive utility with finite states," *Mathematical Social Sciences* 19: 103-115.

[22] Frydman Cary, and Milica Mormann, (2017): "The Role of Salience in Choice under Risk: An Experimental Investigation," mimeo.

[23] Herweg, Fabian, and Daniel Muller, (2021) "A comparison of regret theory and salience theory for decisions under risk," *Journal of Economic Theory*, 193, 1-19.

[24] Konigsheim Christian, Moritz Lukas, and Markus Noth, (2019): "Salience theory: Calibration and Heterogeneity in Probability Distortion," *Journal of Economic Behavior and Organization,* 157, 477-495.

[25] Kontek, Krzysztof, (2016) "A critical note on Salience theory of choice under risk," *Economics Letters,* 149: 168-171.

[26] Koster, Mats (2021): "A Multivariate Salience Theory of Choice under Risk," *mimeo.*

[27] Koszegi, Botond, and Matthew Rabin. (2007) "Reference-Dependent Risk Attitudes," *American Economic Review*, 97, 1047-1063.

[28] Koszegi, Botond, and Matthew Rabin. (2009) "Reference-Dependent Consumption Plans," *American Economic Review*, 99, 909- 936.

[29] Koszegi, Botond, and Adam Szeidl, (2013) "A Model of Focusing in Economic Choice," *Quarterly Journal of Economics,* 53-107.

[30] Krantz, David, Duncan Luce, Patrick Suppes, and Amos Tversky, (1971) "Foundations of measurement, Vol. I: Additive and polynomial representations".

[31] Kreweras, Germaine, (1961). "Sur une possibilite de rationaliser les intransitivites," *La Decision*, 27-32.

[32] Loomes, Graham, and Robert Sugden (1987) "Some Implications of a more General Form of Regret Theory," *Journal of Economic Theory*, 41, 270-287.

[33] Loomes, Graham, and Robert Sugden (1982) "Regret Theory: An Alternative Theory of Rational Choice Under Uncertainty," *The Economic Journal*, 92, 805-824.

[34] Masatlioglu, Yusufcan, and Collin Raymond (2015) "A behavioral analysis of stochastic reference dependence," *American Economic Review*, 106, 2760-82.

[35] Nielsen Carsten, Alexander Sebald, and Peter Sorensen (2021), "Testing for Salience effects in choice under risk," mimeo.

[36] Nishimura, Hiroki, and Efe Ok, (2018) Preference Structures, mimeo, NYU.

[37] Quiggin, John (1994) "Regret theory with general choice sets," *Journal of Risk and Uncertainty*, 8, 153-165.

[38] Smith, Richard (1996): "Is regret theory an alternative basis for estimating the value of health care interventions?", *Health Policy,* vol. 37, pp. 105–15.

[39] Sugden, Robert (1993). "An Axiomatic Foundation for Regret theory," *Journal of Economic Theory*, 60, 159-180.

[40] Sugden, Robert (2003). "Reference-dependent subjective expected utility," *Journal of Economic Theory*, 111, 172-191.

A.6 Main Proofs

Let $\oplus = \left\{(x,y) \in X \times X : \delta_{(x,y)} \in \Pi\right\}$ and $\hat{\oplus} = \left\{(x,y) \in X \times X : \delta_{(x,y)} \in \hat{\Pi}\right\}$.

Proof of Theorem 4

(Necessity of the axioms) Completeness is necessary since the skew symmetry of ϕ guarantees that $\sum_{(x,y)\in X\times X}\pi(x,y)\,\phi(x,y) = -\sum_{(x,y)\in X\times X}\bar{\pi}(x,y)\,\phi(x,y)$. For Strong Independence, let $\pi,\chi\in\Pi$ (resp. $\pi\in\Pi$ and $\chi\in\hat{\Pi}$) and $\lambda\in(0,1)$. Then

$$\begin{aligned}&\sum_{(x,y)\in X\times X}(\lambda\pi+(1-\lambda)\chi)(x,y)\,\phi(x,y)\\ =\;&\lambda\sum_{(x,y)\in X\times X}\pi(x,y)\,\phi(x,y)+(1-\lambda)\sum_{(x,y)\in X\times X}\chi(x,y)\,\phi(x,y)\geq(\text{resp. }\;>)0.\end{aligned}$$

For Archimedean Continuity, let $\pi\in\hat{\Pi},\chi\notin\Pi$. If we define $K:=\sum_{(x,y)\in X\times X}\pi(x,y)\,\phi(x,y) > 0 > \sum_{(x,y)\in X\times X}\chi(x,y)\,\phi(x,y) =: k$, then any $\alpha > \frac{-k}{K-k}$ and $\beta < \frac{-k}{K-k}$ are easily seen to satisfy the requirements.

(Sufficiency of the axioms) We start by establishing some initial claims.

Claim 7. *If* $\operatorname{supp}\pi\subseteq\oplus$*, then* $\pi\in\Pi$.

Proof The claim is proved by induction on the size of $\operatorname{supp}\pi$. The claim is clearly true when $|\operatorname{supp}\pi| = 1$. Suppose the result holds for all the lotteries with support of size $n\in\mathbb{N}$. Let π be such that $|\operatorname{supp}\pi| = n+1$. Choose arbitrarily $(x',y')\in\operatorname{supp}\pi$. Then, we can define $\chi\in\Delta(X\times X)$ as

$$\chi(x,y) = \begin{cases} 0 & \text{if } (x,y) = (x',y')\\ \frac{\pi(x,y)}{1-\pi(x',y')} & \text{otherwise.}\end{cases}$$

Since $|\operatorname{supp}\chi| = n$ and $\operatorname{supp}\chi\subseteq\oplus$, we have $\chi\in\Pi$. Moreover, $\pi = \pi(x',y')\,\delta_{(x',y')} + (1-\pi(x',y'))\,\chi$ and by Strong Independence, we have $\pi\in\Pi$. □

Claim 8. *Let* $\pi\in\hat{\Pi}$, $\chi\notin\Pi$*, there exists a unique* $\lambda\in(0,1)$ *such that* $\lambda\pi+(1-\lambda)\chi\in\Pi\backslash\hat{\Pi}$.

Proof We let $A = \left\{\lambda\in[0,1] : \lambda\pi+(1-\lambda)\chi\in\hat{\Pi}\right\}$, $B = \{\lambda\in[0,1] : \lambda\pi+(1-\lambda)\chi\notin\Pi\}$. By Archimedean Continuity, both A and B have a nonempty intersection with $(0,1)$. Suppose that $\lambda\in A$ and $\mu\in(\lambda,1]$. Then $\mu\pi+(1-\mu)\chi = \frac{\mu-\lambda}{1-\lambda}\pi+\frac{1-\mu}{1-\lambda}(\lambda\pi+(1-\lambda)\chi)$

and Strong Independence implies that $\mu\pi + (1-\mu)\,\chi \in \hat{\Pi}$. This, in turn, implies that $\mu \in A$.

Suppose instead that $\lambda \in B$ and $\mu \in [0, \lambda)$. Then by Completeness $\lambda\bar{\pi} + (1-\lambda)\,\bar{\chi} = \overline{\lambda\pi + (1-\lambda)\,\chi} \in \hat{\Pi}$ and $\mu\bar{\pi} + (1-\mu)\,\bar{\chi} = \frac{\lambda-\mu}{\lambda}\bar{\chi} + \frac{\mu}{\lambda}\,(\lambda\bar{\pi} + (1-\lambda)\,\bar{\chi})$. Therefore, $\mu\bar{\pi} + (1-\mu)\,\bar{\chi} \in \hat{\Pi}$ by Strong Independence, and $\mu\pi + (1-\mu)\,\chi \notin \Pi$. This, in turn, implies that $\mu \in B$.

Summing up, A and B are two intervals in $[0,1]$ with empty intersection. Suppose by contradiction that $A \cup B = [0,1]$. Then, we either have $A = [\lambda^*, 1]$ and $B = [0, \lambda^*)$ or $A = (\lambda^*, 1]$ and $B = [0, \lambda^*]$. In the first case, $\lambda^*\pi + (1-\lambda^*)\,\chi \in \hat{\Pi}$, $\chi \notin \Pi$, and Archimedean Continuity imply the existence of a $\mu \in [0, \lambda^*)$ such that $\mu\pi + (1-\mu)\,\chi \in \hat{\Pi}$, a contradiction. Similarly, we can rule out the other case. Therefore, there exists $\lambda^* \in [0,1] \setminus (A \cup B)$, that is, $\lambda^*\pi + (1-\lambda^*)\,\chi \in \Pi \backslash \hat{\Pi}$.

It only remains to prove uniqueness. Suppose that both λ^* and μ^* have the desired property, and let $\mu^* > \lambda^*$. Then, $\mu^*\pi + (1-\mu^*)\,\chi = \frac{\mu^*-\lambda^*}{1-\lambda^*}\pi + \frac{1-\mu^*}{1-\lambda^*}\,(\lambda^*\pi + (1-\lambda^*)\,\chi)$ and by Strong Independence, $\mu^*\pi + (1-\mu^*)\,\chi \in \hat{\Pi}$, a contradiction. □

Claim 9. *Let $x, y, z, w, t, v \in X$, $\lambda, \mu, \alpha \in (0,1)$ and $\delta_{(x,y)}, \delta_{(z,w)}, \delta_{(t,v)} \in \hat{\Pi}$ with*

$$\begin{aligned} \lambda\delta_{(x,y)} + (1-\lambda)\,\delta_{(w,z)} &\in \Pi \backslash \hat{\Pi}, \\ \mu\delta_{(z,w)} + (1-\mu)\,\delta_{(v,t)} &\in \Pi \backslash \hat{\Pi}, \\ \alpha\delta_{(t,v)} + (1-\alpha)\,\delta_{(y,x)} &\in \Pi \backslash \hat{\Pi}. \end{aligned}$$

Then

$$\frac{\lambda}{1-\lambda} \cdot \frac{\mu}{1-\mu} \cdot \frac{\alpha}{1-\alpha} = 1.$$

Proof Let

$$\gamma = \frac{\mu}{\mu + 1 - \lambda}$$

and

$$\pi = \gamma\left(\lambda\delta_{(x,y)} + (1-\lambda)\,\delta_{(w,z)}\right) + (1-\gamma)\left(\mu\delta_{(z,w)} + (1-\mu)\,\delta_{(v,t)}\right).$$

By Strong Independence, $\pi \in \Pi \backslash \hat{\Pi}$. Since $\gamma\,(1-\lambda) = (1-\gamma)\,\mu$, by Completeness we

have that $\frac{\delta_{(w,z)}+\delta_{(z,w)}}{2} = \frac{\gamma(1-\lambda)\delta_{(w,z)}+(1-\gamma)\mu\delta_{(z,w)}}{\gamma(1-\lambda)+(1-\gamma)\mu} \in \Pi\backslash\hat{\Pi}$. Suppose by way of contradiction that

$$\frac{\gamma\lambda\delta_{(x,y)} + (1-\gamma)(1-\mu)\delta_{(v,t)}}{\gamma\lambda + (1-\gamma)(1-\mu)} \notin \Pi.$$

Then Completeness implies that

$$\frac{\gamma\lambda\delta_{(y,x)} + (1-\gamma)(1-\mu)\delta_{(t,v)}}{\gamma\lambda + (1-\gamma)(1-\mu)} \in \hat{\Pi}$$

and by Strong Independence, $\bar{\pi} \in \hat{\Pi}$. But this leads to the contradiction $\pi \notin \Pi$. Similarly, suppose by contradiction that

$$\frac{\gamma\lambda\delta_{(x,y)} + (1-\gamma)(1-\mu)\delta_{(v,t)}}{\gamma\lambda + (1-\gamma)(1-\mu)} \in \hat{\Pi}.$$

Then, Strong Independence implies that $\pi \in \hat{\Pi}$, another contradiction. Therefore, we have that

$$\frac{\gamma\lambda\delta_{(x,y)} + (1-\gamma)(1-\mu)\delta_{(v,t)}}{\gamma\lambda + (1-\gamma)(1-\mu)} \in \Pi\backslash\hat{\Pi}$$

and by definition of $\hat{\Pi}$

$$\frac{\gamma\lambda\delta_{(y,x)} + (1-\gamma)(1-\mu)\delta_{(t,v)}}{\gamma\lambda + (1-\gamma)(1-\mu)} \in \Pi\backslash\hat{\Pi}.$$

Thus Claim 8 gives $1-\alpha = \frac{\gamma\lambda}{\gamma\lambda+(1-\gamma)(1-\mu)}$ that implies

$$\alpha\mu\lambda = (1-\lambda)(1-\mu)(1-\alpha)$$

proving the statement. □

Claim 10. *If* $\operatorname{supp}\pi \subseteq \oplus$, *and* $\operatorname{supp}\pi \cap \hat{\oplus} \neq \emptyset$ *then* $\pi \in \hat{\Pi}$.

Proof If $\pi = \delta_{(x,y)}$ for some (x,y), the result holds by definition of $\hat{\oplus}$. Therefore, suppose that π is supported at least on two joint outcome realizations, let $(x',y') \in$

$\operatorname{supp}\pi \cap \hat{\oplus}$, and define

$$\chi(x,y) = \begin{cases} 0 & (x,y) = (x',y') \\ \frac{\pi(x,y)}{1-\pi(x',y')} & \text{otherwise.} \end{cases}$$

By Claim 7, $\chi \in \Pi$. Since

$$\pi = \pi(x',y')\,\delta_{(x',y')} + (1-\pi(x',y'))\,\chi$$

Strong Independence implies that $\pi \in \hat{\Pi}$. □

Claim 11. *If $\eta, \chi \in \Pi\backslash\hat{\Pi}$, then for all $\rho \in \Delta(X \times X)$*

$$\lambda\rho + (1-\lambda)\,\chi \in \Pi \Longleftrightarrow \lambda\rho + (1-\lambda)\,\eta \in \Pi.$$

Proof By Strong Independence both statements hold if $\rho \in \Pi$. If $\rho \notin \Pi$, then by Completeness $\bar{\rho} \in \hat{\Pi}$, and by assumption $\bar{\eta}, \bar{\chi} \in \Pi$. Therefore, by Strong Independence, both $\lambda\bar{\rho} + (1-\lambda)\,\bar{\chi}$ and $\lambda\bar{\rho} + (1-\lambda)\,\bar{\eta}$ are in $\hat{\Pi}$. But then, neither $\lambda\rho + (1-\lambda)\,\chi \in \Pi$ nor $\lambda\rho + (1-\lambda)\,\eta \in \Pi$. □

If for every $x, y \in X$, $\delta_{(x,y)} \in \Pi\backslash\hat{\Pi}$, by Claim 7 every $\pi \in \Delta(X \times X)$ is in $\Pi\backslash\hat{\Pi}$, and the statement of the theorem trivially holds by letting $\phi(x,y) = 0$ for all $x, y \in X$. Therefore, by Completeness, we can assume that there exists $(\hat{x}, \hat{y})$ with $\delta_{(\hat{x},\hat{y})} \in \hat{\Pi}$ and let $\phi(\hat{x},\hat{y})$ be an arbitrary strictly positive real number. Moreover, let $\phi(x,y) = 0$ for all $\delta_{(x,y)} \in \Pi\backslash\hat{\Pi}$. If $(x,y) \notin \oplus$, by Claim 8, there exists a unique $\lambda \in (0,1)$ with

$$\lambda\delta_{(\hat{x},\hat{y})} + (1-\lambda)\,\delta_{(x,y)} \in \Pi\backslash\hat{\Pi}.$$

In this case, let

$$\phi(x,y) = -\phi(\hat{x},\hat{y})\,\frac{\lambda}{(1-\lambda)}.$$

It only remains to define ϕ when $(x,y) \in \hat{\oplus}$. We set

$$\phi(x,y) = -\phi(y,x) \quad \forall\,(x,y) \in \hat{\oplus}.$$

We now claim that the previous procedure defines ϕ uniquely up to a positive linear transformation. Given the choice of a particular $(\hat{x}, \hat{y})$, the only degree of freedom is the choice of the (strictly positive) number $\phi(\hat{x}, \hat{y})$, and the values assumed by ϕ on the rest of the domain are linear in $\phi(\hat{x}, \hat{y})$. Suppose instead that we define $\bar{\phi}$ starting from a different $(\bar{x}, \bar{y}) \in \hat{\oplus}$. Since we prove uniqueness only up to a positive linear transformation, we can choose the (strictly positive) value of $\bar{\phi}(\bar{x}, \bar{y})$. In particular, set

$$\bar{\phi}(\bar{x}, \bar{y}) = \phi(\bar{x}, \bar{y}) = \phi(\hat{x}, \hat{y}) \frac{\mu}{(1-\mu)}$$

where

$$\mu\delta_{(\hat{x},\hat{y})} + (1-\mu)\delta_{(\bar{y},\bar{x})} \in \Pi \backslash \hat{\Pi}$$

and consider $(x, y) \notin \oplus$. Then, by Claim 8 there exist unique λ_0, λ_1, such that

$$\begin{aligned} \lambda_0\delta_{(\hat{x},\hat{y})} + (1-\lambda_0)\delta_{(x,y)} &\in \Pi \backslash \hat{\Pi}, \\ \lambda_1\delta_{(\bar{x},\bar{y})} + (1-\lambda_1)\delta_{(x,y)} &\in \Pi \backslash \hat{\Pi}. \end{aligned}$$

Given our definitions,

$$\begin{aligned} \phi(x, y) = \bar{\phi}(x, y) &\iff \phi(\hat{x}, \hat{y}) \tfrac{\lambda_0}{(1-\lambda_0)} = \phi(\bar{x}, \bar{y}) \tfrac{\lambda_1}{(1-\lambda_1)} \\ &\iff \phi(\hat{x}, \hat{y}) \tfrac{\lambda_0}{(1-\lambda_0)} = \phi(\hat{x}, \hat{y}) \tfrac{\mu}{(1-\mu)} \tfrac{\lambda_1}{(1-\lambda_1)} \\ &\iff \tfrac{\lambda_0}{(1-\lambda_0)} = \tfrac{\mu}{(1-\mu)} \tfrac{\lambda_1}{(1-\lambda_1)} \end{aligned}$$

and Claim 9 together with Completeness guarantee that the condition in the last line holds true. Finally, we want to show that

$$\pi \in \Pi \iff \sum_{(x,y) \in \operatorname{supp} \pi} \pi(x, y)\, \phi(x, y) \geq 0.$$

We will consider three possible cases.

(*First Case*) Suppose $\operatorname{supp} \pi \subseteq \oplus$, then by Claim 7, $\pi \in \Pi$, and by definition of ϕ, $\phi(x, y) \geq 0$ for every $(x, y) \in \operatorname{supp} \pi$.

(*Second Case*) Suppose $\operatorname{supp} \bar{\pi} \subseteq \oplus$, and $\operatorname{supp} \bar{\pi} \cap \hat{\oplus} \neq \emptyset$. Then by Claim 10 $\bar{\pi} \in \hat{\Pi}$

and $\pi \notin \Pi$. By definition of ϕ, $\phi(x,y) \leq 0$ for every $(x,y) \in \operatorname{supp}\pi$, and $\phi(x,y) < 0$ for some $(x,y) \in \operatorname{supp}\pi$.

(*Third Case*) Finally, we show that all the other possibilities can be reduced into one of the first two cases. Fix $t \in X$. Suppose we are not in one of the first two cases, that is, there exist $(x_0,y_0),(x_1,y_1) \in \operatorname{supp}\pi$ with $(x_0,y_0),(y_1,x_1) \in \hat{\oplus}$. Then by Claim 8 there exists a unique $\alpha \in (0,1)$ such that $\alpha\delta_{(x_0,y_0)} + (1-\alpha)\delta_{(x_1,y_1)} \in \Pi \backslash \hat{\Pi}$. By Claim 9 and uniqueness up to a positive linear transformation, $\frac{\alpha}{(1-\alpha)}\phi(x_0,y_0) = \phi(y_1,x_1)$. If $\frac{\alpha}{1-\alpha} = \frac{\pi(x_0,y_0)}{\pi(x_1,y_1)}$, then Claim 11 guarantees that $\pi \in \Pi$ if and only if $\pi' \in \Pi$ where[19]

$$\pi'(x,y) = \begin{cases} \pi(x,y) & (x,y) \notin \{(x_0,y_0),(x_1,y_1),(t,t)\} \\ 0 & (x,y) \in \{(x_0,y_0),(x_1,y_1)\} \\ \pi(t,t) + \pi(x_0,y_0) + \pi(x_1,y_1) & (x,y) = (t,t). \end{cases}$$

Moreover,

$$\pi(x_0,y_0)\phi(x_0,y_0) + \pi(x_1,y_1)\phi(x_1,y_1) = 0 = \phi(t,t)(\pi(t,t) + \pi(x_0,y_0) + \pi(x_1,y_1))$$

so that $\sum_{(x,y)\in\operatorname{supp}\pi} \pi(x,y)\phi(x,y) \geq 0 \Leftrightarrow \sum_{(x,y)\in\operatorname{supp}\pi'} \pi'(x,y)\phi(x,y) \geq 0$.

[19] To see this, apply Claim 11 with $\eta = \delta_{(t,t)}$, $\chi = \alpha\delta_{(x_0,y_0)} + (1-\alpha)\delta_{(x_1,y_1)}$, $\lambda = 1 - \pi(x_0,y_0) - \pi(x_1,y_1)$, and $\rho(x,y) = \frac{\pi(x,y)}{1-\pi(x_0,y_0)-\pi(x_1,y_1)}$ if $(x,y) \notin \{(x_0,y_0),(x_1,y_1)\}$ and $\rho(x,y) = 0$ otherwise.

If $\frac{\alpha}{1-\alpha} > \frac{\pi(x_0,y_0)}{\pi(x_1,y_1)}$, Claim 11 guarantees that $\pi \in \Pi$ if and only $\pi' \in \Pi$ where[20]

$$\pi'(x,y) = \begin{cases} \pi(x,y) & (x,y) \notin \{(x_0,y_0),(x_1,y_1),(t,t)\} \\ 0 & (x,y) = (x_0,y_0) \\ \pi(x_1,y_1) - \frac{1-\alpha}{\alpha}\pi(x_0,y_0) & (x,y) = (x_1,y_1) \\ \pi(t,t) + \pi(x_0,y_0) + \frac{1-\alpha}{\alpha}\pi(x_0,y_0) & (x,y) = (t,t). \end{cases}$$

Moreover,

$$\begin{aligned} & \pi(x_0,y_0)\phi(x_0,y_0) + \pi(x_1,y_1)\phi(x_1,y_1) + \phi(t,t)\pi(t,t) \\ = & -\pi(x_0,y_0)\frac{1-\alpha}{\alpha}\phi(x_1,y_1) + \pi(x_1,y_1)\phi(x_1,y_1) + 0 \\ = & \left(\pi(x_1,y_1) - \frac{1-\alpha}{\alpha}\pi(x_0,y_0)\right)\phi(x_1,y_1) + 0 \\ = & \pi'(x_1,y_1)\phi(x_1,y_1) + \phi(t,t)\pi'(t,t) \end{aligned}$$

so that

$$\sum_{(x,y)\in \operatorname{supp}\pi} \pi(x,y)\phi(x,y) \geq 0 \Leftrightarrow \sum_{(x,y)\in \operatorname{supp}\pi'} \pi'(x,y)\phi(x,y) \geq 0.$$

A similar equivalence can be obtained if $\frac{\alpha}{1-\alpha} < \frac{\pi(x_0,y_0)}{\pi(x_1,y_1)}$. In every instance, the resulting π' has strictly fewer elements in the support that do not belong to $\Pi\backslash\hat{\Pi}$ than the original π. Since the support is finite, by repeating this procedure a finite number of times, we will obtain a $\hat{\pi} \in \Delta(X \times X)$ that falls in one of the first two cases, and such that $\pi \in \Pi \Leftrightarrow \hat{\pi} \in \Pi$ and

$$\sum_{(x,y)\in \operatorname{supp}\pi} \pi(x,y)\phi(x,y) \geq 0 \Leftrightarrow \sum_{(x,y)\in \operatorname{supp}\hat{\pi}} \hat{\pi}(x,y)\phi(x,y) \geq 0$$

[20]To see this, apply Claim 11 with $\eta = \delta_{(t,t)}$, $\chi = \alpha\delta_{(x_0,y_0)} + (1-\alpha)\delta_{(x_1,y_1)}$, $\lambda = 1 - \pi(x_0,y_0) - \frac{1-\alpha}{\alpha}\pi(x_0,y_0)$, and

$$\rho(x,y) = \begin{cases} \frac{\pi(x,y)}{1-\pi(x_0,y_0)-\frac{1-\alpha}{\alpha}\pi(x_0,y_0)} & (x,y) \notin \{(x_0,y_0),(x_1,y_1)\} \\ 0 & (x,y) = (x_0,y_0) \\ \frac{\pi(x_1,y_1)-\frac{1-\alpha}{\alpha}\pi(x_0,y_0)}{1-\pi(x_0,y_0)-\frac{1-\alpha}{\alpha}\pi(x_0,y_0)} & (x,y) = (x_1,y_1). \end{cases}$$

concluding the proof. ■

Proof of Proposition 8 We establish Proposition 8 by proving the following more general result.[21]

Claim 12. *If* Π *admits a correlation-sensitive representation then the following are equivalent:*

1. Π *satisfies Transitivity;*

2. $\succsim^{\Pi}$ *satisfies Classic Completeness;*

3. $\succsim^{\Pi}$ *satisfies Classic Completeness, Classic Transitivity, Classic Archimedean Continuity, and Classic Strong Independence;*

4. $\succsim^{\Pi}$ *admits an expected utility representation.*

Proof We define the binary relation $\geq$ over outcomes as $x \geq y \Leftrightarrow \delta_{(x,y)} \in \Pi$. We will be interested in whether ϕ is modular with respect to this binary relation, i.e.,[22]

$$\forall x, x', y, y' \in X \quad \phi\left((x,y) \vee (x',y')\right) + \phi\left((x,y) \wedge (x',y')\right) = \phi(x,y) + \phi(x',y'). \quad \text{(A.9)}$$

The claim is proved by showing that each of the different conditions in the statement is equivalent to Equation (A.9). Notice that since positive linear transformations preserve modularity, it does not matter which representing ϕ we consider.

Equation (A.9) $\Rightarrow$ 4. Let $x_0 \in X$. Define $u(z)$ as $\phi(z, x_0)$. Fix a pair (z, w), with $z \geq w$. There are three cases:

[21] We use the adjective Classic for the conventional versions of Completeness, Transitivity, Archimedean Continuity, and Strong Independence for binary relations. The definition for these standard notions are in Supplementary Appendix A.7.

[22] Note the slight abuse of terminology here, as $\geq$ defined as above is not in general antisymmetric (although it is in our application to salience theory with monetary outcomes) and therefore the join and meet of two elements of the set may be not well defined. In that case everything works even with indifferencies with the understanding that $(x,y) \vee (x',y')$ is any pair (z,w) where $z \in \{x, x'\}$ $z \geq x$, $z \geq x'$ and $w \in \{y, y'\}$ $w \geq y$, $w \geq y'$ and $(x,y) \wedge (x',y')$ is any pair (z,w) where $z \in \{x, x'\}$ $z \leq x$, $z \leq x'$ and $w \in \{y, y'\}$ $w \leq y$, $w \leq y'$.

- $z \geq w \geq x_0$. Applying (A.9) with $x = z$, $y = x' = x_0$ and $y' = w$ we have:

$$\begin{aligned} \phi(z,w) + \phi(x_0,x_0) &= \phi(z,x_0) + \phi(x_0,w) \Leftrightarrow \\ \phi(z,w) &= \phi(z,x_0) - \phi(w,x_0) \Leftrightarrow \phi(z,w) = u(z) - u(w) \end{aligned}$$

  where the first implication follows from the skew symmetry of ϕ.

- $z \geq x_0 \geq w$. Applying (A.9) with $x = z$, $y = w$ and $x_0 = y' = x'$ we have:

$$\phi(z,x_0) + \phi(x_0,w) = \phi(z,w) + \phi(x_0,x_0) \Leftrightarrow \phi(z,w) = u(z) - u(w)$$

  where the implication follows from the skew symmetry of ϕ and the definition of u.

- $x_0 \geq z \geq w$. Applying (A.9) with $x = z$, $y = x' = x_0$ and $y' = w$ we have:

$$\phi(x_0,x_0) + \phi(z,w) = \phi(z,x_0) + \phi(x_0,w) \Leftrightarrow \phi(z,w) = u(z) - u(w)$$

  where the implication follows from the skew symmetry of ϕ and the definition of u.

This proves that $\phi(z,w) = u(z) - u(w)$ whenever $z \geq w$. If $w > z$, by skew-symmetry of ϕ, $\phi(z,w) = -\phi(w,z) = -(u(w) - u(z)) = u(z) - u(w)$ proving that the equality $\phi(z,w) = u(z) - u(w)$ holds for every $z, w \in X$. Therefore, we have $\pi \in \Pi$ if and only if

$$\begin{aligned} \sum_{(x,y)\in X\times X} \pi(x,y)\,\phi(x,y) \geq 0 &\iff \sum_{(x,y)\in X\times X} \pi(x,y)\,(u(x) - u(y)) \geq 0 \\ &\iff \sum_{x\in X} \pi_1(x)\,u(x) \geq \sum_{x\in X} \pi_2(x)\,u(x) \end{aligned}$$

proving that Π admits an EU representation.

4⇒ *Equation (A.9)* If Π admits an EU representation then

$$\pi \in \Pi \Longleftrightarrow \sum_{(x,y)\in X\times X} \pi(x,y)\left(u(x)-u(y)\right) \geq 0.$$

Therefore, if we define $\phi(z,w) = (u(z)-u(w))$, modularity holds: let $x,y,x',y' \in X$

$$\begin{aligned} & \phi((x,y)\vee(x',y'))+\phi((x,y)\wedge(x',y')) \\ = \;& u(x\vee x')-u(y\vee y')+u(x\wedge x')-u(y\wedge y') \\ = \;& u(x)+u(x')-u(y)-u(y') = \phi(x,y)+\phi(x',y'). \end{aligned}$$

$3 \Leftrightarrow 4$ is a version of the vN-M EU theorem.

$4 \Rightarrow 1$ is straightforward given the representation.

$4 \Rightarrow 2$ holds trivially.

$2 \Rightarrow$*Equation (A.9)* and $1 \Rightarrow$*Equation (A.9)* are proved by contradiction. Suppose that there exist $x,y,x',z' \in X$ such that

$$\phi((x,y)\vee(x',y'))+\phi((x,y)\wedge(x',y')) > \phi(x,y)+\phi(x',y')$$

with $(x\vee x') = x$ and $(y\vee y') = y'$. Then the inequality reads

$$\phi(x,y')+\phi(x',y) > \phi(x,y)+\phi(x',y'). \tag{A.10}$$

Choose $(z,w) \in (X\times X)$ and $\alpha \in [0,1]$ such that

$$\alpha\phi(z,w)+(1-\alpha)\left(\frac{\phi(x,y')+\phi(x',y)}{2}\right) > 0 > \alpha\phi(z,w)+(1-\alpha)\left(\frac{\phi(x,y)+\phi(x',y')}{2}\right).$$

The existence of such (z,w) and α is guaranteed by (A.10). Then

$$\alpha\delta_{(z,w)}+\frac{(1-\alpha)\,\delta_{(x,y')}}{2}+\frac{(1-\alpha)\,\delta_{(x',y)}}{2} \in \hat{\Pi} \text{ and } \alpha\delta_{(z,w)}+\frac{(1-\alpha)\,\delta_{(x,y)}}{2}+\frac{(1-\alpha)\,\delta_{(x',y')}}{2} \notin \Pi. \tag{A.11}$$

We now show that (A.11) implies that neither Classic Completeness of $\succsim^{\Pi}$ nor Tran-

sitivity of Π holds. For Classic Completeness notice that (A.11) implies that neither

$$\alpha\delta_z + \frac{(1-\alpha)\,\delta_x}{2} + \frac{(1-\alpha)\,\delta_{x'}}{2} \succsim^{\Pi} \alpha\delta_w + \frac{(1-\alpha)\,\delta_{y'}}{2} + \frac{(1-\alpha)\,\delta_y}{2}$$

nor

$$\alpha\delta_w + \frac{(1-\alpha)\,\delta_{y'}}{2} + \frac{(1-\alpha)\,\delta_y}{2} \succsim^{\Pi} \alpha\delta_z + \frac{(1-\alpha)\,\delta_x}{2} + \frac{(1-\alpha)\,\delta_{x'}}{2}$$

holds, and $\succsim^{\Pi}$ does not satisfy Classic Completeness.

As for Transitivity, let

$$\begin{aligned}
\pi &= \alpha\delta_{(z,z)} + \frac{(1-\alpha)\,\delta_{(x,x)}}{2} + \frac{(1-\alpha)\,\delta_{(x',x')}}{2}, \\
\chi &= \alpha\delta_{(z,w)} + \frac{(1-\alpha)\,\delta_{(x,y')}}{2} + \frac{(1-\alpha)\,\delta_{(x',y)}}{2}, \\
\rho &= \alpha\delta_{(z,w)} + \frac{(1-\alpha)\,\delta_{(x,y)}}{2} + \frac{(1-\alpha)\,\delta_{(x',y')}}{2}.
\end{aligned}$$

Completeness of Π implies that $\pi \in \Pi$, and (A.11) gives $\chi \in \Pi$, $\rho \notin \Pi$. However, since $\pi_1 = \rho_1$, $\pi_2 = \chi_1$, and $\chi_2 = \rho_2$, Transitivity of Π does not hold. Similar arguments can be used to obtain contradictions for other violations of modularity. ■

Proof of Remark 1 (If) Let $x, y, z \in X$ with $x > y$. By assumption, we have $\phi(x,z) > \phi(y,z)$. Therefore, for every $\alpha \in (0,1)$, $z \in X$, and $\pi \in \Delta(X)$

$$\begin{aligned}
\alpha\delta_{(y,z)} + (1-\alpha)\,\pi \in \Pi \iff & \alpha\phi(y,z) + (1-\alpha) \sum_{(x',y')\in X\times X} \pi(x',y')\,\phi(x',y') \geq 0 \\
\Rightarrow \quad & \alpha\phi(x,z) + (1-\alpha) \sum_{(x',y')\in X\times X} \pi(x',y')\,\phi(x',y') > 0 \\
\iff & \alpha\delta_{(x,z)} + (1-\alpha)\,\pi \in \hat{\Pi}.
\end{aligned}$$

(Only if) We first prove that ϕ is strictly increasing in the first argument. Let $x_1, x_2, y \in X$, $x_1 > x_2$. Under a correlation-sensitive representation, we have $\frac{\delta_{(x_2,y)}}{2} + \frac{\delta_{(y,x_2)}}{2} \in \Pi$. Then by Monotonicity $\frac{\delta_{(x_1,y)}}{2} + \frac{\delta_{(y,x_2)}}{2} \in \hat{\Pi}$ and given the correlation-sensitive representation this implies $\phi(x_1,y) > \phi(x_2,y)$. To see that ϕ is strictly

decreasing in the second argument, notice that by skew symmetry:

$$\phi(x_1, y) > \phi(x_2, y) \Rightarrow -\phi(y, x_1) > -\phi(y, x_2) \Rightarrow \phi(y, x_1) < \phi(y, x_2)$$

concluding the proof. ■

Proof of Remark 2 (Only if) If ϕ is always equal to 0 the claim is obvious. Therefore, suppose there exists $z, w \in X$ with $\phi(z, w) > 0$. We prove continuity in the first argument; continuity in the second argument follows from skew-symmetry. Let $(x_n)_{n\in\mathbb{N}} \to x$, and suppose that there exists $y \in X$ such that $\phi(x_n, y) \nrightarrow \phi(x, y)$. There are two cases:

i) There exists an infinite subsequence of $(x_{n_k})_{k\in\mathbb{N}}$ and an $\varepsilon > 0$ such that $\phi(x_{n_k}, y) \geq \phi(x, y) + \varepsilon$ for all $k \in \mathbb{N}$. If $\phi(x, y) \geq -\varepsilon$ notice that we have

$$\begin{aligned}
& \forall k \in \mathbb{N} \quad \frac{\phi(z, w)}{\phi(x, y) + \varepsilon + \phi(z, w)}\phi(x_{n_k}, y) + \frac{\phi(x, y) + \varepsilon}{\phi(x, y) + \varepsilon + \phi(z, w)}\phi(w, z) \geq 0 \\
\Leftrightarrow \quad & \forall k \in \mathbb{N} \quad \frac{\phi(z, w)}{\phi(x, y) + \varepsilon + \phi(z, w)}\delta_{(x_{n_k}, y)} + \frac{\phi(x, y) + \varepsilon}{\phi(x, y) + \varepsilon + \phi(z, w)}\delta_{(w,z)} \in \Pi \\
\Longrightarrow \quad & \frac{\phi(z, w)}{\phi(x, y) + \varepsilon + \phi(z, w)}\delta_{(x,y)} + \frac{\phi(x, y) + \varepsilon}{\phi(x, y) + \varepsilon + \phi(z, w)}\delta_{(w,z)} \in \Pi \\
\Leftrightarrow \quad & \phi(x, y) \geq \phi(x, y) + \varepsilon
\end{aligned}$$

a contradiction. If $\phi(x, y) < -\varepsilon$ notice that we have

$$\begin{aligned}
& \forall k \in \mathbb{N} \quad \frac{\phi(w, z)}{\phi(x, y) + \varepsilon + \phi(w, z)}\phi(x_{n_k}, y) + \frac{\phi(x, y) + \varepsilon}{\phi(x, y) + \varepsilon + \phi(w, z)}\phi(z, w) \geq 0 \\
\Leftrightarrow \quad & \forall k \in \mathbb{N} \quad \frac{\phi(w, z)}{\phi(x, y) + \varepsilon + \phi(w, z)}\delta_{(x_{n_k}, y)} + \frac{\phi(x, y) + \varepsilon}{\phi(x, y) + \varepsilon + \phi(w, z)}\delta_{(z,w)} \in \Pi \\
\Longrightarrow \quad & \frac{\phi(w, z)}{\phi(x, y) + \varepsilon + \phi(w, z)}\delta_{(x,y)} + \frac{\phi(x, y) + \varepsilon}{\phi(x, y) + \varepsilon + \phi(w, z)}\delta_{(z,w)} \in \Pi \\
\Leftrightarrow \quad & \phi(x, y) \geq \phi(x, y) + \varepsilon
\end{aligned}$$

a contradiction.

ii) There exists an infinite subsequence of $(x_{n_k})_{k\in\mathbb{N}}$ and an $\varepsilon > 0$ such that

$\phi(x_{n_k}, y) \leq \phi(x, y) - \varepsilon$ for all $k \in \mathbb{N}$. If $\phi(x, y) \geq \varepsilon$ notice that we have

$$\begin{aligned}
& \forall k \in \mathbb{N} \quad \frac{\phi(z, w)}{\phi(x, y) - \varepsilon + \phi(z, w)} \phi(x_{n_k}, y) + \frac{\phi(x, y) - \varepsilon}{\phi(x, y) - \varepsilon + \phi(z, w)} \phi(w, z) \leq 0 \\
\Leftrightarrow \quad & \forall k \in \mathbb{N} \quad \frac{\phi(z, w)}{\phi(x, y) - \varepsilon + \phi(z, w)} \delta_{(y, x_{n_k})} + \frac{\phi(x, y) - \varepsilon}{\phi(x, y) - \varepsilon + \phi(z, w)} \delta_{(z, w)} \in \Pi \\
\Longrightarrow \quad & \frac{\phi(z, w)}{\phi(x, y) - \varepsilon + \phi(z, w)} \delta_{(y, x)} + \frac{\phi(x, y) - \varepsilon}{\phi(x, y) + \varepsilon + \phi(z, w)} \delta_{(z, w)} \in \Pi \\
\Leftrightarrow \quad & \phi(x, y) - \varepsilon \geq \phi(x, y).
\end{aligned}$$

a contradiction. The case $\phi(x, y) \leq \varepsilon$ is proved along the same lines.

(If) Trivial. ■

A.6.1 Salience Characterization

Proof of Proposition 9 Let Π admit a correlation-sensitive representation, $x, y \in \mathbb{R}$, and $\alpha, \beta \in [0, 1]$ with $x > y$ and $\beta > \alpha$ with at least one between α and β in $(0, 1)$. We have $\left((x, y), \frac{\beta - \alpha}{1 + \beta - \alpha}; (\alpha x + (1 - \alpha) y, \beta x + (1 - \beta) y), \frac{1}{1 + \beta - \alpha}\right) \in \hat{\Pi}$ if and only if $\frac{\beta - \alpha}{1 + \beta - \alpha} \phi(x, y) + \frac{1}{1 + \beta - \alpha} \phi(\alpha x + (1 - \alpha) y, \beta x + (1 - \beta) y) > 0$ that by skew symmetry of ϕ is equivalent to

$$\phi(x, y)(\beta - \alpha) > \phi(\beta x + (1 - \beta) y, \alpha x + (1 - \alpha) y). \tag{A.12}$$

Now, let Π admit a σ-distorted representation. We show that Ordering of Π implies BGS-Ordering of σ, the other direction is trivial.

We first show that if $x \geq z > w \geq y$ with $[y, x] \supset [w, z]$, then $\sigma(x, y) > \sigma(w, z)$. Define $\alpha = \frac{w - y}{x - y}$ and $\beta := \frac{z - y}{x - y}$ and notice that $0 \leq \alpha < \beta \leq 1$ with at least one of the two inequalities being strict. Therefore, (A.12) implies that

$$\begin{aligned}
(\beta - \alpha)(x - y) \sigma(x, y) &> (\beta - \alpha)(x - y) \sigma(\alpha x + (1 - \alpha) y, \beta x + (1 - \beta) y) \\
&= (\beta - \alpha)(x - y) \sigma(w, z),
\end{aligned}$$

and $\sigma(x, y) > \sigma(w, z)$.

Next, let $z = w$, with $x \geq w \geq y$ and at least one of the two inequalities strict, say $x > w \geq y$. Suppose by way of contradiction that $\sigma(w, w) \geq \sigma(x, y)$. By continuity of σ, there exists an $\varepsilon < \frac{x-w}{2}$ with $\sigma(w + \varepsilon, w) > \sigma(x, y)$. But this is a contradiction with what was proved in the previous paragraph. ■

Proof of Proposition 10 Let Π admit a σ-distorted representation and satisfy strict Diminishing Sensitivity. Fix $x > y > 0$, $k > 0$, we have that $\left((x, y), \frac{1}{2}; (y + k, x + k), \frac{1}{2}\right) \in \hat{\Pi}$. Given the σ-distorted representation, this is equivalent to $(x - y)\sigma(x, y) + (y - x)\sigma(y + k, x + k) > 0$. The previous inequality holds if and only if $\sigma(x, y - k) > \sigma(y, x + k) = \sigma(x + k, y)$ proving that σ satisfies BGS-Diminishing Sensitivity. All the steps are reversible. ■

Proof of Proposition 11 (If) Let $x \geq y \geq 0$ and $k \geq 0$. Consider the two marginal distributions $p = \left(x, \frac{1}{2}; y + k, \frac{1}{2}\right)$ and $q = \left(x + k, \frac{1}{2}; y, \frac{1}{2}\right)$. Notice that q is a mean-preserving spread of p, since q can be obtained by further randomizing each realization z of p with the additional random term h_z defined as $h_x = \left(k, \frac{(x-y)}{(x-y)+k}; (y - x), \frac{k}{(x-y)+k}\right)$ and $h_{y+k} = \left((x - y), \frac{k}{(x-y)+k}; -k, \frac{(x-y)}{(x-y)+k}\right)$. Therefore, as risk-averse expected utility DMs dislike mean-preserving spreads:

$$\sum_{z \in X} p(z)\, u(z) \geq \sum_{z \in X} q(z)\, u(z).$$

Rearranging the terms $\frac{1}{2}(u(x) - u(y)) + \frac{1}{2}(u(y + k) - u(x + k)) \geq 0$ or

$$\left((x, y), \frac{1}{2}; (y + k, x + k), \frac{1}{2}\right) \in \Pi$$

and Diminishing Sensitivity holds.

(Only If) Let $x_0 \geq y_0 \geq 0$. By Diminishing Sensitivity

$$\left(\left(\frac{x_0 + y_0}{2}, y_0\right), \frac{1}{2}; \left(\frac{x_0 + y_0}{2}, x_0\right), \frac{1}{2}\right) \in \Pi$$

that is $u\left(\frac{x_0+y_0}{2}\right) \geq \frac{u(x_0)+u(y_0)}{2}$ proving the midpoint concavity of u on the set of positive real numbers. Since u is strictly increasing, it is measurable. Since the Sierpinski theorem implies that a midpoint concave and measurable function is concave, the

DM is risk-averse on that range. ■

Proof of Proposition 12 Let $x, y, w, z \in \mathbb{R}_+$, with $x - y = z - w > 0$. Under a σ-distorted representation

$$\left((x, y), \frac{1}{2}; (w, z), \frac{1}{2}\right) \in \hat{\Pi} \Leftrightarrow \left((-y, -x), \frac{1}{2}; (-z, -w), \frac{1}{2}\right) \in \hat{\Pi}$$

is tantamount to

$$(x - y)\,\sigma(x, y) > (z - w)\,\sigma(w, z) \Leftrightarrow (x - y)\,\sigma(-x, -y) > (z - w)\,\sigma(-w, -z)$$

which is equivalent to $\sigma(x, y) > \sigma(w, z) \Leftrightarrow \sigma(-x, -y) > \sigma(-w, -z)$. The case in which $x-y = z-w < 0$ is completely analogous, and the one in which $x-y = z-w = 0$ immediately follows from the fact that for all $x, w \in \mathbb{R}_+$, $\left((x, x), \frac{1}{2}; (w, w), \frac{1}{2}\right) \in \Pi$ and $\left((-x, -x), \frac{1}{2}; (-w, -w), \frac{1}{2}\right) \in \Pi$. ■

Proof of Proposition 13 We will prove only the case in which Π is risk-averse in (a, b) as the other case is analogous. Let u be a vN-M utility index representing Π such that $u(0) = 0$, and suppose that Π is risk-averse for lotteries with values in $(a, b) \subseteq \mathbb{R}_+$. Let $-b < -x \leq -y < -a$, since u is concave on (a, b), we have $u(x) - u\left(\frac{x+y}{2}\right) \leq u\left(\frac{x+y}{2}\right) - u(y)$ that is $\left(\left(x, \frac{x+y}{2}\right), \frac{1}{2}; \left(y, \frac{x+y}{2}\right), \frac{1}{2}\right) \notin \hat{\Pi}$. By Weak Reflexivity, this means that $\left(\left(-\frac{x+y}{2}, -x\right), \frac{1}{2}; \left(-\frac{x+y}{2}, -y\right), \frac{1}{2}\right) \notin \hat{\Pi}$ or $u\left(-\frac{x+y}{2}\right) \leq \frac{u(-x)+u(-y)}{2}$. This shows that u is mid-point convex on $(-a, -b)$. Since it is also increasing, it is measurable, and by the Sierpinski theorem it is convex on $(-a, -b)$, proving the statement. ■

Proof of Theorem 5 (Only If) Given a smooth salience representation, let $\phi(x, y) = \sigma(x, y)(x - y)$. By the symmetry axiom for σ, we have $\phi(x, y) = \sigma(x, y)(x - y) = \sigma(y, x)(x - y) = -\sigma(y, x)(y - x) = -\phi(y, x)$ proving that ϕ is skew-symmetric. Then Π satisfies Completeness, Strong Independence, and Archimedean Continuity by Theorem 4. It satisfies Ordering, Diminishing Sensitivity, and Weak Reflexivity by Propositions 9, 10, and 12. Since σ satisfies BGS-Ordering, Π satisfies Monotonicity

by Remark 1. To see it, suppose that $y \geq x > x'$. Then $\phi(x,y) = \sigma(x,y)(x-y) \geq \sigma(x',y)(x'-y) = \phi(x,y)$ where the inequality is due to $0 \leq \sigma(x,y) \leq \sigma(x,y')$ with at least one of the two inequalities being strict that in turns is a consequence of BGS Ordering and the fact that a salience function takes positive values by definition. The case $x > x' \geq y$ is proved similarly, and $x > x'$, $y \in (x', x)$ follows immediately from $\phi(x,y) = \sigma(x,y)(x-y) > 0 \geq \sigma(x',y)(x'-y) = \phi(x,y)$. Moreover, Π satisfies Continuity in Outcomes by Remark 2 and since $\phi(y,x)$ is the product of two jointly continuous functions. Finally, let $x \in X$, $(x_n)_{n\in\mathbb{N}}$ be such that $x_n \downarrow x$, $k \in \mathbb{R}$ and $\varepsilon \in \mathbb{R}_{++}$. Then

$$\begin{aligned}
& ((x,x_n),(1-(x_n-x));(k+\varepsilon,k),(x_n-x)) \in \Pi \\
\Leftrightarrow\ & \phi(x,x_n)(1-(x_n-x)) \geq \phi(k,k+\varepsilon)(x_n-x) \\
\Leftrightarrow\ & \sigma(x,x_n)(x-x_n)(1-(x_n-x)) \geq \sigma(k+\varepsilon,k)\,\varepsilon\,(x-x_n) \\
\Leftarrow\ & \sigma(x,x_n)(1-(x_n-x)) \leq \sigma(k+\varepsilon,k)\,\varepsilon
\end{aligned}$$

where the last inequality holds for sufficiently large n by continuity of σ, proving that Π satisfies the first condition of Continuity at Identity. An analogous argument establishes the second part.

(If) Since Π satisfies Completeness, Strong Independence, and Archimedean Continuity by Theorem 4 it admits the representation

$$\pi \in \Pi \Longleftrightarrow \sum_{(x,y)\in X\times X} \phi(x,y)\,\pi(x,y) \geq 0.$$

Define σ by

$$\sigma(x,y) = \frac{\phi(x,y)}{x-y} \qquad \forall x \neq y$$

and $\sigma(x,x) = 0$ for all $x \in X$. We have that σ maps $X \times X$ into positive real numbers by Monotonicity and Remark 1. It is immediate that

$$\pi \in \Pi \Leftrightarrow \sum_{(x,y)\in X\times X} (x-y)\,\sigma(x,y)\,\pi(x,y) \geq 0.$$

Propositions 9, 10, and 12 guarantee that σ satisfies respectively BGS-Ordering, BGS-Diminishing Sensitivity, and BGS-Weak Reflexivity. We now check that σ satisfies symmetry and it is continuous. First, σ satisfies symmetry since ϕ is skew symmetric. Moreover since ϕ is continuous by Remark 2 σ is continuous at every (x, y) such that $x \neq y$. We now show that it is continuous at each $(x, x) \in \mathbb{R} \times \mathbb{R}$. We show that $x_n \downarrow x$ implies $\sigma(x_n, x) \to 0$, the proof for the case in which $x_n \uparrow x$ is completely analogous. Without loss of generality, we can assume that $x_n \neq x$ for all $n \in \mathbb{N}$. By Continuity at Identity, for all $k \in \mathbb{R}$ and $\varepsilon \in \mathbb{R}_{++}$, there exists an $m \in \mathbb{N}$ such that for all $n \geq m$,

$$\begin{aligned}
& ((x, x_n), (1 - (x_n - x)); (k + \varepsilon, k), (x_n - x)) \in \Pi \\
\Leftrightarrow \quad & \phi(x, x_n)(1 - (x_n - x)) \geq \phi(k, k + \varepsilon)(x_n - x) \\
\Leftrightarrow \quad & \sigma(x, x_n)(x - x_n)(1 - (x_n - x)) \geq \sigma(k + \varepsilon, k)\varepsilon(x - x_n) \\
\Leftrightarrow \quad & \sigma(x, x_n)(1 - (x_n - x)) \leq \sigma(k + \varepsilon, k)\varepsilon.
\end{aligned}$$

Since the ε can be chosen arbitrarily small ε and $\sigma(k + \varepsilon, k)$ is decreasing in ε by the BGS-Ordering property established above, this proves that $\sigma(x, x_n)(1 - (x_n - x))$ is converging to 0. This concludes the proof. ■

Proof of Corollary 3 Let $x > z > y$. Then, there exists $\lambda \in (0, 1)$ with $\lambda x + (1 - \lambda) y = z$. Applying Ordering and Proposition 9 with $\alpha = \lambda$ and $\beta = 1$ we get $\phi(x, y)(1 - \lambda) > \phi(x, z)$. Applying Ordering and Proposition 9 with $\beta = \lambda$ and $\alpha = 0$ we get $\phi(x, y)\lambda > \phi(z, y)$. By summing the two inequalities, we get the desired result. ■

A.7 Binary Relations and Preference Sets

Lemma 15. *For every binary relation $\succeq$, we have $\succsim^{\Pi_{\succeq}} = \succeq$.*

We collect the definitions of some standard axioms for binary relations over $\Delta(X)$.

Axiom 21 (Classic Completeness)**.** *For all $p, q \in \Delta(X)$, either $p \succsim q$ or $q \succsim p$ or*

both.

Classic Completeness requires that the DM can (weakly) rank all the marginal lotteries. Our analysis highlights why Classic Completeness may fail: the comparison of some pairs of lotteries may depend on their correlation. The following lemma shows that Completeness of the preference set weakens Classic Completeness. The example discussed in the introduction shows why it may be a strictly weaker requirement.

Lemma 16. *1. Let $\succsim$ be a binary relation. If $\succsim$ satisfies Classic Completeness, then $\Pi_{\succsim}$ satisfies Completeness.*

2. Let Π be a preference set. If $\succsim^{\Pi}$ satisfies Classic Completeness, then Π satisfies Completeness.

That is, the preference set derived from a complete binary relation satisfies Completeness. Moreover, the binary relation induced by a preference set is complete only if the preference set satisfies Completeness.

Axiom 22 (Classic Transitivity)**.** *For all $p, q, r \in \Delta(X)$, if $p \succsim q$ and $q \succsim r$, then $p \succsim r$. Moreover, if either $p \succ q$ or $q \succ r$, then $p \succ r$.*

Classic Transitivity is the other central tenet of rationality.

Axiom 23 (Classic Strong Independence)**.** *For all $p, q, r \in \Delta(X)$ and $\alpha \in (0, 1)$,*

$$p \succsim q \Leftrightarrow \alpha p + (1-\alpha) r \succsim \alpha q + (1-\alpha) r.$$

Classic Strong Independence is the axiom usually paired to Classic Completeness, Classic Transitivity, and Classic Archimedean Continuity to derive the expected utility representation. Since we often work without Classic Transitivity in this chapter, we also need to consider an alternative and stronger form of independence.

Axiom 24 (Classic Strong B-Independence)**.** *For all $p, q, r, s \in \Delta(X)$ and $\alpha \in (0, 1)$,*

$$p \succsim q, r \succsim s \Rightarrow \alpha p + (1-\alpha) r \succsim \alpha q + (1-\alpha) s.$$

Moreover, if $p \succ q$*, then* $\alpha p + (1-\alpha) r \succ \alpha q + (1-\alpha) s$*.*

Classic Strong B-Independence says that convex combinations of preferred alternatives are preferred to the convex combination of the alternatives they dominate. It implies Classic Strong Independence, and the two axioms coincide under Classic Transitivity. The following remark proved in the Supplementary Appendix clarifies that the usual approach that assumes Classic Strong Independence for a binary relation implicitly imposes our notion of Strong Independence for preference sets.

Remark 4. If a binary relation $\succsim$ satisfies Classic Strong B-Independence, then $\Pi_{\succsim}$ satisfies Strong Independence.

The next axiom is a standard and weak form of continuity imposed on preferences defined over a convex set.

Axiom 25 (Classic Archimedean Continuity)**.** *For all* $p, q, r \in \Delta(X)$ *such that* $p \succ q$ *and* $q \succ r$*, there exist* $\alpha, \beta \in (0,1)$ *such that* $\alpha p + (1-\alpha) r \succ q$ *and* $q \succ \beta p + (1-\beta) r$*.*

A slightly more demanding version of Classic Archimedean Continuity is needed when dealing with nontransitive and incomplete preferences.

Axiom 26 (Classic Archimedean B-Continuity)**.** *For all* $p, q, r, s \in \Delta(X)$ *such that* $p \succ q$ *and* $r \not\succsim s$*, there exist* $\alpha, \beta \in (0,1)$ *such that*

$$\alpha p + (1-\alpha) r \succ \alpha q + (1-\alpha) s \text{ and } \beta p + (1-\beta) r \not\succsim \beta q + (1-\beta) s.$$

Therefore, under Classic Completeness, Classic Archimedean Continuity is the particular case of Classic Archimedean B-Continuity in which $s = q$.

Axiom 27 (Classic Sequential Continuity)**.** *For each pair of sequences* $(p_n)_{n\in\mathbb{N}}$ *and* $(q_n)_{n\in\mathbb{N}}$ *in* $\Delta(X)$ *such that* $(p_n)_{n\in\mathbb{N}} \to p_0$ *and* $(q_n)_{n\in\mathbb{N}} \to q_0$

$$p_n \succsim q_n \text{ for all } n \in \mathbb{N} \implies p_0 \succsim q_0.$$

Classic Sequential Continuity implies Classic Archimedean B-Continuity under Classic Completeness.

Lemma 17. *If $\succsim$ satisfies Classic Sequential Continuity and Classic Completeness, then $\succsim$ satisfies Classic Archimedean B-Continuity.*

The following remark shows that the usual approach that assumes Classic Archimedean B-Continuity for a binary relation implicitly imposes our notion of Archimedean Continuity for preference sets.

Remark 5. If $\succsim$ satisfies Classic Archimedean B-Continuity, then $\Pi_{\succsim}$ satisfies Archimedean Continuity.

The next result verifies the asserted link between Classic Transitivity and Transitivity of the preference sets.

Lemma 18. *1. If $\succsim$ satisfies Classic Transitivity, $\Pi_{\succsim}$ satisfies Transitivity.*

2. If Π satisfies Transitivity, then $\succsim^{\Pi}$ satisfies Classic Transitivity.

3. $\succsim^{\Pi}$ satisfies Classic Transitivity and Classic Completeness if and only if Π satisfies Transitivity and Completeness.

Proposition 8 can be used to highlight an additional benefit of our preference sets approach: we obtain a new set of axioms that are one to one with expected utility theory.

Theorem 6. *For every $\Pi \subseteq \Delta(X \times X)$ the following are equivalent:*

1. $\succsim_{\Pi}$ satisfy Classic Completeness, Classic Strong B-Independence, and Classic Archimedean B-Continuity;

2. $\succsim_{\Pi}$ satisfies Classic Completeness, Classic Transitivity, Classic Archimedean Continuity, and Classic Strong Independence;

3. $\succsim_{\Pi}$ admits an expected utility representation.

Proof It immediately follows by combining Lemmata 15, 16, and 18, Remarks 4 and 5, Claim 12 and the vN-M expected utility theorem. ■

A.8 Choice from arbitrary sets

We now turn to an important question left open by the previous analysis: how the DM chooses from a finite set A of more than two alternatives. In general, since the correlation-sensitive decision criterion is intransitive, it is possible that, given a choice set A, no element of A is (weakly) preferred to all the other options.

To describe the decision maker's preferences when multiple alternatives are available, we will need to generalize the concept of choice rule. In particular, let $\Delta(X^n)$ be the joint distribution over the n dimensional outcomes, and $\Delta = \bigcup_{n\in\mathbb{N}} \Delta(X^n)$ be the set of all the joint distribution over a finite number of outcomes. A choice rule is a $\mathcal{C} : \Delta \rightrightarrows \mathbb{N}$ such that $\pi \in \Delta(X^n)$ implies that $\emptyset \neq \mathcal{C}(\pi) \subseteq \{1, ..., n\}$. The interpretation is that the choice rule takes as an input a $\pi \in \Delta(X^n)$ that describes the joint distribution over outcomes of n alternatives, and gives back the subset of these alternatives preferred by the DM given the correlation structure.

A natural question is whether, given a preference set Π that satisfies Completeness, Strong Independence, and Archimedean Continuity, there always exists a consistent choice rule. Formally, given a preference set Π that admits a correlation-sensitive representation with skew symmetric function ϕ a choice rule $\mathcal{C}$ is *consistent* with Π if for all $n \in \mathbb{N}$, for all $\pi \in \Delta(X^n)$, if $i \in \mathcal{C}(\pi)$, then

$$\forall j \in \{1, ..., n\} \qquad \sum_{(x,y)} \pi_{i,j}(x, y)\, \phi(x, y) \geq 0$$

where $\pi_{i,j}$ is the marginal distribution over alternatives i and j. That is, if the DM prefers to be paid according to the i alternative, then the preference set Π deems i preferable to every other j in their pairwise comparison.

It is immediate that if Π admits an expected utility representation, then there exists a choice function consistent with Π. Unfortunately, given the more general criterion's intransitivity, this may not be the case for some preference sets that satisfy Completeness, Strong Independence, and Archimedean Continuity, as shown in the example below.

However, in some situations, the DM may be able to randomize over the alternatives with a randomization device independent of the alternative under consideration. A stochastic choice rule is a $\mathcal{S} : \Delta \to \Delta(\mathbb{N})$ such that $\pi \in \Delta(X^n)$ implies that $\operatorname{supp} \mathcal{S}(\pi) \subseteq \{1, ..., n\}$. Since the randomization performed by the DM does not introduce any additional correlation, we extend the notion of consistency in a linear way. Given a preference set Π that admits a correlation-sensitive representation with skew symmetric function ϕ a stochastic choice rule $\mathcal{S}$ is *consistent* with Π if for all $n \in \mathbb{N}$, for all $\pi \in \Delta(X^n)$, if $\mathcal{S}(\pi) = \nu$, then

$$\forall \nu' \in \Delta(\{1, ..., n\}) \qquad \sum_{i,j \in \{1,...,n\}} \nu(i)\, \nu(j) \sum_{(x,y)} \pi_{i,j}(x,y)\, \phi(x,y) \geq 0.$$

Fortunately, we can extend a result by Kreweras (1961) to show that a consistent stochastic choice rule always exists.

Proposition 14. *If* Π *satisfies Completeness, Strong Independence and Archimedean Continuity, then there exists a stochastic choice rule* $\mathcal{S}$ *that is consistent with* Π*.*

Proof By Theorem 4 Π admits a correlation-sensitive representation with skew symmetric function ϕ. Let $n \in \mathbb{N}$ and $\pi \in \Delta(X^n)$. We have that

$$\begin{aligned}
& \max_{\nu \in \Delta(\{1,...,n\})} \min_{\nu' \in \Delta(\{1,...,n\})} \sum_{i,j \in \{1,...,n\}} \sum_{(x,y)} \nu(i)\, \nu'(j)\, \pi_{i,j}(x,y)\, \phi(x,y) \\
= & \min_{\nu' \in \Delta(\{1,...,n\})} \max_{\nu \in \Delta(\{1,...,n\})} \sum_{i,j \in \{1,...,n\}} \nu(i)\, \nu'(j) \sum_{(x,y)} \pi_{i,j}(x,y)\, \phi(x,y) \\
= & \min_{\nu' \in \Delta(\{1,...,n\})} \max_{\nu \in \Delta(\{1,...,n\})} \sum_{i,j \in \{1,...,n\}} \nu(i)\, \nu'(j) \left(- \sum_{(x,y)} \pi_{j,i}(x,y)\, \phi(x,y) \right) \\
= & - \max_{\nu \in \Delta(\{1,...,n\})} \min_{\nu' \in \Delta(\{1,...,n\})} \left(\sum_{i,j \in \{1,...,n\}} \sum_{(x,y)} \nu(i)\, \nu'(j)\, \pi_{i,j}(x,y)\, \phi(x,y) \right)
\end{aligned}$$

where the first equality follows from von Neumann's min-max theorem, the second

by skew symmetry of ϕ, and the last by simple algebra. Therefore,

$$\max_{\nu\in\Delta(\{1,...,n\})}\min_{\nu'\in\Delta(\{1,...,n\})}\sum_{i\in\{1,...,n\}}\sum_{j\in\{1,...,n\}}\sum_{(x,y)}\nu(i)\,\nu'(j)\,\pi_{i,j}(x,y)\,\phi(x,y)=0,$$

that is there exists $\nu_\pi\in\Delta(\{1,...,n\})$ such that for all $\nu'\in\Delta(\{1,...,n\})$,

$$\sum_{i\in\{1,...,n\}}\sum_{j\in\{1,...,n\}}\sum_{(x,y)}\nu_\pi(i)\,\nu'(j)\,\pi_{i,j}(x,y)\,\phi(x,y)\geq 0.$$

The result then follows by letting $\mathcal{S}(\pi)=\nu_\pi$ for all $\pi\in\Delta$. ■

Example 8 (The effect of salience on random choice)**.** *Suppose that the preference set* Π *admits a salience representation with salience function* $\sigma(x,y)=|x-y|$*. The DM faces three symmetric alternatives:* $\pi\in\Delta(X^3)$ *with* $\pi(3,1,2)=\pi(2,3,1)=\pi(1,2,3)=\frac{1}{3}$*. Here, choosing to be deterministically paid according to a single alternative is not consistent with* Π*, since for such a salience-sensitive DM* $\pi_{1,2},\pi_{2,1},\pi_{3,1}\in\hat{\Pi}$*. However, it is easy to see that the unique randomization consistent with* Π *sees the DM randomizing uniformly over the three acts. Next, suppose that the DM faces the joint distribution* $\pi'\in\Delta(X^4)$ *with* $\pi'(3,1,2,5)=\pi'(2,3,1,0)=\pi'(1,2,3,0)=\frac{1}{3}$*. Notice that for this salience-sensitive DM* $\pi'_{1,4},\pi'_{3,4}\in\hat{\Pi}$*, but* $\pi'_{4,2}\in\Pi$ *since when the second alternative is compared to the fourth, the realization where the fourth alternative pays* 5 *and the second pays* 1 *results sufficiently salient to tilt the comparison in favor of the fourth alternative. It is easy to see that when faced with the choice set* π'*, uniform randomization over the first three alternatives is no longer optimal for the agent and that the unique optimal randomization is* $\left(\frac{1}{2},0,\frac{1}{6},\frac{1}{3}\right)$*. Here, the fourth alternative plays a "stochastic decoy" effect: the probability of the other three alternatives are distorted to favor the ones that perform better in the salient state in which the decoy pays* 5*.* ▲

A.9 Analysis of the Rank-Based Version

In this section, we analyze the relative weaknesses of the alternative rank-based salience theory proposed in BGS. First, note that every function $\sigma : X \times X \to \mathbb{R}$ induces a rank on the support of π. More precisely, if for all $(x, y) \in \operatorname{supp} \pi$ we let

$$\hat{\sigma}_\pi (x, y) = |\{(x', y') \in \operatorname{supp} \pi : \sigma (x', y') > \sigma (x, y)\}|,$$

we obtain $|\operatorname{supp} \pi| > \hat{\sigma}_\pi (x, y) \geq 0$ with $\hat{\sigma}_\pi (x, y) = 0$ for the most salient pair of outcomes. Given these definitions, we can say when a preference relation admits a rank-based salience theory representation.

Definition 19. A preference set Π admits a rank-based salience representation if there exist a salience function $\sigma : \mathbb{R}^2 \to \mathbb{R}$ and $\beta \in (0, 1]$ such that

$$\pi \in \Pi \Leftrightarrow \sum_{(x,y) \in \operatorname{supp} \pi} (x - y) \, \beta^{\hat{\sigma}_\pi (x,y)} \pi (x, y) \geq 0. \tag{A.13}$$

Since $\beta \leq 1$, and $\hat{\sigma}_\pi$ is decreasing in the salience of a pair of outcomes, the decision criterion is overweighting the most salient joint realizations. Therefore, this criterion has the advantage of suggesting the main features of a salience-sensitive DM: she probabilistically aggregates the difference between what is paid by the two alternatives, with additional weight given to salient pairs of rewards. Notice that if $\beta = 1$, the agent is a risk-neutral EU maximizer.

A.9.1 Weakness

Rank-based salience theory captures the idea that the distortion in evaluating an event depends only on its relative salience. Hence, small perturbations in the amount paid in a state can dramatically change its evaluation. As outlined above, this decision criterion is intransitive, and it does not satisfy the weaker axiom of Transitive Consistency.[23] For a joint distribution π define the conditional row distribution of π

[23] For an in-depth analysis of Transitive Consistency, see Cerreia-Vioglio and Ok (2018), and Nishimura and Ok (2018). For examples of intransitive binary relations satisfying this axiom,

given $y \in \operatorname{supp} \pi_2$ as

$$\pi_y(x) = \frac{\pi(x,y)}{\sum_{x \in X} \pi(x,y)}.$$

Axiom 28 (Transitive Consistency)**.** *Let π, χ be such that $\pi_2 = \chi_2$ and for all $y \in \operatorname{supp} \pi_2$*

$$\pi_y \geq_{FOSD} \chi_y$$

then $\chi \in \Pi$ implies $\pi \in \Pi$.

Transitive Consistency is a minimum rationality requirement imposed on an intransitive DM. The underlying idea is that, under the joint distribution π, the row marginal has been improved *conditional* on every possible realization of the column marginal. This implies that $\pi_1 \geq_{FOSD} \chi_1$, and Transitive Consistency is satisfied both by regret theory and the smooth salience theory.

The following example illustrates the possible transitive inconsistencies of rank-based salience theory.

Example 9. *Let π and χ be*

π	5	11.5
7	1/3	0
9	0	2/3

χ	5	11.5
7	1/3	0
8.8	0	1/3
9	0	1/3

Suppose that we use the leading example of salience function proposed in BGS

$$\sigma(x,y) = \frac{|x-y|}{|x|+|y|}$$

and we set $\beta = 1/2$. We obtain $\sigma(7,5) > \sigma(9,11.5)$. Therefore $\pi \notin \Pi$ since

$$\frac{1}{3}[7-5] + \beta\frac{2}{3}[9-11.5] = \frac{1}{3} \cdot 2 - \frac{1}{2} \cdot \frac{2}{3} \cdot 2.5 < 0.$$

see Cerreia-Vioglio, Giarlotta, Greco, Maccheroni, and Marinacci (2020). Se also Kontek (2016) for a related critique of the rank-based model.

On the other hand, $\chi \in \Pi$ *since*

$$\sigma(7,5) > \sigma(8.8, 11.5) > \sigma(9, 11.5).$$

and

$$\begin{aligned} & \frac{1}{3}[7-5] + \beta\frac{1}{3}[8.8-11.5] + \beta^2\frac{1}{3}[9-11.5] \\ = \quad & \frac{1}{3}\cdot 2 - \frac{1}{2}\cdot\frac{1}{3}\cdot 2.7 - \frac{1}{4}\cdot\frac{1}{3}\cdot 2.5 > 0. \end{aligned}$$

A.10 Minor Proofs

Proof of Lemma 15 Let $p \succeq q$. Then, if $\pi \in \Delta(X \times X)$ and $(\pi_1, \pi_2) = (p, q)$, $\pi \in \Pi_{\succeq}$ by definition of $\Pi_{\succeq}$. However, since π was an arbitrary joint lottery with marginals p and q, by definition of $\succsim^{\Pi_{\succeq}}$, we have $p \succsim^{\Pi_{\succeq}} q$.

Let $p \succsim^{\Pi_{\succeq}} q$. Then, by definition of $\succsim^{\Pi_{\succeq}}$, $p \times q \in \Pi_{\succeq}$. But by definition of $\Pi_{\succeq}$ this means that $p \succeq q$. ■

Proof of Lemma 16 (1) Let $\pi \in \Delta(X \times X)$. Since $\succsim$ satisfies Classic Completeness, at least one between $\pi_1 \succsim \pi_2$ and $\pi_2 \succsim \pi_1$ holds. By definition of $\Pi_{\succsim}$ this implies that at least one between $\pi \in \Pi$ and $\bar{\pi} \in \Pi$ holds.

(2) Let $\pi \in \Delta(X \times X)$. Since $\succsim^{\Pi}$ satisfies Classic Completeness at least one between $\pi_1 \succsim^{\Pi} \pi_2$ and $\pi_2 \succsim^{\Pi} \pi_1$ holds, and this implies that at least one between $\pi \in \Pi$ and $\bar{\pi} \in \Pi$ holds. ■

Proof of Remark 4 Let $\pi, \chi \in \Pi_{\succsim}$ (resp. $\chi \in \hat{\Pi}_{\succsim}$) and $\lambda \in (0,1)$. By definition of $\Pi_{\succsim}$, $\pi_1 \succsim \pi_2$ and $\chi_1 \succsim \chi_2$ (resp. $\chi_1 \succ \chi_2$). Since $\succsim$ satisfies Classic Strong B-Independence, $\lambda\pi_1 + (1-\lambda)\chi_1 \succsim \lambda\pi_2 + (1-\lambda)\chi_2$ (resp. $\lambda\pi_1 + (1-\lambda)\chi_1 \succ \lambda\pi_2 + (1-\lambda)\chi_2$), and by definition of $\Pi_{\succsim}$, we have $\lambda\pi + (1-\lambda)\chi \in \Pi_{\succsim}$ (resp. $\lambda\pi + (1-\lambda)\chi \in \hat{\Pi}_{\succsim}$). ■

Proof of Lemma 17 Let $p, q, r, s \in \Delta(X)$ be such that $p \succ q$ and $r \not\succsim s$. We first show that there exists $\alpha \in (0,1)$ such that $\alpha p + (1-\alpha)r \succ \alpha q + (1-\alpha)s$.

Define $r_n = \left(1 - \frac{1}{n}\right) p + \frac{1}{n} r$ and $q_n = \left(1 - \frac{1}{n}\right) q + \frac{1}{n} s$. If $r_n \succ q_n$ for some $n \in \mathbb{N}$, the result follows by setting $\alpha = 1 - \frac{1}{n}$. Otherwise, by Classic Completeness of $\succsim$, we have $q_n \succsim r_n$ for all $n \in \mathbb{N}$, but by Classic Sequential Continuity this implies that $\lim_n q_n = q \succsim \lim_n r_n = p$, a contradiction.

The existence of $\beta \in (0,1)$ such that $\beta p + (1-\beta) r \not\succsim \beta q + (1-\beta) s$ follows from the first part and noticing that under Classic Completeness $r \not\succsim s \iff s \succ r$ and $\beta p + (1-\beta) r \not\succsim \beta q + (1-\beta) s \iff \beta q + (1-\beta) s \succ \beta p + (1-\beta) r$. ■

Proof of Remark 5 Let $\pi \in \hat{\Pi}_{\succsim}$ and $\chi \notin \Pi_{\succsim}$. By definition of $\Pi_{\succsim}$, this means that $\pi_1 \succ \pi_2$ and $\chi_1 \not\succsim \chi_2$. But then, by Classic Archimedean B-Continuity, there exists $\alpha \in (0,1)$ and $\beta \in (0,1)$ such that $\alpha\pi_1 + (1-\alpha)\chi_1 \succ \alpha\pi_2 + (1-\alpha)\chi_2$ and $\beta\pi_1 + (1-\beta)\chi_1 \not\succsim \beta\pi_2 + (1-\beta)\chi_2$. By definition of $\Pi_{\succsim}$, this means that $\alpha\pi + (1-\alpha)\chi \in \hat{\Pi}_{\succsim}$ and $\beta\pi + (1-\beta)\chi \notin \Pi_{\succsim}$. ■

Proof of Lemma 18

(1) Let $\pi, \chi, \rho \in \Delta(X \times X)$, with $\pi_2 = \chi_1$, $\rho_1 = \pi_1$, and $\rho_2 = \chi_2$, and $\pi \in \Pi_{\succsim}, \chi \in \Pi_{\succsim}$. By definition of $\Pi_{\succsim}$, we have $\rho_1 = \pi_1 \succsim \pi_2 = \chi_1$ and $\chi_1 \succsim \chi_2 = \rho_2$. Since $\succsim$ is transitive, this implies that $\rho_1 \succsim \rho_2$, and by definition of $\Pi_{\succsim}$, we have $\rho \in \Pi_{\succsim}$.

(2) Let $p, q, r \in \Delta(X)$ with $p \succsim^{\Pi} q$ and $q \succsim^{\Pi} r$. Let $\pi = p \times q$, $\chi = q \times r$, and let ρ be such that $\rho_1 = p$ and $\rho_2 = r$. Then π,$\chi \in \Pi$ by definition of $\succsim^{\Pi}$, and $\rho \in \Pi$ by Transitivity of Π. Since ρ was chosen arbitrarily among the joint lotteries with marginals p and r, $p \succsim^{\Pi} r$, and the result follows.

(3) ($\succsim^{\Pi}$ *satisfies Classic Transitivity and Classic Completeness*$\Rightarrow$ Π *satisfies Transitivity and Completeness*) Let $\pi, \chi, \rho \in \Delta(X \times X)$, with $\pi_2 = \chi_1$, $\rho_1 = \pi_1$, and $\rho_2 = \chi_2$, and $\pi \in \Pi, \chi \in \Pi$. Then, Classic Completeness of $\succsim^{\Pi}$ implies that $\pi_1 \succsim^{\Pi} \pi_2 \succsim^{\Pi} \chi_2$, and Classic Transitivity of $\succsim^{\Pi}$ implies $\rho_1 \succsim^{\Pi} \rho_2$, and the definition of $\succsim^{\Pi}$ implies $\rho \in \Pi$, that is Π satisfies Transitivity. Moreover, Π satisfies Completeness by Lemma 16.

(Π *satisfies Transitivity and Completeness*$\Rightarrow\succsim^{\Pi}$ *satisfies Classic Transitivity and Classic Completeness*) That $\succsim^{\Pi}$ satisfies Classic Transitivity follows from the part (2). For Classic Completeness, let $p, q \in \Delta(X)$. Define π as the product measure

$\pi = p \times q \in \Delta(X \times X)$. By Completeness of Π, either $\pi \in \Pi$ or $\bar{\pi} \in \Pi$. If $\pi \in \Pi$, let $\rho \in \Delta(X \times X)$ be an arbitrary element of $\Delta(X \times X)$ such that $\rho_1 = p$ and $\rho_2 = q$, and define $\chi = q \times q$. By Completeness, $\chi \in \Pi$, and by Transitivity $\pi \in \Pi$ and $\chi \in \Pi$ together imply that $\rho \in \Pi$. Since ρ was chosen arbitrarily among the joint lotteries with marginals p and q, $p \succsim^{\Pi} q$. Suppose $\bar{\pi} = q \times p \in \Pi$. Let $\rho \in \Delta(X \times X)$ be an arbitrary element of $\Delta(X \times X)$ such that $\rho_1 = q$ and $\rho_2 = p$, and define $\chi = p \times p$. By Completeness, $\chi \in \Pi$, and by Transitivity $\bar{\pi} \in \Pi$ and $\chi \in \Pi$ together imply that $\rho \in \Pi$. Since ρ was chosen arbitrarily among the joint lotteries with marginals q and p, $q \succsim^{\Pi} p$. Therefore, $\succsim^{\Pi}$ satisfies Classic Completeness. ■

Appendix B

Dynamic Opinion Aggregation

B.1 Introduction

In recent years, studying people's opinion dynamics and reciprocal influence has become of utmost importance for economic research. This is mainly due to the significant increase in social media usage and the formation of global social networks. Under the classical Bayesian approach, agents act as statisticians who try to estimate a fundamental parameter based on their neighbors' opinions. An alternative approach, which is more descriptive and tractable, considers agents who assign fixed weights to their neighbors and repeatedly take weighted averages of the opinions they observe. This is commonly known as the DeGroot linear updating rule. However, even among stationary updating rules, the DeGroot model is still quite unrealistic. In real life, individuals are often attracted to extreme or intermediate stances, and the set of their influencers varies and is not given by a fixed network of connections. These and other relevant properties are incompatible with simple repeated averaging and have made generalizing the DeGroot model and its insights the primary theoretical challenge in this literature.

While the convergence and long-run consensus properties of particular nonlinear rules have been widely studied in applied mathematics, computer science, and economics, a general treatment of convergence and consensus with nonlinear updating rules that naturally extends the tools of the DeGroot model is still missing. More-

over, the central question of information aggregation in large networks, which is the wisdom of the crowd hypothesis, has received much less attention for nonlinear updating rules, primarily due to technical challenges. These gaps in the social learning literature require new methodologies and mathematical tools to be closed.

This chapter addresses these gaps by analyzing a general and unifying class of stationary nonlinear updating rules. It answers questions on convergence, consensus, and information aggregation by developing new mathematical tools that are well suited to study nonlinear (and often nondifferentiable) rules and that generalize the ones used for the DeGroot model. We show that most of the insights of the DeGroot model can be generalized to this class of updating rules. However, we also highlight qualitative insights that are peculiar to nonlinear rules. For example, due to certain patterns of nonlinearities, agents may sometimes disregard some of their neighbors' opinions, reducing the number of effective connections and inducing long-run disagreement for finite populations. Moreover, for the wisdom of the crowd in large populations, we point out a trade-off between how connected society is and the nonlinearity of the opinion aggregator.

Robust opinion aggregators We consider agents on a network whose initial opinions equal a common fundamental parameter plus some agent-specific noise. Agents observe their neighbors' opinions and repeatedly incorporate them to update their own through functions that we call *robust opinion aggregators*. These aggregators map the last-period opinions of the neighbors of each agent into her current stance and satisfy the following natural properties:

1. **Normalization**: If the agents have reached a consensus, then none of them further updates her opinion.

2. **Monotonicity**: If two opinion profiles are such that the first coordinatewise dominates the second, then this relation is preserved after aggregation.

3. **Translation invariance**: If the same constant shifts each agent's opinion, then the updates are shifted accordingly.

The first two properties are straightforwardly interpreted as minimal trust in the neighbors' opinions. Translation invariance is equivalent to assuming that agents only care about the opinions' differences rather than their intrinsic levels and rules out explosive/chaotic dynamics. This property is a consequence of a robust loss-minimization procedure that provides a foundation and an interpretation of the updating rule proposed (cf. Section B.5). Importantly, the recent field studies that compare Bayesian to non-Bayesian social learning models have obtained evidence consistent with our properties. For instance, Chandrasekhar et al. (2020) find that if the sampled subjects reach a consensus, they remain stuck on their beliefs even when such behavior is objectively suboptimal: this is consistent with normalization. Similarly, they also find that most subjects respond monotonically to changes in their neighbors' opinions.

The properties of robust opinion aggregators imply that the influence among agents depends on their current opinions. This simple feature makes our model the first unifying framework to capture many documented descriptive phenomena that we illustrate in Section B.3.[1] In such framework, our main results deal with the long-run stability of opinions across two complementary dimensions. We first provide graph-theoretic conditions on robust opinion aggregators for different forms of convergence of opinions in finite populations. We then derive structural properties of robust opinion aggregators that either guarantee or prevent the identification of the fundamental parameter as the population grows.

The dynamics of robust opinion aggregation We first show that the opinions' *time averages* induced by *any* robust opinion aggregator *uniformly* converge so that a profile of long-run opinions always exists. This first benchmark result implies that an external agent can test the long-run learning properties of the updating procedure by computing time averages, a feature that we exploit in our results on large networks.

Moreover, this is the stepping stone for deriving convergence and consensus formation from the properties of the network structures associated with robust opinion

[1]We postpone the comparison with the existing models to the related literature.

aggregators. We say that an agent is *strongly influenced* by another if the former *always* reacts to variations in the latter's opinion, regardless of the current opinion profile in the society. We show that if each agent has at least one strong link and the induced *strong network* is aperiodic, then opinions converge. This result is powerful for two reasons. First, it guarantees that, in a comprehensive class of models, the sole iteration of the aggregation procedure always leads to a stable distribution of opinions in the population (i.e., a Nash equilibrium under a best-response dynamics interpretation). Second, it highlights the critical role of strong ties in society to stabilize opinions in the long run.

Alternatively, we say that an agent is *weakly influenced* by another if the former reacts to variations in the latter's opinion for *at least one opinion profile*, and we show that opinions always converge only if the *weak network* is aperiodic. Therefore, whenever these extreme networks coincide (for example, in the DeGroot model), opinions' convergence is characterized by network aperiodicity. However, whenever behavioral biases or robustness concerns in the updating rules induce a wedge between the two extreme networks, we cannot dispense from studying both to have a complete picture of the opinions' long-run behavior.

Instead, our contribution to convergence to *consensus* is more conceptual than technical. It illustrates how the strong and weak networks are the key objects for nonlinear opinion aggregation since extra conditions on them buy extra convergence properties. We show that if the strong network has a unique, strongly connected, and closed group, which is aperiodic, convergence to consensus always obtains. Moreover, a necessary condition for forming consensus, regardless of the initial opinions, is that the weak network has a unique, strongly connected, and closed group, which is aperiodic. Whenever the two networks coincide, convergence to consensus is fully characterized by the previous property. However, if they differ, then even in societies where every two agents share some form of connection, we might observe persistent disagreement in the long run due to the weakness of these connections. Compared to the existing literature on convergence to consensus, we are the first to link a network structure derived from a given normalized, monotone, and translation invariant

aggregator to convergence to consensus. However, several important works, such as Moreau (2005), provide sufficient and necessary conditions given a fixed vector of initial opinions that can be used as part of an alternative route to our result about consensus. We postpone to Section B.6 a detailed comparison with these works.

Vox populi, vox Dei? We next study the information-aggregation properties of robust opinion aggregators. In particular, we study whether the *wisdom of the crowd* is achieved, i.e., if, in large networks, the agents' opinions converge to a true fundamental parameter (cf. Golub and Jackson, 2010).

We define a given robust opinion aggregator's strong and weak influence vector. These objects respectively capture the minimal and maximal influences among agents in the long run and give us a tool to study the limit opinions' variability. If the long-run *weak influence* of every agent vanishes sufficiently fast as the population grows, then the variance of their opinions vanishes as well. Conversely, if the long-run *strong influence* of at least one agent remains positive, then the aggregation procedure does not wash out all the idiosyncratic variability. Vanishing variability and symmetry of the robust opinion aggregator and the errors guarantee that the long-run opinions coincide with the true parameter in the large population limit.

Notably, our large-population limit analysis does not presume convergence or consensus. Therefore, the previous finite-population conclusions determine how the opinions concentration in the large-population limit should be understood. When only convergence of time averages obtains, these results should be interpreted in terms of wisdom *from* the crowd; an external observer can identify the parameter by computing time averages of opinions. If standard convergence obtains, we have the usual wisdom *of* the crowd interpretation. In particular, even if consensus does not obtain for finite population sizes, a typical outcome in our model, our results still yield a form of "stochastic" consensus for large populations.

Even if the conditions above are interpretable, they might be computationally challenging to verify since they are expressed in terms of long-run influence. Therefore, we combine graph-theoretic conditions on the weak networks and a nonlinearity index

of the aggregators into more primitive sufficient conditions for the wisdom of the crowd under the maintained symmetry assumptions. First, the aggregators are wise when the nonlinearity index is bounded across population sizes and the degrees in the weak network are growing sufficiently fast. Second, even if the degrees are bounded, but their distribution is balanced, and the connectivity of the weak network (measured by its second-largest eigenvalue in modulus) is high relative to the nonlinearity index, wisdom obtains. For example, the former condition is satisfied in an Erdős–Rényi model with (sufficiently) slowly decreasing linking probability. In turn, the latter condition is satisfied by expander graphs with a sufficiently high (finite) degree or by the island model of Golub and Jackson (2012) with a moderate level of homophily.

Foundation of robust opinion aggregators The properties of robust opinion aggregators arise from the natural generalization of two foundations for non-Bayesian opinions' dynamics: repeated estimation of the underlying parameter with naive agents (cf. DeMarzo et al. (2003)) and best-response dynamics in coordination games (cf. Golub and Jackson (2012)). In particular, an opinion aggregator is robust if and only if there is a profile of distance-based loss functions with positive complementaries whose unique solution map coincides with the aggregator itself. Moreover, natural convexity and smoothness properties of the loss functions yield robust opinion aggregators with the sufficient (and necessary) conditions for convergence and consensus obtained in our main results. Therefore, it is possible to reinterpret these results in terms of convergence to Nash equilibria and the consistency of iterated robust estimation.

B.2 The model

This section introduces our model of opinion aggregation in social networks. Let $N = \{1, ..., n\}$, with $n \in \mathbb{N}$, denote a finite set of agents and let I be an *arbitrary* closed interval of $\mathbb{R}$ with nonempty interior denoting the set of possible opinions. Let $B = I^n \subseteq \mathbb{R}^n$ denote the set of opinion profiles $x = (x_i)_{i=1}^n$. For example, the opinion

profile may be the agents' subjective probability assessments of an event, and in this case, $I = [0, 1]$. In this chapter, we consider different (directed) networks. We identify them with an $n \times n$ adjacency matrix A', that is, $a'_{ij} = 1$ if there is a directed link from agent i to agent j, and $a'_{ij} = 0$ otherwise.

Time is discrete, $t \in \mathbb{N}$, and the initial opinion of agent $i \in N$ at period 0 is given by a signal $X_i^0 = \mu + \varepsilon_i$, where $\mu \in \mathbb{R}$ is an underlying fundamental parameter and each $\varepsilon_i : \Omega \to \mathbb{R}$ is a random variable defined over a common probability space $(\Omega, \mathcal{F}, P)$.[2] Let A denote the *observation network* with $N_i = \{j \in N : a_{ij} = 1\}$ denoting the *neighborhood* of agent i. The interpretation is that agent i can only observe the current opinions of her neighbors $j \in N_i$.

Let x_i^0 denote the realization of the period-0 opinion of agent i. We model the evolution of opinions in the following periods through an *opinion aggregator* $T : B \to B$ that for each profile of period-t opinions $x^t \in B$ returns the profile of period-$(t+1)$ updates $x^{t+1} = T(x^t)$. We let $T_i : B \to I$ denote the i-th component of T, the updating rule of agent i.[3] Let $e \in \mathbb{R}^n$ denote the vector whose components are all 1s.

Definition 20. Let T be an opinion aggregator. We say that:

1. T is *normalized* if and only if $T(ke) = ke$ for all $k \in I$.

2. T is *monotone* if and only if for each $x, y \in B$

$$x \geq y \implies T(x) \geq T(y).$$

3. T is *translation invariant* if and only if

$$T(x + ke) = T(x) + ke \qquad \forall x \in B, \forall k \in \mathbb{R} \text{ s.t. } x + ke \in B.$$

[2] For completeness, we present the stochastic structure of initial opinions here. However, this does not have a relevant role in the analysis until Section B.4 on the wisdom of the crowd.

[3] The network structure (N, A) can be reflected in the opinion aggregator T by assuming that for each $i \in N$ and for each $x, x' \in B$

$$x_j = x'_j \qquad \forall j \in N_i \implies T_i(x) = T_i(x').$$

It is a natural assumption satisfied by all our illustrations, but it can be dispensed with for the general analysis.

We say that T is *robust* if and only if T is normalized, monotone, and translation invariant.

Normalization requires that whenever all the agents share the same opinion, each of the next-period updates coincides with that opinion. Monotonicity embodies a form of trust of the agents in the opinions observed by others. Translation invariance naturally arises when agents only care about their opinions' differences, as we show in Section B.5. In our related work (2023), we provide a game-theoretic foundation that relaxes this property to translation subinvariance, that is, agents react less than proportionally to uniform shifts. All our main convergence results continue to hold.[4]

Robust opinion aggregators are rich enough to describe several behavioral phenomena that we illustrate below: aversion/attraction to extreme opinions, rank-dependent social influence, confirmatory bias, and pure right/left bias. Moreover, they nest the widely studied DeGroot model, where T is also linear: $T(x) = Wx$, for all $x \in B$. Here, $W \in \mathcal{W}$ is the matrix collecting the vectors of weights, and $\mathcal{W}$ denotes the collection of stochastic matrices. This simple aggregation rule arises from either best-response dynamics in coordination games with quadratic payoffs or naive repeated maximum-likelihood estimation of a location parameter under Gaussian signal. In both cases, each $T_i(x)$ is the minimizer over $c \in \mathbb{R}$ of the loss function

$$\sum_{j=1}^{n} w_{ij} (x_j - c)^2, \tag{B.1}$$

where, $w_i \in \Delta = \left\{ p \in \mathbb{R}^n_+ : \sum_{j=1}^n p_j = 1 \right\}$ is the i-th row of W. In Section B.5.1, we derive robust opinion aggregators from a more general *robust* loss-minimization problem that removes the quadratic and Gaussian assumptions. For this reason and the unifying role of the properties in Definition 20, we have called robust the aggregators

[4]A careful inspection of the proofs shows that our convergence result will continue to hold for opinion aggregators which are normalized, monotone, and Lipschitz continuous of order 1. Under normalization and monotonicity, this latter property is equivalent to translation subinvariance. A natural concern is that for some opinion domains, the shift from, e.g., $\frac{1}{4}$ and $\frac{1}{2}$ is perceived as larger than the shift from $\frac{1}{2}$ and $\frac{3}{4}$. If all the agents share this perception, all our results continue to hold after rescaling I according to the perceived differences. We thank an anonymous referee for this observation.

we analyze. Although natural, these properties exclude some extremely discontinuous behavior patterns, such as agents listening to each other only when their opinions are closer than some threshold. They also exclude updating rules where agents always give some weight to an exogenously fixed opinion, as in Friedkin and Johnsen (1990).

Turning to the analysis of opinions' dynamics, we deal with two kinds of limit of $\{T^t(x)\}_{t\in\mathbb{N}}$, the standard one induced by the supnorm $\|\ \|_\infty$ and the one of Cesaro (i.e., time-average limit):

$$\text{C-}\lim_t T^t(x) = \lim_\tau \frac{1}{\tau}\sum_{t=1}^{\tau} T^t(x)$$

where the limit on the right-hand side of the definition is the standard one.

Definition 21. Let T be an opinion aggregator. We say that T is *Cesaro convergent* if and only if $\text{C-}\lim_t T^t(x)$ exists for all $x \in B$. We say that T is *convergent* if and only if $\lim_t T^t(x)$ exists for all $x \in B$.

Given the initial opinions x^0, if the updates converge, then it is well known that Cesaro convergence obtains, and the time-average and the standard limit coincide. When T is Cesaro convergent, we define the *long-run opinion aggregator* $\bar{T} : B \to \mathbb{R}^n$ by

$$\bar{T}(x) = \text{C-}\lim_t T^t(x) \qquad \forall x \in B. \tag{B.2}$$

If convergence obtains, we study whether the profile of long-run opinions is represented by a unique consensus across all agents or by several coexisting conventions, i.e., long-run disagreement. We denote by $D \subseteq B$ the consensus subset, that is, $x \in D$ if and only if $x_i = x_j$ for all $i, j \in N$.

Definition 22. Let T be an opinion aggregator. We say that *convergence to consensus always obtains under* T if and only if T is convergent and $\bar{T}(x) \in D$ for all $x \in B$.

B.3 The dynamics of robust opinion aggregation

This section studies the long-run properties of opinions for a given population size.

B.3.1 Convergence of the time averages

Our first result shows that even if the updates of a robust opinion aggregator might not converge, their time averages always stabilize in the long run.

Theorem 7. *If T is a robust opinion aggregator, then T is Cesaro convergent. Moreover, the long-run opinion aggregator $\bar{T}$ is a robust opinion aggregator such that $\bar{T} \circ T = \bar{T}$, and if $\hat{B}$ is a bounded subset of B, then*

$$\lim_{\tau} \left(\sup_{x \in \hat{B}} \left\| \frac{1}{\tau} \sum_{t=1}^{\tau} T^t(x) - \bar{T}(x) \right\|_{\infty} \right) = 0. \tag{B.3}$$

The *Cesaro limit* is described by the long-run opinion aggregator $\bar{T}$ that, for each initial profile of stances $x \in B$, returns the long-run average opinion of each agent. In particular, $\bar{T}$ is robust and satisfies the fixed point equation $\bar{T} \circ T = \bar{T}$, hence generalizing the well-known notion of *eigenvector centrality* of the DeGroot model. Finally, whenever the initial opinions of the agents are known to belong to a bounded set, the initial realizations of their signals do not affect the *rate of convergence* of the time averages.

Median aggregator We now illustrate the content of Theorem 7 with a natural alternative to opinion aggregation via weighted means: the median aggregator. Assume that the agents best respond to the previous opponents' opinions, but instead of minimizing a weighted quadratic loss function (B.1), they minimize the weighted absolute deviations:

$$\sum_{j=1}^{n} w_{ij} |x_j - c| \qquad \forall x \in B, \forall c \in I \tag{B.4}$$

where the values w_{ij} are the entries of a stochastic matrix W. It is well known that the solution correspondence admits as a selection the robust opinion aggregator T,

$$T_i(x) = \min\left\{c \in \mathbb{R} : \sum_{j:x_j \leq c} w_{ij} \geq 0.5\right\} \qquad \forall x \in B, \forall i \in N, \tag{B.5}$$

that is, $T_i(x)$ is the (weighted) median of x.

Example 10. A group of agents $N = \{1, 2, 3, 4\}$ share their opinions $x^0 \in B = [0, 1]^4$. The weights assigned to the other agents are represented by the matrix

$$W = \begin{pmatrix} 0.4 & 0.3 & 0.3 & 0 \\ 0.3 & 0.4 & 0.3 & 0 \\ 0.1 & 0.1 & 0.2 & 0.6 \\ 0 & 0 & 0.6 & 0.4 \end{pmatrix}.$$

Aggregation through weighted *averages* would achieve consensus in the limit. However, the dynamics induced by using the median are qualitatively different.

If $x^0 = (x_1^0, 1, 1, 1)$, then the block of agents agreeing on the higher opinion is sufficiently large to attract agent 1 to the same opinion, and the limit (consensus) opinion of $(1, 1, 1, 1)$ is reached in one round of updating. Note that the initial opinion of agent 1 is irrelevant given the agreement of the other agents. Similarly, the same limit consensus obtains if agent 2 disagrees with the initial consensus, that is if $x^0 = (1, x_2^0, 1, 1)$.

Instead, convergence to consensus fails if the initial opinions of *both* agents 1 and 2 fall. If $x^0 = (0, 1/2, 1, 1)$, then the first round of updating is $x^1 = (1/2, 1/2, 1, 1)$, and this opinion segregation will be the limit outcome: a strongly connected society fails to reach consensus without a sufficiently large block of initial agreement. This highlights how with median aggregation, a *joint* deviation from consensus by a group of agents might be necessary to destabilize an initial consensus.[5]

If $x^0 = (0, 1/2, 0, 1)$, then the agents' first update is $x^1 = (0, 0, 1, 0)$ and agents 1

[5]Note that in the corresponding DeGroot model with matrix W, both an individual and a joint deviation would still lead to a consensus but on a different opinion.

and 2 never change their opinions again, whereas agents 3 and 4 keep on reciprocally switching their opinions. This shows that even convergence may not be guaranteed. However, given Theorem 7, we obtain that $\bar{T}(x^0) = (0, 0, 1/2, 1/2)$. ▲

On the one hand, the robust opinion aggregator defined in equation (B.5), with $w_{ii} = 0$ for all $i \in N$, yields a natural process of best-response dynamics under the payoffs of equation (B.4). In this case, Theorem 7 always guarantees that actions are going to stabilize on average over time, even when they do not converge. On the other hand, there is no compelling reason to assume that each agent has the same attraction for relatively central opinions.

For example, assume that the agents best respond to the previous opponents' opinions by computing a convex linear combination of an optimistic and a pessimistic aggregation. Formally, for each $i \in N$, consider a convex and closed set of probability weights $C_i \subseteq \Delta$, a weight $\alpha_i \in [0, 1]$, and let

$$T_i(x) = \alpha_i \min_{w_i \in C_i} \sum_{j=1}^{n} w_{ij} x_j + (1 - \alpha_i) \max_{w_i \in C_i} \sum_{j=1}^{n} w_{ij} x_j \qquad \forall x \in B. \tag{B.6}$$

In words, agent i is uncertain about the relative importance of the opinions of the other agents and this subjective uncertainty is represented by the set of possible weights C_i, while α_i measures the relative attractiveness toward lower stances. This opinion aggregator is robust. Thus, Theorem 7 still guarantees convergence of time averages. To obtain standard convergence, as for the linear case, we need extra graph-based conditions. But, differently from the DeGroot model, given the nonlinearity of T, there is no obvious notion of graph associated with it. In the next section, we show that two natural graphs $\underline{A}$ and $\bar{A}$ associated with T determine the long-run behavior of the agents' opinions. Indeed, for the aggregator in (B.6), we could either say that i is influenced by j if $w_{ij} > 0$ for *all* $w_i \in C_i$ or if $w_{ij} > 0$ for *some* $w_i \in C_i$. Intuitively, the resulting graphs $\underline{A}$ and $\bar{A}$ collect the links relevant under *every* scenario and those relevant under *some* scenario. In stark contrast with the linear case, T is not always convergent to consensus even if every two agents are directly connected under $\bar{A}$, that is, $\overline{a}_{ij} = 1$ for all $i, j \in N$. Nevertheless, Theorem 8 provides necessary and sufficient

conditions for convergence in terms of $\bar{A}$ and $\underline{A}$.

B.3.2 Stable long-run opinions

In the standard DeGroot model, convergence is tied to the properties of an underlying network structure. The latter can either be implicit and given by the indicator matrix $A(W)$ of W or be explicit and given by a primitive observation network.[6] Here, we follow the first approach and derive different network structures from a robust opinion aggregator T. The generalization of the second approach is postponed to Section B.5.2.

We recall some common terminology from the network literature first. Consider an arbitrary network A' and let $\emptyset \neq M \subseteq N$ denote an arbitrary group. The network A' is *nontrivial* if and only if for each $i \in N$ there exists $j \in N$ such that $a'_{ij} = 1$. A path in M is a finite sequence of agents $i_1, i_2, ..., i_K \in M$ with $K \geq 2$, not necessarily distinct, such that $a'_{i_k i_{k+1}} = 1$ for all $k \in \{1, ..., K-1\}$. In this case, the length of the path is $K-1$. A cycle in M is a path in M such that $i_1 = i_K$. A cycle is simple if and only if the only repeated index in the sequence is the starting (and ending) one.[7] We say that M is *strongly connected* if and only if for each $i, j \in M$ there exists a path in M such that $i_1 = i$ and $i_K = j$. We say that M is *closed* if and only if for each $i \in M$, $a'_{ij} = 1$ implies $j \in M$. We say that M is aperiodic if and only if the greatest common divisor of the lengths of its simple cycles is 1. Finally, we say that A' is aperiodic if and only if each closed group M is *aperiodic*.

In principle, there are multiple networks corresponding to the same robust aggregator T. We now give two natural definitions that formalize two extreme networks among agents induced by T. A piece of notation: $e^j \in \mathbb{R}^n$ denotes the j-th vector of the canonical basis.

Definition 23. Let T be an opinion aggregator. We say that j *strongly influences* i if and only if there exists $\varepsilon_{ij} \in (0,1)$ such that for each $x \in B$ and for each $h > 0$

[6]Formally, the indicator matrix $A(W)$ of an arbitrary $W \in \mathcal{W}$ is such that its ij-th entry is equal to 1 if w_{ij} is strictly positive and 0 otherwise.

[7]More formally, a cycle (of length $K-1$) is simple if and only if for each $k, k' \in \{1, ..., K-1\}$: $i_k = i_{k'} \implies k = k'$.

with $x + he^j \in B$

$$T_i\left(x + he^j\right) - T_i\left(x\right) \geq \varepsilon_{ij} h. \tag{B.7}$$

We say that $\underline{A}\left(T\right)$ is the *network of strong ties* of T if and only if for each $i, j \in N$ the ij-th entry is such that

$$\underline{a}_{ij} = \begin{cases} 1 & \text{if } j \textit{ strongly} \text{ influences } i \\ 0 & \text{otherwise} \end{cases}.$$

We say that j *weakly influences* i if and only if there exist $x \in B$ and $h > 0$ such that $x + he^j \in B$ and

$$T_i\left(x + he^j\right) - T_i\left(x\right) > 0.$$

We say that $\bar{A}\left(T\right)$ is the *network of weak ties* of T if and only if for each $i, j \in N$ the ij-th entry is such that

$$\bar{a}_{ij} = \begin{cases} 1 & \text{if } j \textit{ weakly} \text{ influences } i \\ 0 & \text{otherwise} \end{cases}.$$

Equation (B.7) reflects uniform responsiveness of i to j: no matter what is the current opinion profile, the update of i increases at least linearly in the opinion of j. In actual social networks, strong links characterize only a subset of all the connections: close friends, own past opinions (anchoring effect), or an extremely reliable source (more generally, the relational "strong ties" as in Granovetter, 1973 and Centola and Macy, 2007).

In principle, there might be additional links (i.e., relational "weak ties") not in $\underline{A}\left(T\right)$ that are active only under particular circumstances. For instance, a person can completely discard a distant friend's opinion when this is too extreme compared to the ones of the rest of her neighbors. In contrast, for topics involving potential high stakes risks (e.g., vaccinations), a person may well be influenced by the opinion of someone outside her personal network, especially when the latter reports an extremely negative stance (e.g., isolated serious adverse reactions to vaccines). These examples

motivate the second part of Definition 23. Intuitively, i is weakly influenced by j if there are circumstances under which a change in j's opinion affects her update.

It is plain to see that $\underline{A}(T) \leq \bar{A}(T)$, and if T is linear with matrix W, then $A(W) = \underline{A}(T) = \bar{A}(T)$. Therefore, it is impossible to separate these two extreme networks in the DeGroot model. For a general robust opinion aggregator T, the strong directed network $\underline{A}(T)$ is the *minimal* network underlying T, while the weak directed network $\bar{A}(T)$ is the *maximal*. As such, they are instrumental in providing respectively *sufficient* and *necessary* conditions for convergence.

Theorem 8. *Let T be a robust opinion aggregator. The following statements are true:*

1. *If the network of strong ties $\underline{A}(T)$ is aperiodic and nontrivial, then T is convergent.*

2. *If T is convergent, then the network of weak ties $\bar{A}(T)$ is aperiodic and nontrivial.*

Therefore, if $\underline{A}(T) = \bar{A}(T)$, then T is convergent if and only if $\underline{A}(T)$ is aperiodic and nontrivial.

The first part of the result builds on the uniform convergence of the time averages of T updates to obtain standard convergence. Specifically, we need to use a Tauberian condition for T that turns uniform Cesaro convergence into standard convergence. We show that such a condition can be expressed in terms of the network of strong ties, and in particular, it requires that it is aperiodic and nontrivial. We postpone to Section B.6 a more detailed sketch of the proof that also elaborates on the technical contributions of each step of the proof.

Even if an agent does not strongly influence another, this does not always prevent communication between the two. Coherently, the second part of Theorem 8 states that if there exists a cyclic behavior in a group that is closed with respect to weak ties, then there exists a profile of initial opinions such that the updates of this group will

not stabilize. Indeed, since the agents in this group are never affected by outsiders, the cycle cannot be broken.

The third part of the result significantly generalizes Golub and Jackson (2010), which states that aperiodicity of $A(W)$ characterizes convergence for linear aggregators. The class of robust opinion aggregators such that $\underline{A}(T) = \bar{A}(T)$ is much larger (see Proposition 18), but, as we illustrate with rank-dependent aggregators right below, in general, there exists a wedge between the two extreme networks $\underline{A}(T)$ and $\bar{A}(T)$.

Theorem 8 has important implications for our game-theoretic interpretation. Even if multiple closed groups do not strongly influence each other, simple best-response dynamics converge to a Nash equilibrium, provided that these groups are aperiodic under $\underline{A}(T)$. Instead, when T captures a process of pure information aggregation, it is natural to assume that information gathered in the past is not entirely dismissed in light of new evidence. This translates into the property that each agent strongly influences herself, a condition that guarantees convergence. Notably, in the empirical social learning literature, Chandrasekhar et al. (2020) find that most subjects' behavior is consistent with a form of own-history dependence, even when it is objectively suboptimal.

Corollary 4. *Let T be a robust opinion aggregator. If T is* self-influential*, that is $\underline{a}_{ii} = 1$ for all $i \in N$, then T is convergent.*

We next introduce a general class of robust opinion aggregators which illustrates both the flexibility of our model and our convergence results. Their distinctive feature is rank-dependent influence across agents: a property that we have already encountered with the median aggregator.

Rank-dependent influence Consider a stochastic matrix W whose positive entries implicitly define the observation network. Formally, we say that T^f is a *rank-*

dependent aggregator if and only if for each $i \in N$

$$T_i^f(x) = \sum_{j=1}^{n} x_{\pi(j)} \left[f_i \left(\sum_{l=1}^{j} w_{i\pi(l)} \right) - f_i \left(\sum_{l=1}^{j-1} w_{i\pi(l)} \right) \right] \qquad \forall x \in B, \tag{B.8}$$

where π is a permutation of N such that $x_{\pi(1)} \leq ... \leq x_{\pi(n)}$ and $f_i : [0,1] \to [0,1]$ is a weakly increasing *distortion function* such that $f_i(0) = 0$ and $f_i(1) = 1$.[8]

A flexible parametric distortion function is given by

$$f_i(s) = q_i^{\left(\frac{\ln s}{\ln q_i}\right)^{\alpha_i}} \qquad \forall s \in (0,1] \tag{B.9}$$

where $q_i \in (0,1)$ and $\alpha_i \in \mathbb{R}_{++}$.[9] The parameter α_i captures the attitudes of agent i with respect to extreme opinions: (relative to q_i) attraction ($\alpha_i \in (0,1)$) or aversion ($\alpha_i \in (1,\infty)$). The parameter q_i captures the relative concern of agent i for stating an opinion that is higher ($q_i \in (0,1/2)$) or lower ($q_i \in (1/2,1)$) than the opinions of her neighbors. To see why the parameter q_i captures the asymmetric concerns for disagreement of agent i, note that, as the aversion to extreme opinions increases ($\alpha_i \to \infty$), under a mild assumption, the corresponding rank-dependent aggregator converges pointwise to

$$T_i^{q_i}(x) = \min \left\{ c \in \mathbb{R} : \sum_{j:x_j \leq c} w_{ij} \geq q_i \right\} \qquad \forall x \in B, \tag{B.10}$$

that is, the weighted q_i-quantile. In particular, we get back to the weighted median in (B.5) when $q_i = 0.5$. The q_i-quantiles capture the idea of an extreme truncation of the sample of opinions effectively taken into account. Indeed, the essential feature of these

[8]The map $T_i^f : B \to I$ is a Choquet integral against the capacity obtained by distorting the probability vector $w_i \in \Delta$ with respect to the conjugated distortion $\tilde{f}_i(\cdot) = 1 - f_i(1 - \cdot)$ hence, T^f is robust. Note in particular that the functional form of T_i^f is analogous to the decision criterion in rank-dependent utility theory.

[9]Clearly, f_i is defined only on $(0,1]$, but it also admits a unique continuous extension to $[0,1]$. The extension takes value 0 in 0. In particular, we obtain Prelec's probability weighting function (1998) when $q_i = 1\backslash e$. More generally, using an f_i different from the identity map is a way to introduce a *perception bias* a la Banerjee and Fudenberg (2021) in a model of naive and nonequilibrium learning.

particular rank-dependent aggregators is the extreme flatness of the corresponding weight distortion function $f_i(s) = 1_{[q_i,1]}(s)$ for all $s \in [0,1]$. With this, for each opinion profile $x \in B$, agent i is only influenced by the neighbor with the opinion corresponding to the q_i-quantile of the distribution of opinions induced by the profile x and the weights $w_i \in \Delta$. In the case of continuous opinions, a less extreme form of truncation might be desirable. For example, agent i aggregates opinions with a trimmed mean with thresholds $\underline{q}_i, \bar{q}_i \in [0,1]$, $\underline{q}_i < \bar{q}_i$, if her distortion function is

$$f_i(s) = \begin{cases} 0 & \text{if} \quad s < \underline{q}_i \\ \frac{s-\underline{q}_i}{\bar{q}_i - \underline{q}_i} & \text{if} \quad \underline{q}_i \leq s \leq \bar{q}_i \\ 1 & \text{if} \quad s > \bar{q}_i \end{cases} \qquad \forall s \in [0,1]. \tag{B.11}$$

The q_i-quantile is the limit case in which both $\underline{q}_i$ and $\bar{q}_i$ converge to $q_i \in (0,1)$. Notice that flat regions of f_i imply that agent i disregards the opinions of some of her neighbors depending on the current ranking of opinions. For example, suppose that the opinion of j is currently the lowest among the opinions of the neighbors of agent i. If the weight that agent i puts on j's opinion is not too high, that is $w_{ij} < \underline{q}_i$, then i completely ignores j's opinion. Differently, whenever the weight on the opinion of j is high enough, that is $w_{ij} > \max\left\{\underline{q}_i, 1 - \bar{q}_i\right\}$, agent i will always be influenced by j regardless of the current opinion profile. We illustrate this point in a particular example.

Example 11 (The islands model)**.** Suppose that the agents are partitioned in m groups $\{M_p\}_{p=1}^m$, that is, $N = \cup_{p \in G} M_p$, where $M_p \cap M_{p'} = \emptyset$ for all $p, p' \in G = \{1, ..., m\}$ such that $p \neq p'$. For example, these groups might capture the agents' similar cultural or social backgrounds. Also, consider a *strongly connected* observation network A with $a_{ii} = 1$ for all $i \in N$. So far, there is no relation between the neighborhood N_i of an agent i and the only group she belongs to, denoted M_{p_i}. In order to relate these two objects, let us define the *internal linking fraction* of $i \in N$ as

$$\ell_i = \frac{|\{j \in M_{p_i} : a_{ij} = 1\}| - 1}{|N_i|}.$$

According to our interpretation of the groups, the ℓ_is capture the degree of homophily in the given network structure: agents with a high ℓ_i are connected with many neighbors belonging to their own group M_{p_i}. A stylized picture of real-world networks that has been fruitfully used in the literature (cf. Golub and Jackson, 2012) is the islands structure with a large internal linking fraction for each agent.

Let each N_i be such that $|N_i| \geq 3$. Consider the stochastic matrix W such that $w_{ii} = \beta \in (1/|N_i|, 1/2)$, $w_{ij} = \frac{1-\beta}{|N_i|-1}$ if $j \in N_i \setminus \{i\}$, and $w_{ij} = 0$ otherwise, for all $i \in N$. Suppose that each agent $i \in N$ aggregates the opinions she observes in her neighborhood using a trimmed mean T_i with weights given by W and $\underline{q}_i = 1 - \bar{q}_i = \alpha/2$ where $\alpha \in [0, 2\beta)$. In words, every agent computes the weighted average of the opinions she observes, discarding both the $\alpha/2$ highest and lowest opinions and never fully discarding her own previous opinion, that is, $\underline{A}(T) \geq I$. Therefore, T is convergent by Corollary 4. The DeGroot model, obtained as a particular case by setting $\alpha = 0$, would still predict convergence to consensus in the long run. However, if there is sufficiently high homophily, that is, $\ell_i > 1 - \alpha/2$ for all $i \in N$, then disagreement is a typical outcome for the long-run dynamics. We next illustrate this point by studying the opinions' evolution in the society when, starting from a consensus $ke \in B$, the stances of a nonempty subset $M \subseteq N$ of agents are shifted upwards, that is,

$$x_i^0 = \begin{cases} k + \delta & \text{if} \quad i \in M \\ k & \text{otherwise,} \end{cases} \qquad \forall i \in N$$

with $\delta > 0$ such that $k + \delta \in I$. For example, we can interpret this shock as follows: a subset of agents M is targeted by a marketing campaign and persuaded to increase the use of a certain technology. Crucially, the extent of opinion segregation in the new long-run dynamics will depend on the agents' identities in the subgroup in relation to the islands structure. If the shock is local, that is, $M = M_p$ for some $p \in G$, then the long-run limit will be such that $\lim_t T_i^t(x^0) > k$ if $i \in M$, and $\lim_t T_i^t(x^0) = k$ if instead $i \notin M$. Differently, if the shock is dispersed, that is $|M \cap M_p| \leq 1$ for all $p \in G$, and the self-influentiality β is low enough, then the long-run limit will be such that $\lim_t T_i^t(x^0) = k$ for all $i \in N$.

If the number of islands m is much greater than the size of each island $|M_p|$, then the dispersed shock involves a much larger subgroup of agents. Nevertheless, the deviation of each subgroup member is washed out within each island, and the original consensus is restored. Instead, the original consensus is broken if the targeted set of agents M is smaller but more inward-looking, as in the first case. This phenomenon resembles the so-called "complex contagion" theory of Centola and Macy (2007), whereby a few "long ties" are not sufficient to spread an increased opinion globally. It is supported by the evidence on technology adoption in developing countries. In contrast, in the DeGroot model, both shocks lead to the formation of a new higher consensus. ▲

Even if the observation network is strongly connected, there is no global convergence to consensus due to the wedge between the observation and the strong network. It is easy to see that whenever $\ell_i \geq 1 - \alpha/2$ for each $i \in N$, no agent strongly influences any agent, apart for herself. In general, the strong and the weak networks for rank-dependent aggregators are completely characterized by the distortion functions $(f_i)_{i=1}^n$ and the matrix of weights W. Agent j strongly influences i if and only if her incremental weight, $f_i\left(\sum_{l\in M\cup\{j\}} w_{il}\right) - f_i\left(\sum_{l\in M} w_{il}\right)$, with respect to *any* baseline group $M \subseteq N\setminus\{j\}$ of agents is strictly positive. Similarly, agent j weakly influences i if and only if her incremental weight with respect to *some* baseline group of agents is strictly positive. This shows that convergence of opinions to disagreement is a much more natural outcome for robust opinion aggregators even in completely connected societies.

Remark 6. *Suppose that the agents use a rank-dependent aggregator T^f with matrix of weights $W \in \mathcal{W}$. Consider two disjoint groups $\overline{N}, \underline{N} \subseteq N$. If the members of both groups distort sufficiently toward zero the total weights of the outsiders, that is,*

$$f_i\left(\sum_{j\in N\setminus\overline{N}} w_{ij}\right) = 0 \quad \forall i \in \overline{N} \quad \textit{and} \quad f_l\left(\sum_{j\in\underline{N}} w_{lj}\right) = 1 \quad \forall l \in \underline{N}, \tag{B.12}$$

then convergence to consensus does not *always obtain under T^f. For example, long-*

run disagreement arises whenever there is initial agreement within $\bar{N}$ *on* $b \in I$*, initial agreement within* $\underline{N}$ *on* $a < b$*, and all the other agents have intermediate opinions* $x_i \in [a, b]$*. In particular, equation (B.12) is compatible with an observation and a weak network,* $A(W)$ *and* $\bar{A}\left(T^f\right)$*, that are both* strongly connected. ▲

The remark shows that it is not possible to resort to known results on convergence to consensus for nonlinear opinion aggregation models to analyze this kind of long-run behavior (e.g., (2005)). In turn, Theorem 8 gives easy-to-check sufficient conditions, in terms of strong links, to assess convergence of opinions. Finally, as we can easily see in Example 11, the exact composition of these groups is flexible and might change depending on their initial stances.

B.3.3 Long-run consensus

Our following result shows that if we cannot partition the strong network into multiple strongly connected and closed groups, then convergence to consensus always obtains. Conversely, convergence to consensus implies that the weak network does not admit such a partition.

Proposition 15. *Let* T *be a robust opinion aggregator. The following statements are true:*

1. *If the network of strong ties* $\underline{A}(T)$ *is nontrivial, has a unique strongly connected and closed group* M*, and* M *is aperiodic under* $\underline{A}(T)$*, then convergence to consensus always obtains.*

2. *If convergence to consensus always obtains, then the network of weak ties* $\bar{A}(T)$ *is nontrivial, has a unique strongly connected and closed group* M*, and* M *is aperiodic under* $\bar{A}(T)$*.*

Therefore, if $\underline{A}(T) = \bar{A}(T)$*, then convergence to consensus always obtains if and only if* $\underline{A}(T)$ *is nontrivial, has a unique strongly connected and closed group* M*, and* M *is aperiodic.*

Point 1 states that if there exists a unique strongly connected set of agents in the society that do not have strong connections with the outsiders, then all the agents will eventually conform to this group. Instead, if even the weak ties are not sufficient to connect two disjoint subgroups, then *long-run disagreement* can occur. It is then critical to identify strong and weak ties in the society to understand whether an intervention might generate a global consensus or just a localized one. However, the last part of the result confirms a general principle for robust opinion aggregators: if weak and strong ties coincide, then the results for convergence and consensus of the DeGroot model extend plainly. We next completely characterize the long-run opinion aggregator for a case with this property.

Quasi-arithmetic biased aggregation and opinions' dispersion Consider agents that best respond to the previous opinions of the opponents at each period. Within this interpretation of our dynamics, a restriction imposed by the quadratic loss in (B.1) is that upward and downward discrepancies are felt as equally harming by every agent. It might be the case that (some) agents are more concerned with one or the other. A smooth and tractable robust opinion aggregator that takes into account these asymmetries is obtained by minimizing

$$\phi_i^\theta (x - ce) = \sum_{j=1}^{n} w_{ij} \left[\exp\left(\theta \left(x_j - c\right)\right) - \theta \left(x_j - c\right)\right] \qquad \forall x \in \mathbb{R}^n, \forall c \in \mathbb{R} \qquad \text{(B.13)}$$

where $\theta \neq 0$ and the values w_{ij} are the entries of a stochastic matrix W. In particular, whenever $\theta > 0$, upward deviations from i's current opinion are more penalized than downward deviations and vice versa whenever $\theta < 0$.

We next show that there exists a unique solution function T_i^θ for each minimization problem induced by ϕ_i^θ. In particular, for this parametric class, we derive an explicit formula for the induced robust long-run opinion aggregator.

Proposition 16. *Let I be bounded and let ϕ be the profile of loss functions $\left(\phi_i^\theta : \mathbb{R}^n \to \mathbb{R}_+\right)_{i=1}^n$ as in (B.13) with $W \in \mathcal{W}$ and $\theta \in \mathbb{R} \setminus \{0\}$. The following statements are true:*

1. *For each $i \in N$ we have that*

$$T_i^{\theta}(x) = \operatorname{argmin}_{c \in \mathbb{R}} \phi_i^{\theta}(x - ce) = \frac{1}{\theta} \ln \left(\sum_{j=1}^{n} w_{ij} \exp(\theta x_j) \right) \qquad \forall x \in B \quad \text{(B.14)}$$

and T^{θ} is a robust opinion aggregator with $\underline{A}\left(T^{\theta}\right) = \bar{A}\left(T^{\theta}\right) = A(W)$.

2. *For each $i \in N$ we have that*

$$\lim_{\theta \to \hat{\theta}} T_i^{\theta}(x) = \begin{cases} \max_{j:w_{ij}>0} x_j & \text{if } \hat{\theta} = \infty \\ \sum_{j=1}^{n} w_{ij} x_j & \text{if } \hat{\theta} = 0 \\ \min_{j:w_{ij}>0} x_j & \text{if } \hat{\theta} = -\infty \end{cases} \qquad \forall x \in B.$$

3. *If there exists a vector $s \in \Delta$ such that*

$$\lim_{t} W^t x = \left(\sum_{i=1}^{n} s_i x_i \right) e \qquad \forall x \in \mathbb{R}^n, \qquad \text{(B.15)}$$

then convergence to consensus always obtains under T^{θ} and

$$\bar{T}^{\theta}(x) = \frac{1}{\theta} \ln \left(\sum_{i=1}^{n} s_i \exp(\theta x_i) \right) e \qquad \forall x \in B.$$

Point 1 gives an explicit functional form for the opinion aggregator, proving that the time-invariant version of the Log-Sum-Exp model of Tahbaz-Salehi and Jadbabaie (2006) is also a robust opinion aggregator.[10] Point 2 shows that this functional form encompasses the linear case as a limit and allows for nonneutral behaviors toward the direction of disagreement. Equation (B.15) in point 3 is satisfied if and only if $A(W)$ has a unique strongly connected and closed group M and M is aperiodic under $A(W)$. In this case, we see how not just the network structure determines the limit influence of each agent, but the initial opinion also plays a key role. Indeed, the marginal contribution to the limit of agent i's initial opinion is proportional to $s_i \exp(\theta x_i)$. Therefore, when $\theta > 0$, the higher the initial signal realization of an in-

[10] Differently from us, (2006) allow for time changing connections, but they assume uniform weights for all neighbors.

dividual, the higher her marginal contribution to the limit is. This fact has extremely relevant consequences. For example, consider one of the classical applications of non-Bayesian learning, technology adoption in a village of a developing country, with an opinion vector representing how much the agents have invested in the new technology (e.g., the share of land cultivated with the new technology). There, $\theta > 0$ captures the idea that the most innovative members of the society have a disproportionate influence on the others, maybe because their performance attracts relatively more attention. If resources are limited, i.e., if the external actor can only increase adoption for an agent directly, relying on the network aggregation for the rest, the policy prescription is qualitatively different. Indeed, she should choose the agent j for which $s_j \exp(\theta x_j)$ is maximized, combining the standard eigenvector centrality s_j with a distortion increasing in the initial opinion x_j of agent j.

B.4 Vox populi, vox Dei?

In the previous section, we considered a given deterministic profile of initial opinions and studied their evolution. However, for any given population size, the stochastic nature of the vector of initial opinions $X = \mu + \varepsilon$ implies that the long-run outcome $\bar{T}(X)$ will be stochastic as well. This section considers large networks to study the aggregate variability and the accuracy of long-run opinions under robust opinion aggregation, following the approach pioneered by Golub and Jackson (2010). Their question is whether the long-run opinions approach the true mean in large networks, i.e., if a "law of large numbers" holds under DeGroot opinion aggregation. We take up that question for robust opinion aggregators. Remarkably, even seemingly very basic questions about this were unresolved. For example, take a large Erdős–Rényi network and assume that everyone uses a nonlinear rule such as rank-dependent influence. On the one hand, it seems that in a large network where everyone's neighborhood is small and influence locally look symmetric, there is no channel for anyone's idiosyncratic noise to become influential enough to disrupt a law of large numbers. On the other hand, existing techniques seem basically powerless against this question. We next

provide sufficient and necessary conditions for this concentration around the true parameters to hold.

Formally, we keep the same setup of Sections B.2 and B.3, with the caveat that here everything is parametrized by the size n of the population.

Assumptions In this section, we maintain the following assumptions:

1. $I = \mathbb{R}$.

2. For each $n \in \mathbb{N}$ we assume that $X_i(n) = \mu + \varepsilon_i(n)$ for all $i \in N$, where $\{\varepsilon_i(n)\}_{i \in N, n \in \mathbb{N}}$ is an array of uniformly bounded and independent random variables such that $\inf_{i \in N, n \in \mathbb{N}} \operatorname{Var}(\varepsilon_i(n)) \geq \sigma^2 > 0$.

Some additional notation is helpful for the following analysis.

Notation With $\hat{I}$, we denote a bounded open interval such that $X_i(n)(\omega) \in \hat{I}$ for all $\omega \in \Omega$, $i \in N$, and $n \in \mathbb{N}$. We denote by $\ell \overset{\text{def}}{=} \sup \hat{I} - \inf \hat{I}$ the *signal range*. Moreover, we denote the collection of probability vectors in $\mathbb{R}^n$ by Δ_n.

We are interested in whether a growing society becomes wise (cf. Golub and Jackson (2010)), that is, whether there is an efficient aggregation of the information available in the network in the limit.

Definition 24. Let $\{T(n)\}_{n \in \mathbb{N}}$ be a sequence of robust opinion aggregators. The sequence $\{T(n)\}_{n \in \mathbb{N}}$ has *vanishing variance* if and only if, for each $\iota \in \mathbb{N}$,[11]

$$\operatorname{Var}\left(\bar{T}_\iota(n)(X_1(n), ..., X_n(n))\right) \to 0. \tag{B.16}$$

The sequence $\{T(n)\}_{n \in \mathbb{N}}$ is *wise* if and only if, for each $\iota \in \mathbb{N}$,

$$\bar{T}_\iota(n)(X_1(n), ..., X_n(n)) \xrightarrow{P} \mu. \tag{B.17}$$

[11] Note the following innocuous abuse of notation (given our interest in limit results): for each $\iota \in \mathbb{N}$, the sequences in equations (B.16) and (B.17) are well defined only starting from $n \geq \iota$. In fact, an agent with position ι can only belong to a society with size n greater than or equal to ι. A similar observation applies throughout the section, in particular, in Theorem 9.

When equation (B.16) holds, the aggregation procedure neutralizes the idiosyncratic variability of the agents' opinions. If, in addition, the agents' limit opinions are unbiased, then they concentrate around μ, and equation (B.17) holds. If $T(n)$ is *linear* with *strongly connected* matrix $W(n)$, then $\bar{T}(n)$ is linear and represented by a matrix $\bar{W}(n)$ whose rows all coincide with the left Perron-Frobenius eigenvector $s(T(n)) \in \Delta_n$ of $W(n)$: a standard measure of network centrality. DeMarzo et al. (2003) as well as Golub and Jackson (2010) call $s(T(n))$ the *influence vector* and the latter show that $\{T(n)\}_{n\in\mathbb{N}}$ is wise if and only if $\lim_n \max_{k\in N} s_k(T(n)) = 0$, provided the errors $\varepsilon_i(n)$ have 0 mean. In this case, the vector $s(n)$ coincides with the gradient of $\bar{T}_i(n)$, thereby capturing the idea of the "marginal contributions" of the agents to the limit opinion of i.[12]

As suggested by Theorem 7, for robust opinion aggregators, the marginal contributions to the limit opinion are captured by the partial derivatives of $\bar{T}_i(n)$. Even if our opinion aggregators might not be (Frechet) differentiable, they are Lipschitz continuous,[13] hence almost everywhere differentiable by Rademacher's Theorem. Let $\mathcal{D}\left(\bar{T}(n)\right) \subseteq \hat{I}^n$ be the subset of $\hat{I}^n$ where $\bar{T}(n)$ is differentiable.

Definition 25. Let $T(n) : \mathbb{R}^n \to \mathbb{R}^n$ be a robust opinion aggregator and $i \in N$. We say that $\underline{s}_i(T(n)) \in \mathbb{R}^n$ is the *strong influence vector* for i given $T(n)$ if and only if

$$\underline{s}_{ij}(T(n)) = \inf_{x\in\mathcal{D}\left(\bar{T}(n)\right)} \frac{\partial \bar{T}_i(n)}{\partial x_j}(x) \qquad \forall j \in N.$$

We say that $\bar{s}_i(T(n)) \in \mathbb{R}^n$ is the *weak influence vector* for i given $T(n)$ if and only if

$$\bar{s}_{ij}(T(n)) = \sup_{x\in\mathcal{D}\left(\bar{T}(n)\right)} \frac{\partial \bar{T}_i(n)}{\partial x_j}(x) \qquad \forall j \in N.$$

As for the notions of networks associated with a robust opinion aggregator, there

[12] Observe that, compared to the results of the wisdom of the crowd result in(2010), we are allowing for a sequence of opinion aggregators that do not necessarily induce a consensus from every starting opinion, and so we may have $\bar{T}_\iota(n) \neq \bar{T}_{\iota'}(n)$ for some $n \in \mathbb{N}$ and $\iota, \iota' \in \{1, ..., n\}$. In that case, our definitions of vanishing variance and wise require that, for each fixed agent ι, respectively the variance of the long-run opinion is going to 0 and the long-run opinion is converging in probability to μ. This definition collapses to the one of(2010) under their additional assumption that $\bar{T}_\iota(n) = \bar{T}_{\iota'}(n)$ for all $n \in \mathbb{N}$ and $\iota, \iota' \in \{1, ..., n\}$

[13] See Lemma 20 in Appendix B.8.

are two natural definitions of influence vector. The values $\underline{s}_{ij}(T(n)), \bar{s}_{ij}(T(n)) \in \mathbb{R}$ are respectively the minimal and maximal influence that, under the opinion aggregator $T(n)$, the initial opinion of j exerts on the limit opinion of i. Observe that, whenever $T(n)$ is a robust opinion aggregator that satisfies 1 of Proposition 15, for each $i, l \in N$, we have $\underline{s}_i(T(n)) = \underline{s}_l(T(n))$ and $\bar{s}_i(T(n)) = \bar{s}_l(T(n))$, since $\bar{T}_i = \bar{T}_l$. Moreover, both definitions of influence vector above coincide with the one of Golub and Jackson whenever $T(n)$ is linear and strongly connected since $\underline{s}_i(T(n)) = \bar{s}_i(T(n)) = s(T(n))$ for all $i \in N$.

These objects are crucial to providing sufficient and necessary conditions for vanishing variance. To obtain also the wisdom of the crowd, the following additional symmetry assumptions are needed. We say that the array $\{\varepsilon_i(n)\}_{i\in N, n\in\mathbb{N}}$ is *symmetric* if and only if for each $i \in N$ and for each $n \in \mathbb{N}$, $\varepsilon_i(n)$ and $-\varepsilon_i(n)$ have the same distribution under P. Moreover, we say that the sequence $\{T(n)\}_{n\in\mathbb{N}}$ is *odd* if and only if $T(n)(-x) = -T(n)(x)$ for all $x \in \mathbb{R}^n$ and for all $n \in \mathbb{N}$.[14]

Theorem 9. *Let $\{T(n)\}_{n\in\mathbb{N}}$ be a sequence of robust opinion aggregators. The following statements are true:*

1. *If $\lim_n \sum_{j=1}^n \bar{s}_{\iota j}(T(n))^2 = 0$ for all $\iota \in \mathbb{N}$, then $\{T(n)\}_{n\in\mathbb{N}}$ has vanishing variance. If in addition $\{T(n)\}_{n\in\mathbb{N}}$ is odd and $\{\varepsilon_i(n)\}_{i\in N, n\in\mathbb{N}}$ is symmetric, then $\{T(n)\}_{n\in\mathbb{N}}$ is wise.*

2. *If $\limsup_n \max_{j\in N} \underline{s}_{\iota j}(T(n)) > 0$ for some $\iota \in \mathbb{N}$, then $\{T(n)\}_{n\in\mathbb{N}}$ does not have vanishing variance. In particular, $\{T(n)\}_{n\in\mathbb{N}}$ is not wise.*

Given $\iota \in \mathbb{N}$, the quantity $\sum_{j=1}^n \bar{s}_{\iota j}(T(n))^2$ is an upper bound for the sensitivity of $\bar{T}_\iota(n)$ to changes in the initial opinions of small subsets of agents. As long as this measure vanishes, the variance of the limit opinion of ι is going to 0. It is easy to show that this condition is implied by $\max_{j\in N} \bar{s}_{\iota j}(T(n)) = o\left(\frac{1}{\sqrt{n}}\right)$, that is, the *maximum weak influence* on ι is vanishing fast enough. Conversely, if the *maximum strong influence* on some agent ι is not vanishing, then the variability of her limit opinion

[14] In the foundation of robust opinion aggregators that we propose in Section B.5.1, loss functions that are symmetric with respect to opinions' deviations (i.e., even) induce odd opinion aggregators.

does not disappear, preventing agent ι from learning μ. Therefore, the wisdom of the crowd is achieved only if $\lim_n \max_{j\in N} \underline{s}_{\iota j}\left(T\left(n\right)\right) = 0$ for all $\iota \in \mathbb{N}$, paralleling the linear case.

Observe that, whenever each $T\left(n\right)$ is linear and strongly connected, the sufficient and necessary conditions for the wisdom of the crowd in points 1 and 2 are equivalent to $\lim_n \max_{j\in N} s_j\left(T\left(n\right)\right) = 0$: the condition of Golub and Jackson(2010) which characterizes the wisdom of the crowd for the DeGroot model.[15] Thus, we obtain their characterization as a particular case of our result. In general, there are two other conceptual differences between the previous results about the wisdom of the crowd and ours. First, we neither impose any parametric structure on the opinion aggregators nor assume that agents aggregate opinions according to functionals belonging to the same subclass (e.g., the median, quantiles, rank-dependent, quasi-arithmetic). Second, our results encompass the case of nonconvergent robust opinion aggregators. In such a case, $\bar{T}\left(n\right)$ is the limit of the updates' time averages. This extra layer of generality is helpful for the following question: can an external observer learn μ by observing only part of the updating dynamics of a subset of the agents, i.e., can she achieve the *wisdom from the crowd*? We have a positive answer under the conditions of point 1: the external observer can use $\bar{T}_\iota\left(n\right)$ as a consistent estimator of the underlying parameter, even if the agents' opinions are not converging. In addition, when $T\left(n\right)$ is also convergent for all $n \in \mathbb{N}$, we have the *wisdom of the crowd*: all agents learn the true parameter. Finally, as the proof of Theorem 9 clarifies, our results are not only qualitative, but also *quantitative*. For example, in point 1, not only do we prove that there is vanishing variance, but we provide an estimate of the variance, given a fixed population of size n.

The proof of Theorem 9 has the following steps. For point 1, we treat each $\bar{T}_i\left(n\right)$ as an estimator of μ and borrow techniques from large-deviation theory. In particular, we observe that McDiarmid's concentration inequality can be used to bound the variance of $\bar{T}_i\left(n\right)$ whenever its variations with respect to the signal realizations can

[15]Indeed, given $n \in \mathbb{N}$ and $i \in N$, if $\bar{s}_i\left(T\left(n\right)\right) \in \Delta_n$ (as in(2010)), then $\sum_{j=1}^n \bar{s}_{ij}\left(T\left(n\right)\right)^2 \leq \max_{j\in N} \bar{s}_{ij}\left(T\left(n\right)\right)$.

be bounded. Intuitively, these variations are proportional to the partial derivatives of $\bar{T}_i(n)$ with respect to the initial opinions of the other agents when these derivatives are defined. We can formalize this idea by using a version of the Mean Value Theorem for Lipschitz functions to show that each $\bar{s}_{ij}(T(n))$ bounds the changes of $\bar{T}_i(n)$ as X_j varies. With this, we obtain a bound on the variance of $\bar{T}_i(n)$ that vanishes as $\sum_{j=1}^{n} \bar{s}_{\iota j}(T(n))^2$ does, yielding the first part of point 1. Next, we show that if both the errors and the opinion aggregator are symmetric, then $\bar{T}_i(n)$ is an unbiased estimator, so it converges in probability to μ.

For point 2, we show that the assumption on the strong influence vector implies that the variance of the long-run opinion of agent ι remains bounded away from zero for every n. This happens because (up to selecting a subsequence) for every n, there exists an agent j_n with a strong influence of at least $\alpha \in (0,1)$ on ι. In turn, this implies that we can decompose the long-run opinion of ι as the convex linear combination of $X_{j_n}(n)$ (with weight α) and a monotone function of the opinions of all agents. By Harris inequality, the covariance between $X_{j_n}(n)$ and this monotone function is nonnegative. Therefore the overall variance of agent ι long-run opinion is at least $\alpha^2 \operatorname{Var}(X_{j_n}(n)) \geq \alpha^2 \sigma^2 > 0$.

B.4.1 Weak networks and the wisdom of the crowd

Point 1 of Theorem 9 provides an easy-to-interpret sufficient condition on the sequence of long-run opinion aggregators for both absence of aggregate variability and wisdom. However, it is important to have properties of the primitive sequence of robust opinion aggregators that induce long-run wisdom. To address this point via Theorem 9, we need to control the derivatives of the sequence of robust opinion aggregators $\{T(n)\}_{n\in\mathbb{N}}$ with their weak networks $\{\bar{A}(n)\}_{n\in\mathbb{N}}$. For each $n \in \mathbb{N}$ and $i \in N$, we denote the *degree* of i in $\bar{A}(n)$ by $\bar{d}_i(n) = \sum_{j\in N} \bar{a}_{ij}(n)$. We define the maximum and minimum degrees by $\bar{d}_{\max}(n) = \max_{i\in N} \bar{d}_i(n)$ and $\bar{d}_{\min}(n) = \min_{i\in N} \bar{d}_i(n)$, respectively. Recall that $\hat{I}^n$ is the set of possible initial opinion vectors. Similar to before, we denote by $\mathcal{D}(T(n)) \subseteq \hat{I}^n$ the subset of $\hat{I}^n$ where $T(n)$ is differentiable.

Definition 26. Let $\{T(n)\}_{n\in\mathbb{N}}$ be a sequence of robust opinion aggregators and $\kappa \geq 1$. The sequence $\{T(n)\}_{n\in\mathbb{N}}$ is κ-dominated if and only if

$$\frac{\partial T_i(n)}{\partial x_j}(x) \leq \frac{\kappa}{\bar{d}_i(n)} \qquad \forall x \in \mathcal{D}(T(n)) \tag{B.18}$$

for all $i, j \in N$ and for all $n \in \mathbb{N}$.

For a *fixed* $n \in \mathbb{N}$, since each $T(n)$ is Lipschitz continuous, we can always satisfy the inequality in (B.18) by choosing $\kappa(n) = \bar{d}_{\max}(n)$.[16] Therefore, a sufficient condition for the sequence $\{T(n)\}_{n\in\mathbb{N}}$ to be κ-dominated for some $\kappa \geq 1$ is that $\sup_{n\in\mathbb{N}} \bar{d}_{\max}(n) < \infty$. Here, κ measures the deviation of $T(n)$ from the uniform linear aggregation of the opinions of the weak neighbors. This deviation can take two forms: i) some neighbors may be more important than others; and ii) the relative weights may depend on the current opinion. The first form is already present in the linear model with nonuniform weights, while the second one is specific to robust opinion aggregators, as we next illustrate.

Example 12. Let $\left\{T^f(n)\right\}_{n\in\mathbb{N}}$ denote the sequence of rank-dependent aggregators with matrices of weights $\{W(n)\}_{n\in\mathbb{N}}$ and distortions $\{f_\iota\}_{\iota\in\mathbb{N}}$, with each f_ι continuous and locally Lipschitz on $(0,1)$.[17] This implies that there exists a set $F \subseteq (0,1)$ of measure 1 where each f_ι is differentiable. We assume that the weights are uniform over the (nontrivial) observation network, that is, for each $n \in \mathbb{N}$ and $i, j \in N$, it holds $w_{ij}(n) \in \{0, 1/|N_i(n)|\}$. In this case, we have that the inequality in (B.18) holds with $\kappa = \sup_{\iota\in\mathbb{N}} \sup_{x\in F} f'_\iota(x)$. If $\kappa < \infty$, then the sequence $\left\{T^f(n)\right\}_{n\in\mathbb{N}}$ is κ-dominated. For example, if all agents use the same distortion $f_\iota = \hat{f}$ which belongs to any of the cases in Figure 1, except for quantiles, then κ is finite. Alternatively, if all agents are using trimmed means with symmetric, but potentially heterogenous

[16] In general, we can choose a much smaller $\kappa(n)$ (cf. Example 12). That said, since $T(n)$ is monotone and translation invariant, observe that the gradient $\nabla T_i(n)(x)$ is a probability vector for all $i \in N$ and for all $x \in \mathcal{D}(T(n))$. This implies that $\kappa(n)$ can never be chosen to be smaller than 1. Moreover, it can be chosen to be 1 if and only if $T(n)(x) = W(n)x$ for all $x \in \mathbb{R}^n$, where $W(n)$ is the stochastic matrix of uniform weights associated with $\bar{A}(n)$. Intuitively, the less the derivative of T can change, the closer T is to being linear, and the smaller κ can be chosen. For these reasons, we interpret κ as an index of nonlinearity.

[17] For example, this is the case if each f_ι is continuous on $[0,1]$ and either convex or concave.

trimming cutoffs $\left(\underline{q}_\iota, 1-\underline{q}_\iota\right)_{\iota\in\mathbb{N}}$ such that $\sup_{\iota\in\mathbb{N}}\underline{q}_\iota < 1/2$, then $\left\{T^f(n)\right\}_{n\in\mathbb{N}}$ is κ-dominated with $\kappa = 1/\left(1-2\sup_{\iota\in\mathbb{N}}\underline{q}_\iota\right)$ and each $T^f(n)$ is odd. ▲

We now give two difference conditions under which a a κ-dominated sequence of odd robust opinion aggregators is wise. For each $n\in\mathbb{N}$, if $\bar{A}(n)$ is strongly connected and undirected, the stochastic matrix of uniform weights associated with $\bar{A}(n)$ (i.e., the matrix whose ij-th entry is $\bar{a}_{ij}(n)/\bar{d}_i(n)$) has n real eigenvalues. We denote by $\lambda_2(n)$ the second largest eigenvalue in modulus of this matrix (henceforth, SLEM): a standard measure of connectivity.

Proposition 17. *Let $\{T(n)\}_{n\in\mathbb{N}}$ be a κ-dominated sequence of odd robust opinion aggregators and $\{\varepsilon_i(n)\}_{i\in N,n\in\mathbb{N}}$ be symmetric. The following statements are true:*

1. *If $\lim_n \frac{\sqrt{n}}{\bar{d}_{\min}(n)} = 0$, then $\{T(n)\}_{n\in\mathbb{N}}$ is wise.*
2. *If the weak networks $\left\{\bar{A}(n)\right\}_{n\in\mathbb{N}}$ are undirected and strongly connected, $\sup_{n\in\mathbb{N}}\frac{\bar{d}_{\max}(n)}{\bar{d}_{\min}(n)} < \infty$, and $\sup_{n\in\mathbb{N}}\lambda_2(n) < \frac{1}{\kappa^2}$, then $\{T(n)\}_{n\in\mathbb{N}}$ is wise.*

The first part of the propostion shows that a sequence of odd robust opinion aggregators which is κ-dominated is wise, provided that the weak degree of each agent is increasing fast enough. On the one hand, the degree-growth condition in this statement is satisfied with high probability in standard random graph models such as the Erdős–Rényi model with (sufficiently) slowly decreasing linking probability.

On the other hand, many real-world networks exhibit bounded degrees, even when the population size grows. In these cases, we can still obtain the wisdom of the crowd at the cost of requiring a high level of connectivity in the weak networks compared to the nonlinearity index κ. We now observe that this joint condition is satisfied by multiple graph models. For example, within the class of the $\bar{d}(n)$-regular graphs, where each agent has exactly $\bar{d}(n)$ links, *Ramanujan graphs* have particularly high connectivity, with $\lambda_2(n) \leq 2/\sqrt{\bar{d}(n)}$. Importantly, for fixed $\bar{d}\in\mathbb{N}$, random graphs that are uniformly distributed over $\bar{d}$-regular graphs are "almost Ramanujan", in the sense that, with probability converging to 1, their SLEM will be lower than $2/\sqrt{\bar{d}}$, as n grows. Therefore, under this graph model, the connectivity condition reduces

to $\bar{d} > 4\kappa^4$. In the context of Example 12 with agents using trimmed means with symmetric cutoffs, this condition amounts to $\bar{d} > 4\left(\frac{1}{1-2\sup_{\iota\in\mathbb{N}}\underline{q}_\iota}\right)^4$, which is satisfied with reasonable parameters such as $\sup_{\iota\in\mathbb{N}}\underline{q}_\iota \leq 1/8$ and $\bar{d} \geq 13$.

Even if regular graphs constitute a benchmark structure given their balancedness properties, they still fail to capture the clustering of many real-world networks. The multi-type random graph model of Golub and Jackson (2012) is an example that overcomes this limitation allowing for homophily between agents of the same type. Notably, the realized degrees distribution is balanced, and the SLEM of the realized network is close to the SLEM of the associated deterministic network of types.[18] Therefore, in order to guarantee the wisdom of the crowd, we need that the SLEM of the type network generating the weak networks of $\{T(n)\}_{n\in\mathbb{N}}$ is small enough compared to their coefficient of nonlinearity $1/\kappa^2$. Moreover, in their leading case of an island model, this condition is always satisfied when the homophily index is low enough.

In Example 13 in Section B.5, we illustrate how to use the sufficient conditions of Proposition 17 to obtain the wisdom of the crowd in a model where agents repeatedly solve an estimation problem for the fundamental parameter μ.

Point 2 of Theorem 9 establishes that the persistent limit influence, of at least an individual, is sufficient to preserve the opinions' variability, even for large populations. It is not difficult to show that a more structural sufficient condition for persistent influence in terms of prominent families as in Golub and Jackson (2010) can be given.

B.5 Foundation of robust opinion aggregators

In this section, we give a microfoundation of robust opinion aggregators and their convergence and information-aggregation properties.

[18] The second statement is the content of their Theorem 2, while the balance condition is implied by their Lemma A.4. Golub and Jackson (2012) also point out that a small SLEM guarantees that convergence speed to μ does not explode as the population size increases.

B.5.1 A characterization of robust opinion aggregators

Here, we characterize robust opinion aggregators as the solution to a distance minimization problem. Formally, we endow each agent i with a loss function $\phi_i : \mathbb{R}^n \to \mathbb{R}_+$ and we assume that at each period the agent solves

$$\min_{c \in \mathbb{R}} \phi_i (x - ce) \tag{B.19}$$

where $x \in B$ is the opinion profile of the previous period. Intuitively, in choosing her current opinion c, agent i minimizes a loss function that penalizes the disagreement (i.e., differences of opinions) with the last-period opinions of her neighbors. We next impose two minimal restrictions on the profile of loss functions $\phi = (\phi_i)_{i=1}^n$.

Definition 27. The profile of loss functions ϕ is *sensitive* if and only if $\phi_i (he) > \phi_i (0)$ for all $i \in N$ and for all $h \in \mathbb{R} \setminus \{0\}$.

If agent i observes a unanimous opinion (including herself), then her loss is minimized by declaring that same opinion. In particular, under a best-response dynamics interpretation, sensitivity implies that all the constant profiles of actions are Nash equilibria of the induced game.

Definition 28. The profile of loss functions ϕ has *increasing shifts* if and only if for each $i \in N$, $z, v \in \mathbb{R}^n$, and $h \in \mathbb{R}_{++}$

$$z \geq v \implies \phi_i (z + he) - \phi_i (z) \geq \phi_i (v + he) - \phi_i (v) .$$

It has *strictly increasing shifts* if and only if the above inequality is strict whenever $z \gg v$.

The property of increasing shifts is a form of complementarity in disagreeing with two or more agents from the same side. It is implied by stronger properties usually required on supermodular games played on networks, such as degree complementarity.

We call *robust* a profile of loss functions that is sensitive and has increasing shifts. The collection of all these profiles is denoted by Φ_R. Given a robust profile of loss

functions ϕ, we denote with $T^{\phi} : B \to B$ an arbitrary selection of the argmin correspondence

$$T^{\phi}(x) \in \prod_{i=1}^{n} \operatorname{argmin}_{c \in \mathbb{R}} \phi_i (x - ce) \qquad \forall x \in B. \tag{B.20}$$

The selfmap T^{ϕ} is an opinion aggregator and describes one possible updating rule induced by ϕ. The next theorem shows that our loss-function-based updating procedure naturally generalizes the one of the DeGroot model without committing to any specific functional form (e.g., quadratic) of the loss function.[19]

Theorem 10. *Let T be an opinion aggregator. The following statements are equivalent:*

(i) There exists $\phi \in \Phi_R$ which has strictly increasing shifts and is such that $T = T^{\phi}$, that is, for each $i \in N$

$$T_i(x) = \operatorname{argmin}_{c \in \mathbb{R}} \phi_i (x - ce) \qquad \forall x \in B; \tag{B.21}$$

(ii) T is a robust opinion aggregator.

The property of strictly increasing shifts guarantees that $\operatorname{argmin}_{c \in \mathbb{R}} \phi_i (x - ce)$ is a singleton. However, it is violated in some interesting specifications of ϕ (see, e.g., equation (B.4)). In Proposition 23 in Appendix B.10, we show that the solution correspondence of problem (B.19) always admits a selection which is a robust opinion aggregator.

This theorem also suggests that, as in DeMarzo et al. (2003), we can interpret the induced opinion dynamics as repeated estimation of μ given the last-period neighbors' opinions. In particular, (2003) only studied the case of maximum likelihood updating with Gaussian initial signals. Instead, we follow the general robust statistics approach: the agents minimize a loss function.

[19] In particular, it is always possible to derive a DeGroot aggregator via the loss function (B.1).

B.5.2 Loss functions and long-run dynamics

Next, we illustrate how our foundation is linked to the convergence and wisdom results for robust opinion aggregators. We focus on the familiar and particularly tractable class of loss functions given by

$$\phi_i(z) = \sum_{j=1}^{n} w_{ij}\rho_i(z_j) \qquad \forall z \in \mathbb{R}^n, \forall i \in N$$

where $W \in \mathcal{W}$ is a stochastic matrix whose positive entries implicitly define the observation network, and $\rho = (\rho_i : \mathbb{R} \to \mathbb{R}_+)_{i=1}^n$ is a profile of positive functions. The weight w_{ij} captures the relative importance of the opinion of j as perceived by i. We call such a profile *additively separable* and write $\phi = (W, \rho)$. We denote the set of *robust* and *additively separable* profiles of loss functions with Φ_A. Easy computations yield that $(W, \rho) \in \Phi_A$ if and only if each ρ_i is convex, strictly decreasing on $\mathbb{R}_-$, and strictly increasing on $\mathbb{R}_+$. Additionally, if each ρ_i is strictly convex, then there exists a unique robust opinion aggregator T^ϕ that satisfies (B.20). Three relevant examples of robust opinion aggregators stemming from additively separable loss functions are the DeGroot aggregators, the quantile aggregators, and the opinion aggregator of Proposition 16.

Natural conditions on the profile of loss functions $\phi = (W, \rho)$ yield that both the strong network $\underline{A}\left(T^\phi\right)$ and the weak network $\bar{A}\left(T^\phi\right)$ coincide with the observation network given by W.[20]

Proposition 18. *Let $\phi = (W, \rho) \in \Phi_A$. If I is compact and ρ_i is twice continuously differentiable and strongly convex for all $i \in N$, then there exists a unique T^ϕ that satisfies (B.20) and $\underline{A}\left(T^\phi\right) = \bar{A}\left(T^\phi\right) = A(W)$.*

Note that Proposition 18, paired with Theorem 8 and Proposition 15, characterizes convergence and convergence to consensus in terms of the observation network $A(W)$, provided that each ρ_i is sufficiently smooth and convex.

[20] In general, we can prove a similar result for profiles of loss functions which are not additively separable. In this case, the assumptions of differentiability and strong convexity can also be weakened and replaced with a coercivity condition and a Lipschitz property of the difference quotients.

Finally, we illustrate how Proposition *17* can be applied to check the wisdom of the crowd in terms of the profile of loss functions. As a by-product, we obtain that, under Assumptions 1-3 of Section B.4, the wisdom of the crowd can be achieved as long as the minimum degree of connections gets larger as the population size increases.

Example 13. Consider a sequence $\{T(n)\}_{n\in\mathbb{N}}$ of odd robust opinion aggregators as in Section B.4 such that:

$$T_i(n)(x) \in \operatorname{argmin}_{c\in\mathbb{R}} \sum_{j\in N_i(n)} \frac{\rho_i(n)(x_j - c)}{|N_i(n)|} \qquad \forall x \in \mathbb{R}^n$$

where the profile of loss functions $\phi(n) = (W(n), \rho(n)) \in \Phi_A$ used by the agents satisfies the assumptions in Proposition 18 and is such that $\rho_i(n)(-z) = \rho_i(n)(z)$ for all $z \in \mathbb{R}$, for all $i \in N$, and for all $n \in \mathbb{N}$.[21] In this case, the weights $w_{ij}(n)$ of each $W(n)$ are uniform over their (nonempty) neighborhoods $N_i(n)$. Moreover, let $\{\varepsilon_i(n)\}_{i\in N, n\in\mathbb{N}}$ be symmetric and assume that there exists $\kappa \in \mathbb{R}$ such that

$$\frac{\rho_i''(n)(z)}{\rho_i''(n)(z')} \leq \kappa \quad \forall i \in N, \forall n \in \mathbb{N}, \forall z, z' \in [-\ell, \ell].$$

In particular, this condition is satisfied if $\rho_i(n) = \bar{\rho}$ for all $i \in N$ and for all $n \in \mathbb{N}$. By the Implicit Function Theorem, we have that $T(n)$ is differentiable and

$$\frac{\partial T_i(n)}{\partial x_j}(x) \leq \frac{\kappa}{|N_i(n)|} \leq \frac{\kappa}{\min_{k\in N}|N_k(n)|} \qquad \forall i, j \in N, \forall x \in \hat{I}^n, \forall n \in \mathbb{N}.$$

In words, the uniform bound on the sensitivity of the loss functions implies that the reciprocal weak influence among the agents can be bounded using the size of the minimal neighborhood in the growing network. By Proposition 18, we have that $\{T(n)\}_{n\in\mathbb{N}}$ is κ-dominated.

By Proposition 17, wisdom is reached if the minimal degree in the society is

[21] In this case, I is the closure of $\hat{I}$.

growing sufficiently fast, that is,

$$\frac{1}{\min_{k\in N}|N_k(n)|} = o\left(\frac{1}{\sqrt{n}}\right). \tag{B.22}$$

Alternatively, if each $A(W(n))$ is undirected and strongly connected, $\sup_{n\in\mathbb{N}} \frac{\max_{k\in N}|N_k(n)|}{\min_{k\in N}|N_k(n)|} < \infty$, and $\sup_{n\in\mathbb{N}} \lambda_2(n) < \frac{1}{\kappa^2}$, then $\{T(n)\}_{n\in\mathbb{N}}$ is wise. For example, when the signal range is 1 and $\bar{\rho}(z) = \alpha z^4 + (1-\alpha)z^2$ for some $\alpha \in (0,1)$, the SLEM condition becomes $\sup_{n\in\mathbb{N}} \lambda_2(n) < \left(\frac{1-\alpha}{5\alpha+1}\right)^2$. ▲

B.6 Related literature

The linear model This chapter belongs to the literature on non-Bayesian opinion aggregation and nests the benchmark DeGroot model (1974). Within this model, Golub and Jackson (2010) fully characterize convergence, convergence to consensus, and the wisdom of the crowd in terms of the network structure. For convergence, we significantly extend the scope of the conditions of Golub and Jackson (2010). We show that in our nonlinear model they are still sufficient for convergence and convergence to consensus when imposed on the strong network, while they are necessary when imposed on the weak network. For the wisdom of the crowd, we derive a general law of large numbers for robust opinion aggregators specializing to the one of(2010) for the linear case. Here the three main novelties are that: i) the maximal influence in the network, which generalizes the notion of maximal eigenvector centrality, has to vanish *sufficiently fast*; ii) both the noise distribution and the opinion aggregators must satisfy a symmetry property without which we only obtain the bias of the crowd; and iii) the necessary and sufficient conditions for the wisdom of the crowd must be expressed respectively in terms of the strong and the weak network, possibly creating a wedge that is not present in the linear model.

Convergence and the mathematics literature Our most novel contribution in terms of convergence is Theorem 8. Compared to the opinion aggregation literature in computer science and economics, our techniques are completely functional ana-

lytic. This is natural since our aggregators are nonlinear. Formally, this creates an immediate overlap with the literature of maps iteration and fixed point theory where the iterates $\{T^t(x)\}_{t\in\mathbb{N}}$ and their convergence are studied in order to find the fixed points of T. Using functional analysis in place of linear algebra comes at a cost. On the one hand, it is a language that is richer but not immediately amenable to graph-theoretic notions which are better expressed in terms of matrices. On the other hand, graph-theoretic properties are instead primitive within our framework. Thus, as a general contribution, our notions of networks of weak and strong ties build a useful link between nonlinear analysis and graph theory.

More in detail, the proof of point 1 of Theorem 8 relies on five major steps. We next comment on each step in relation to the literature. Given uniform Cesaro convergence of Theorem 7 and using Lorentz's Theorem, the first step (Lemma 22) observes that convergence of T is equivalent to asymptotic regularity. This technique seems to have first appeared in Bruck (1978), who applied it to the case of nonexpansive maps in Hilbert spaces. Because of this observation, showing that T is asymptotically regular is important. Conceptually, it poses the issue of what asymptotic regularity might mean at a graph-theoretic level. The second step moves to address these points. Proposition 19 is a quite simple yet new observation: if $\underline{A}(T)$ is nontrivial, then T admits a decomposition $T(x) = \varepsilon Wx + (1-\varepsilon)S(x)$ where $\varepsilon \in (0,1)$, W is a stochastic matrix such that $A(W) = \underline{A}(T)$, and S is a robust opinion aggregator. This grain of linearity is what allows us to bridge graph notions to the convergence properties of the operator T. Indeed, the third step (Lemma 23 and Proposition 20) shows that when W is a $\{0,1\}$-valued stochastic matrix that partitions the agents in m classes of agents that share the only individual in the class they observe (see Definition 29), then T is asymptotically regular. The third step thus offers an example of a graph-theoretic property encoded by W, which yields asymptotic regularity. In proving this step, we generalize the techniques of Edelstein and O'Brien (1978).[22] The

[22]Their case is more general in terms of the domain of T in that B can be any convex subset of a normed vector space. However, their generality comes at a cost. In our jargon, they are only studying the case in which T is self-influential, which in our case would only yield the intermediate step needed to derive Corollary 4.

decomposition used in the third step yields convergence, but it is a very special one. This concern is partially tamed by the fourth step (Lemma 24): if $\underline{A}(T)$ is aperiodic and nontrivial, then there exists $t \in \mathbb{N}$ such that T^t and T^{t+1} possess such a special decomposition, making T^t and T^{t+1} convergent. In the final step (proof of point 1 of Theorem 8), we prove that if T^t and T^{t+1} are convergent, so is T. To our knowledge, the second point of Theorem 8 does not have a counterpart in the literature.

Convergence to consensus and the computer science literature The multidisciplinary literature on repeated averaging procedures is mostly focused on convergence to consensus: a relevant question which we study in Section B.3.3. We now discuss the most important contributions to this issue. The closest paper to our functional approach is Moreau (2005), who considers the iteration of a nonlinear and time-varying operator on a Euclidean space. Neither our results nor the ones in (2005) nest the others. We restrict ourselves to time-homogeneous operators on a one-dimensional space and impose the additional condition of translation invariance (both papers assume normalization and monotonicity). The first two restrictions are substantial, and make our approach less useful for some engineering applications considered in (2005). Instead, the requirement of translation invariance only boils down to different continuity assumptions between the two papers. Indeed, as we mentioned in the text, the only implication of translation invariance used in our convergence result is Lipschitz continuity of order 1. Assumption 1.4 of Moreau (2005) imposes a different continuity condition on an ancillary function that controls the shrinking rate of the operator. More generally, Moreau (2005) can only be used, after some additional steps, to derive point 1 of Proposition 15, which we obtain from Theorem 8. However, Moreau (2005) does not address issues which are relevant to us such as convergence without consensus and the wisdom of the crowd. These questions significantly complicate the analysis and we need to resort to completely different techniques coming from functional analysis as discussed above. In addition, since our opinion aggregators are microfounded, under mild conditions, they inherit the primitive observation network structure of the foundation (see Proposition 18). This

imposes a strong discipline on the averaging process that allows us to provide bounds on the rate of convergence to consensus which are function of the underlying network.

Wisdom of the crowd and asymptotic learning Among the recent papers, the one closest to our wisdom of the crowd results is Molavi et al. (2018). However, both the questions and the methodology are rather different. First, they follow Jadbabaie et al. (2012) in considering social learning when agents both repeatedly receive external signals about an underlying state of the world and naively combine the beliefs of their neighbors. Instead, we follow the wisdom of the crowd approach of Golub and Jackson (2010), and we study the long-run opinions as the size of the society grows to infinity. Therefore, we single out the role of the network structure and the opinion aggregator in efficiently combining the agents' *initial* information as the network's size increases. For the questions we explore, log-linear aggregators a la Molavi et al. (2018) can be studied in an equivalent linear system, thus making use of the results developed for the DeGroot model and its time-varying versions. So, our results cover their aggregators too after a suitable transformation.

Other related contributions Both Mueller-Frank (2018) and Arieli et al. (2021) address different robustness concerns in a social learning setting: in Mueller-Frank (2018)it is with respect to external manipulation of the initial opinions, while in Arieli et al. (2021) it is with respect to the initial information structure of the agents.

Finally, our results also make use of some techniques coming from decision theory, and in particular Ghirardato et al. (2004), Maccheroni et al. (2006), and Schmeidler (1989). Ghirardato et al. (2004), Maccheroni et al. (2006) are the first to study functionals that satisfy normalization, monotonicity, and translation invariance, using nonstandard differential techniques. These techniques turn out to be particularly useful when we discuss the wisdom of the crowd. The third paper introduces the class of comonotonic additive functionals that include rank-dependent aggregators. Compared to these papers, we instead consider (iterations of) operators as opposed to functionals. However, even under the usual decision theoretic interpretation, our

machinery and convergence results turn out to be useful, as shown in Cerreia-Vioglio et al. (2023).

B.7 Conclusion

We see our results on the wisdom of the crowd as a natural starting point for further work. In Section B.4.1, we considered a sequence of robust opinion aggregators $\{T(n)\}_{n\in\mathbb{N}}$ and a derived sequence of (uniform) DeGroot aggregators $\{W(n)\}_{n\in\mathbb{N}}$. Each $W(n)$ was constructed from the networks of weak ties $\bar{A}(n)$ which we assumed to be undirected. In a nutshell, we showed that if the Jacobian of each $T(n)$, whenever defined, is uniformly dominated by the corresponding $W(n)$, then the wisdom of the crowd holds, provided the dominating graphs exhibit enough connectivity. A careful inspection of the proof shows that $W(n)$ does not need to be induced by the network of weak ties. For example, it can be induced by any undirected multigraph and the result would still hold. In both cases, connectivity is measured by the second largest eigenvalue in modulus, which can be computed thanks to the graphs being undirected. It remains an open question if the same type of result holds true when the graph is not assumed to be undirected, for example, by replacing the eigenvalue measure with another coefficient of ergodicity.

On a more applied side, our results can be important tools for studying the transmission of idiosyncratic shocks to aggregate fluctuations in large economies. Even if we derived $\bar{T}$ as the operator mapping initial opinions to long-run opinions, our Theorem 9 would apply to any nonlinear operator with the same properties. For example, we might consider a standard macroeconomic model of production networks and derive the equilibrium output and prices as functions of the idiosyncratic shocks of the firms. In their seminal paper, Acemoglu et al. (2010) obtain *linear* equilibrium maps and provide sufficient conditions for the persistence of aggregate fluctuations in large economies. In our language, this means a non-zero asymptotic variance as $n \to \infty$. Under more general specifications of the production functions or, perhaps more interestingly, under endogenous network formation (see, e.g., Acemoglu and Azar, 2020),

the equilibrium maps might well be nonlinear, but still satisfy our properties. Therefore, our results would be the first step to extend and test the results of (2010) in these more general and realistic settings. In all these cases, it would be interesting to derive the sufficient and necessary conditions for persistent aggregate fluctuations on the equilibrium operators from properties of the primitives, in the spirit of our Proposition 17. This is the subject of current investigation.

Another avenue for future work explores the role of robust opinion aggregators as a bridge between DeGroot-style continuous opinion aggregators and diffusion/contagion of a binary behavior such as adopting new technology. Indeed, (generalizations of) the discrete-opinion models of Morris (2000), Kempe et al. (2003), Centola and Macy (2007), and Muller-Frank and Neri (2021) can be obtained by considering a subclass of robust opinion aggregators with the property that each agent's updated opinion exactly coincides with one of the neighbors' opinions observed in the last period, a property that linear aggregators rule out. In the working paper Cerreia-Vioglio, Corrao, and Lanzani (2020), we show how our framework can deal with discrete (e.g., binary) opinions and obtain a result about convergence in that case. Obtaining sharper results on the wisdom of the crowd for such aggregators is an interesting open question.

B.8 Appendix: convergence

All the missing proofs are in the Supplementary Appendix (see Section B.11.1). The next three ancillary lemmas highlight the properties of T and the limiting operator $\bar{T}$, whenever it exists. Their proofs are based on routine arguments.

Lemma 19. *Let T be an opinion aggregator. The following statements are true:*

1. *If T is robust, then it admits an extension $S : \mathbb{R}^n \to \mathbb{R}^n$ which is also robust.*

2. *If T is normalized and monotone, then $\|T^t(x)\|_\infty \leq \|x\|_\infty$ for all $x \in B$ and for all $t \in \mathbb{N}$.*

Lemma 20. *If T is a robust opinion aggregator, then T^t is nonexpansive (i.e., Lipschitz continuous of order 1) for all $t \in \mathbb{N}$. In particular, T is nonexpansive.*

Despite being easy to derive, the property of nonexpansivity plays an important role in what follows and it also rules out the presence of chaotic behavior. The proof of next lemma instead relies on the property of "being a limit". It thus shows that the properties of T are often inherited by $\bar{T}$, provided the latter exists.

Lemma 21. *Let T be an opinion aggregator. If T is Cesaro convergent, then $\bar{T} : B \to B$, as defined in equation (B.2), is well defined and $\bar{T} \circ T = \bar{T}$. Moreover,*

1. *If T is nonexpansive, so is $\bar{T}$. In particular, $\bar{T}$ is continuous.*
2. *If T is normalized and monotone, so is $\bar{T}$.*
3. *If T is robust, so is $\bar{T}$.*
4. *If T is odd, so is $\bar{T}$, provided I is a symmetric interval, that is, $k \in I$ if and only if $-k \in I$.*

We can now prove that any sequence of updates of a robust opinion aggregator converges a la Cesaro and this convergence is uniform on bounded subsets of B.

Proof of Theorem 7. Consider $x \in B$. By point 2 of Lemma 19, we have that $\{T^t(x)\}_{t\in\mathbb{N}}$ is a bounded sequence and, in particular, relatively compact. By Lemma 20, T is nonexpansive. By Baillon et al. (1978), we can conclude that C-$\lim_t T^t(x)$ exists for all $x \in B$. By Lemma 21, $\bar{T}$ is a robust opinion aggregator such that $\bar{T} \circ T = \bar{T}$. Next, consider a bounded subset $\hat{B}$ of B. Define by $\tilde{B}$ the closed convex hull of $\hat{B}$. Since $\hat{B}$ is bounded and B is closed and convex, $\tilde{B}$ is a closed and bounded subset of B and, in particular, compact. For each $\tau \in \mathbb{N}$ define $S_\tau : \tilde{B} \to \mathbb{R}^n$ by

$$S_\tau(x) = \frac{1}{\tau}\sum_{t=1}^{\tau} T^t(x) \qquad \forall x \in \tilde{B}.$$

By Lemma 20, S_τ is well defined and nonexpansive for all $\tau \in \mathbb{N}$. The collection $\{S_\tau\}_{\tau\in\mathbb{N}}$ belongs to the space $C\left(\tilde{B}, \mathbb{R}^n\right)$ of continuous functions from $\tilde{B}$ to $\mathbb{R}^n$. This

space is a Banach space once endowed with the supnorm: $\|f\|_* = \sup_{x\in\tilde{B}} \|f(x)\|_\infty$ for all $f \in C\left(\tilde{B}, \mathbb{R}^n\right)$. Since $\{S_\tau\}_{\tau\in\mathbb{N}}$ is a collection of nonexpansive maps, this implies that the sequence $\{S_\tau\}_{\tau\in\mathbb{N}} \subseteq C\left(\tilde{B}, \mathbb{R}^n\right)$ is equicontinuous. By contradiction, assume that $S_\tau \overset{\| \|_*}{\nrightarrow} \bar{T}_{\left|\tilde{B}\right.}$. This would imply that there exist $\varepsilon > 0$ and a subsequence $\{S_{\tau_m}\}_{m\in\mathbb{N}} \subseteq \{S_\tau\}_{\tau\in\mathbb{N}}$ such that $\left\|S_{\tau_m} - \bar{T}_{\left|\tilde{B}\right.}\right\|_* \geq \varepsilon$ for all $m \in \mathbb{N}$. By the Arzela-Ascoli Theorem and since $\{S_{\tau_m}\}_{m\in\mathbb{N}}$ is equicontinuous and $\{S_{\tau_m}(x)\}_{m\in\mathbb{N}} \subseteq \mathbb{R}^n$ is bounded for all $x \in \tilde{B}$, this would imply that there exists a subsequence $\left\{S_{\tau_{m(l)}}\right\}_{l\in\mathbb{N}}$ and a function $\hat{S} \in C\left(\tilde{B}, \mathbb{R}^n\right)$ such that $\lim_l \left\|S_{\tau_{m(l)}} - \hat{S}\right\|_* = 0$. By the previous part of the proof, recall that $\lim_\tau S_\tau(x) = \bar{T}(x)$ for all $x \in \tilde{B}$. By definition of $\| \|_*$, it would follow that $\bar{T}(x) = \lim_l S_{\tau_{m(l)}}(x) = \hat{S}(x)$ for all $x \in \tilde{B}$, that is, $\bar{T} = \hat{S}$ on $\tilde{B}$. This would imply that $0 < \varepsilon \leq \lim_l \left\|S_{\tau_{m(l)}} - \bar{T}_{\left|\tilde{B}\right.}\right\|_* = 0$, a contradiction. We can conclude that

$$0 \leq \lim_\tau \sup_{x\in\hat{B}} \left\|\frac{1}{\tau}\sum_{t=1}^{\tau} T^t(x) - \bar{T}(x)\right\|_\infty \leq \lim_\tau \sup_{x\in\tilde{B}} \left\|\frac{1}{\tau}\sum_{t=1}^{\tau} T^t(x) - \bar{T}(x)\right\|_\infty = \lim_\tau \left\|S_\tau - \bar{T}_{\left|\tilde{B}\right.}\right\|_* = 0,$$

proving the last part of the statement. ■

We next prove our first result on standard convergence: Theorem 8. We begin by presenting few facts which are useful for proving point 1. First, we identify a technical property, termed asymptotic regularity, which characterizes convergence. Second, we show how $\underline{A}(T)$ being nontrivial is equivalent to T having a useful decomposition. Finally, via this decomposition, we show that aperiodicity of $\underline{A}(T)$ yields asymptotic regularity, hence convergence. We then prove point 2 of Theorem 8 for an important special case: N strongly connected under $\bar{A}(T)$. The general case then follows by observing that a robust opinion aggregator can be restricted to any strongly connected component of $\bar{A}(T)$ and retain its properties, including convergence.

Lemma 22. *Let T be a robust opinion aggregator. The following statements are equivalent:*

(i) T is asymptotically regular, that is, $\lim_t \|T^{t+1}(x) - T^t(x)\|_\infty = 0$ for all $x \in B$;

(ii) T is convergent.

Proposition 19. *Let T be a robust opinion aggregator. The following statements are equivalent:*

(i) $\underline{A}(T)$ is nontrivial;

(ii) There exist $W \in \mathcal{W}$ and $\varepsilon \in (0,1)$ such that

$$T(x) = \varepsilon W x + (1-\varepsilon) S(x) \qquad \forall x \in B \tag{B.23}$$

where S is a robust opinion aggregator.

Moreover, we have that W in (ii) can be chosen to be such that $A(W) = \underline{A}(T)$.

Proof. (i) implies (ii). For each $i, j \in N$ if j strongly influences i, consider $\varepsilon_{ij} \in (0,1)$ as in (B.7) otherwise let $\varepsilon_{ij} = 1/2$. Define $\tilde{W}$ to be such that $\tilde{w}_{ij} = \underline{a}_{ij}\varepsilon_{ij}$ for all $i, j \in N$ where $\underline{a}_{ij}$ is the ij-th entry of $\underline{A}(T)$. Since each row of $\underline{A}(T)$ is not null, for each $i \in N$ there exists $j \in N$ such that $\underline{a}_{ij} = 1$ and, in particular, $\tilde{w}_{ij} > 0$. This implies that $\sum_{l=1}^{n} \tilde{w}_{il} > 0$ for all $i \in N$. Define also $\varepsilon = \min\{\min_{i \in N} \sum_{l=1}^{n} \tilde{w}_{il}, 1/2\} \in (0,1)$. Define $W \in \mathcal{W}$ to be such that $w_{ij} = \tilde{w}_{ij} / \sum_{l=1}^{n} \tilde{w}_{il}$ for all $i, j \in N$. Clearly, we have that for each $i, j \in N$

$$w_{ij} > 0 \iff \tilde{w}_{ij} > 0 \iff \underline{a}_{ij} = 1. \tag{B.24}$$

This yields that $A(W) = \underline{A}(T)$. Next, consider $x, y \in B$ such that $x \geq y$. Define $y^0 = y$. For each $t \in \{1, ..., n-1\}$ define $y^t \in B$ to be such that $y_i^t = x_i$ for all $i \leq t$ and $y_i^t = y_i$ for all $i \geq t+1$. Define $y^n = x$. Note that $x = y^n \geq ... \geq y^1 \geq y^0 = y$. It follows that for each $i \in N$

$$T_i(x) - T_i(y) = \sum_{j=1}^{n} \left[T_i\left(y^j\right) - T_i\left(y^{j-1}\right) \right] \geq \sum_{j=1}^{n} \underline{a}_{ij}\varepsilon_{ij}\left(y_j^j - y_j^{j-1}\right) = \sum_{j=1}^{n} \tilde{w}_{ij}(x_j - y_j)$$

$$= \left(\sum_{l=1}^{n} \tilde{w}_{il} \right) \left(\sum_{j=1}^{n} \frac{\tilde{w}_{ij}}{\sum_{l=1}^{n} \tilde{w}_{il}} (x_j - y_j) \right) = \left(\sum_{l=1}^{n} \tilde{w}_{il} \right) \left(\sum_{j=1}^{n} w_{ij}(x_j - y_j) \right) \geq \varepsilon \sum_{j=1}^{n} w_{ij}(x_j - y_j).$$

It follows that

$$x \geq y \implies T(x) - T(y) \geq \varepsilon W(x-y) = \varepsilon(Wx - Wy). \tag{B.25}$$

Define $S : B \to \mathbb{R}^n$ by

$$S(x) = \frac{T(x) - \varepsilon W x}{1-\varepsilon} \qquad \forall x \in B. \tag{B.26}$$

By definition of S and since $W \in \mathcal{W}$ and T is normalized and translation invariant, it is immediate to see that $S(ke) = ke$ for all $k \in I$ and that S is translation invariant. Since (B.25) holds and $\varepsilon \in (0,1)$, routine computations yield that S is monotone. Since S is normalized and monotone, then $S(B) \subseteq B$, that is, S is a selfmap and, in particular, S is a robust opinion aggregator. By rearranging (B.26), (B.23) follows.

(ii) implies (i). Consider $i \in N$. Since W is a stochastic matrix, there exists $j \in N$ such that $w_{ij} > 0$. Let $x \in B$ and $h > 0$ be such that $x + he^j \in B$. By (B.23) and since S is monotone, we have that $T_i(x + he^j) - T_i(x) = \varepsilon w_{ij} h + (1-\varepsilon) S_i(x + he^j) - (1-\varepsilon) S_i(x) \geq \varepsilon w_{ij} h$, proving that j strongly influences i and $\underline{a}_{ij} = 1$. It follows that the i-th row of $\underline{A}(T)$ is not null. Since i was arbitrarily chosen, the statement follows.

Finally, by (B.24), note that W in (ii) can be chosen to be such that $A(W) = \underline{A}(T)$. ■

Point 1 of Theorem 8 builds on two assumptions: i) the matrix of strong ties $\underline{A}(T)$ has no null row; ii) each closed group of $\underline{A}(T)$ is aperiodic. The first assumption allows for a decomposition of T into a convex linear combination of a linear opinion aggregator with matrix W and a robust opinion aggregator S (cf. Proposition 19). We next show that if W takes a very particular form, which we dub partition matrix, then T is asymptotically regular and, in particular, convergent (Lemma 23 and Proposition 20 below). The second assumption yields that W can be always chosen such that W^t eventually "contains" a partition matrix. This will prove point 1 of Theorem 8.

Definition 29. Let $J : B \to B$ be an opinion aggregator. We say that J is a *partition operator/matrix* if and only if there exists a family of disjoint nonempty

subsets $\left\{\hat{N}_l\right\}_{l=1}^{m}$ of N such that $\cup_{l=1}^{m}\hat{N}_l = N$ and for each $l \in \{1, ..., m\}$ there exists $k_l \in \hat{N}_l$ such that $J_i(x) = x_{k_l}$ for all $i \in \hat{N}_l$.

Note that a partition operator is linear. With a small abuse of notation, we will denote the matrix and the operator by the same symbol.

Lemma 23. *Let T be a robust opinion aggregator such that $T = \varepsilon J + (1-\varepsilon) S$ where $\varepsilon \in (0,1)$, J is a partition operator, and $S : B \to B$ is a robust opinion aggregator. Let C be a nonempty subset of B such that there exists $k > 0$ satisfying*

$$\|T(x) - x\|_\infty < k \qquad \forall x \in C. \tag{B.27}$$

If there exists $\delta > 0$ such that for each $t \in \mathbb{N}_0$ there exists $x \in C$ satisfying

$$\left\|T^{t+1}(x) - T^t(x)\right\|_\infty \geq \delta, \tag{B.28}$$

then $\{T^t(x) : x \in C \text{ and } t \in \mathbb{N}_0\}$ is unbounded.

Proposition 20. *Let T be a robust opinion aggregator. If T is such that $T = \varepsilon J + (1-\varepsilon) S$ where $\varepsilon \in (0,1)$, J is a partition operator, and S is a robust opinion aggregator, then T is asymptotically regular and, in particular, convergent.*

Proof. Fix $x \in B$. In Lemma 23, set $C = \{x\}$. Clearly, there exists $k > 0$ that satisfies $\|T(x) - x\|_\infty < k$. By point 2 of Lemma 19 and since T is a robust opinion aggregator, it follows that $\{T^t(x)\}_{t\in\mathbb{N}_0}$ is bounded. By Lemma 23, we have that for each $\delta > 0$ there exists $\bar{t} \in \mathbb{N}_0$ such that

$$\left\|T^{\bar{t}+1}(x) - T^{\bar{t}}(x)\right\|_\infty < \delta. \tag{B.29}$$

Since T is nonexpansive, $\{\|T^{t+1}(x) - T^t(x)\|_\infty\}_{t\in\mathbb{N}_0}$ is a decreasing sequence. By (B.29) and since $\{\|T^{t+1}(x) - T^t(x)\|_\infty\}_{t\in\mathbb{N}_0}$ is a decreasing sequence, we have that for each $\delta > 0$ there exists $\bar{t} \in \mathbb{N}$ such that $\|T^{t+1}(x) - T^t(x)\|_\infty < \delta$ for all $t \geq \bar{t}$, that is, $\lim_t \|T^{t+1}(x) - T^t(x)\|_\infty = 0$. Since x was arbitrarily chosen, it follows that T is asymptotically regular. By Lemma 22, this implies that T is convergent. ■

Lemma 24 below shows that if $\underline{A}(T)$ is aperiodic and nontrivial, then there exists $\bar{t} \in \mathbb{N}$ such that $T^{\bar{t}} = \gamma J + (1-\gamma) S$ (resp. $T^{\bar{t}+1} = \gamma J + (1-\gamma) S$) where J is a partition operator, $\gamma \in (0,1)$, and S is a robust opinion aggregator. The operator J only depends on $\underline{A}(T)$ while γ and S both depend on $\bar{t}$ (resp. $\bar{t}+1$). In turn, Proposition 20 yields that $T^{\bar{t}}$ and $T^{\bar{t}+1}$ are convergent. This will be sufficient to imply the convergence of T.

Lemma 24. *Let T be a robust opinion aggregator. If $\underline{A}(T)$ is aperiodic and nontrivial, then there exists $\bar{t} \in \mathbb{N}$ such that $T^{\bar{t}}$ and $T^{\bar{t}+1}$ are convergent.*

Proof. By Proposition 19 and since $\underline{A}(T)$ is nontrivial, we have that there exists $W \in \mathcal{W}$, $\varepsilon \in (0,1)$, and a robust opinion aggregator $S : B \to B$ such that

$$T(x) = \varepsilon W x + (1-\varepsilon) S(x) \quad \forall x \in B. \tag{B.30}$$

Moreover, W can be chosen to be such that $A(W) = \underline{A}(T)$. By Theorems 2 and 3 of Golub and Jackson (2010) and since $\underline{A}(T)$ is aperiodic, this implies that there exist $\bar{t} \in \mathbb{N}$ and a partition $\left\{\hat{N}_l\right\}_{l=1}^{m}$ of N such that for each $l \in \{1,...,m\}$ there exists $k_l \in \hat{N}_l$ satisfying $w_{ik_l}^{(\bar{t})}, w_{ik_l}^{(\bar{t}+1)} > 0$ for all $i \in \hat{N}_l$.[23] It follows that

$$W^{\bar{t}} = \delta_{\bar{t}} J + (1-\delta_{\bar{t}}) \tilde{W}_{\bar{t}} \text{ and } W^{\bar{t}+1} = \delta_{\bar{t}+1} J + (1-\delta_{\bar{t}+1}) \tilde{W}_{\bar{t}+1} \tag{B.31}$$

where $\delta_{\bar{t}}, \delta_{\bar{t}+1} \in (0,1)$, J is a partition operator/matrix,[24] and $\tilde{W}_{\bar{t}}$ as well as $\tilde{W}_{\bar{t}+1}$ are stochastic matrices. By (B.30) and induction, we also have that $T^{\bar{t}}(x) = \varepsilon^{\bar{t}} W^{\bar{t}} x + \left(1-\varepsilon^{\bar{t}}\right) \tilde{S}_{\bar{t}}(x)$ and $T^{\bar{t}+1}(x) = \varepsilon^{\bar{t}+1} W^{\bar{t}+1} x + \left(1-\varepsilon^{\bar{t}+1}\right) \tilde{S}_{\bar{t}+1}(x)$ for all $x \in B$, where $\tilde{S}_{\bar{t}}$ and $\tilde{S}_{\bar{t}+1}$ are robust opinion aggregators. By (B.31), it follows that $T^{\bar{t}} = \gamma_{\bar{t}} J + (1-\gamma_{\bar{t}}) \hat{S}_{\bar{t}}$ and $T^{\bar{t}+1} = \gamma_{\bar{t}+1} J + (1-\gamma_{\bar{t}+1}) \hat{S}_{\bar{t}+1}$ where $\gamma_{\bar{t}} = \varepsilon^{\bar{t}} \delta_{\bar{t}}$ (resp. $\gamma_{\bar{t}+1} = \varepsilon^{\bar{t}+1} \delta_{\bar{t}+1}$) and $\hat{S}_{\bar{t}}(x) = \frac{\varepsilon^{\bar{t}}(1-\delta_{\bar{t}})}{1-\varepsilon^{\bar{t}}\delta_{\bar{t}}} \tilde{W}_{\bar{t}} x + \frac{1-\varepsilon^{\bar{t}}}{1-\varepsilon^{\bar{t}}\delta_{\bar{t}}} \tilde{S}_{\bar{t}}(x)$ (resp. $\hat{S}_{\bar{t}+1}(x) = \frac{\varepsilon^{\bar{t}+1}\left(1-\delta_{\bar{t}+1}\right)}{1-\varepsilon^{\bar{t}+1}\delta_{\bar{t}+1}} \tilde{W}_{\bar{t}+1} x + \frac{1-\varepsilon^{\bar{t}+1}}{1-\varepsilon^{\bar{t}+1}\delta_{\bar{t}+1}} \tilde{S}_{\bar{t}+1}(x)$) for all $x \in B$. It follows that $\gamma_{\bar{t}}, \gamma_{\bar{t}+1} \in (0,1)$ and $\hat{S}_{\bar{t}}$ as well as $\hat{S}_{\bar{t}+1}$ are robust opinion

[23] As usual, we denote by $w_{ik_l}^{(\bar{t})}$ (resp. $w_{ik_l}^{(\bar{t}+1)}$) the entry in the i-th row and k_l-th column of the matrix $W^{\bar{t}}$ (resp. $W^{\bar{t}+1}$).

[24] That is, $J_i(x) = x_{k_l}$ for all $i \in \hat{N}_l$ and for all $l \in \{1,...,m\}$ where $\left\{\hat{N}_l\right\}_{l=1}^{m}$ and $\{k_l\}_{l=1}^{m}$ have been defined above.

aggregators. By Proposition 20, this implies that $T^{\bar{t}}$ and $T^{\bar{t}+1}$ are convergent. ■

We next present two results which are instrumental to prove point 2 of Theorem 8. To this end, we focus on the network of weak ties $\bar{A}(T)$. Assume that $\left\{C_{[r]}\right\}_{r\in\{0,...,d-1\}}$ is a family of disjoint nonempty subsets of N such that $\cup_{r=0}^{d-1}C_{[r]} = N$ with $d \geq 1$. Given $\left\{x^{[r]}\right\}_{r\in\{0,...,d-1\}} \subseteq B$, we denote by $x = \sum_{r=0}^{d-1} x^{[r]}1_{C_{[r]}} \in B$ the vector whose i-th generic component is such that $x_i = x_i^{[r']}$ when $i \in C_{[r']}$ and $C_{[r']}$ is the only element in $\left\{C_{[r]}\right\}_{r\in\{0,...,d-1\}}$ containing i.

Lemma 25. *Let T be an opinion aggregator and $\left\{C_{[r]}\right\}_{r\in\{0,...,d-1\}}$ a family of disjoint nonempty subsets of N such that $\cup_{r=0}^{d-1}C_{[r]} = N$ with $d \geq 1$. If T is normalized and monotone, then $\bar{A}(T)$ is nontrivial. Moreover, if $\bar{\imath} \in N$ and $\{j \in N : \bar{a}_{\bar{\imath}j} = 1\} \subseteq C_{[r_{\bar{\imath}}]}$ for some $r_{\bar{\imath}} \in \{0, ..., d-1\}$, then*

$$x = \sum_{r=0}^{d-1} x^{[r]}1_{C_{[r]}} \implies T_{\bar{\imath}}(x) = T_{\bar{\imath}}\left(x^{[r_{\bar{\imath}}]}\right). \tag{B.32}$$

Proposition 21. *Let T be a robust opinion aggregator such that N is strongly connected under $\bar{A}(T)$. If T is convergent, then the network of weak ties $\bar{A}(T)$ is aperiodic and nontrivial.*

Proof. By Lemma 25 and since T is normalized and monotone, $\bar{A}(T)$ is nontrivial. By contradiction, assume that $\bar{A}(T)$ is not aperiodic, that is, there exists a closed group M which is not aperiodic under $\bar{A}(T)$. Since N is strongly connected under $\bar{A}(T)$, we have that N is the only closed group, yielding that the greatest common divisor of the lengths of the simple cycles in N is $d \geq 2$. For each $i \in N$ define $\bar{N}_i = \{j \in N : \bar{a}_{ij} = 1\}$. It follows that there exists a partition of N in cyclic classes $\left\{C_{[r]}\right\}_{r\in\{0,...,d-1\}}$ such that $\cup_{i\in C_{[r]}}\bar{N}_i \subseteq C_{[r]\oplus[1]}$ for all $r \in \{0, ..., d-1\}$ where $[r]$ are the elements of $\mathbb{Z}_d$ and $\oplus$ is the standard sum in $\mathbb{Z}_d$. Since I has nonempty interior, there exist $a, b \in I$ such that $a > b$. Define the vector $x \in B$ to be such that $x = \sum_{r=0}^{d-1}\left(k_{[r]}e\right)1_{C_{[r]}}$, where $k_{[0]} = a$ and $k_{[r]} = b$ for all $r \in \{1, ..., d-1\}$. By Lemma

25 and induction and since $\cup_{i\in C_{[r]}}\bar{N}_i \subseteq C_{[r]\oplus[1]}$ for all $r \in \{0,...,d-1\}$, we have that

$$T^t(x) = \sum_{r=0}^{d-1} \left(k_{[r]\oplus t[1]}e\right) 1_{C_{[r]}} \qquad \forall t \in \mathbb{N}.$$

This implies that $\|T^{t+1}(x) - T^t(x)\|_\infty \geq a - b > 0$ for all $t \in \mathbb{N}$, a contradiction with Lemma 22 and T being convergent. ■

Proof of Theorem 8. 1. We adopt the usual convention $T^0(x) = x$ for all $x \in B$. By Lemma 24 and since $\underline{A}(T)$ is aperiodic and nontrivial, there exists $\bar{t} \in \mathbb{N}$ such that $T^{\bar{t}}$ and $T^{\bar{t}+1}$ are convergent. We next show that this implies that T is convergent. Fix $x \in B$. Since $T^{\bar{t}}$ is convergent, we can conclude that $\lim_k T^{k\bar{t}}(x)$ exists. Denote $\bar{x} = \lim_k T^{k\bar{t}}(x)$. Since T is continuous and so is $T^{\bar{t}}$, it is plain that $T^{\bar{t}}(\bar{x}) = \bar{x}$. This implies that

$$T^{\bar{t}}(T^s(\bar{x})) = T^{\bar{t}+s}(\bar{x}) = T^{s+\bar{t}}(\bar{x}) = T^s\left(T^{\bar{t}}(\bar{x})\right) = T^s(\bar{x}) \qquad \forall s \in \mathbb{N}_0.$$

By induction on k, this yields that for each $s \in \mathbb{N}_0$

$$T^{(k+1)\bar{t}}(T^s(\bar{x})) = T^{k\bar{t}}\left(T^{\bar{t}}(T^s(\bar{x}))\right) = T^{k\bar{t}}(T^s(\bar{x})) = T^s(\bar{x}) \quad \forall k \in \mathbb{N}.$$

In particular, by setting $k = s$, we obtain that for each $s \in \mathbb{N}$

$$T^{s(\bar{t}+1)}(\bar{x}) = T^{s\bar{t}}(T^s(\bar{x})) = T^s(\bar{x}). \tag{B.33}$$

Since $T^{\bar{t}+1}$ is convergent, we have that $\lim_s T^{s(\bar{t}+1)}(\bar{x})$ exists. By (B.33), this implies that $\lim_s T^s(\bar{x})$ exists. Denote $\hat{x} = \lim_s T^s(\bar{x})$. Since T is continuous, it is plain that $T(\hat{x}) = \hat{x}$. Since $\left\{T^{k\bar{t}}(\bar{x})\right\}_{k\in\mathbb{N}} \subseteq \{T^s(\bar{x})\}_{s\in\mathbb{N}}$ and $T^{k\bar{t}}(\bar{x}) = \bar{x}$ for all $k \in \mathbb{N}$, we have that

$$\bar{x} = \lim_k T^{k\bar{t}}(\bar{x}) = \lim_s T^s(\bar{x}) = \hat{x} \text{ and } T(\hat{x}) = \hat{x}. \tag{B.34}$$

We can now prove that $\{T^t(x)\}_{t\in\mathbb{N}}$ converges too. By (B.34) and since T is nonex-

pansive, we have that

$$\left\|\bar{x}-T^{t+1}(x)\right\|_{\infty}=\left\|T(\bar{x})-T\left(T^{t}(x)\right)\right\|_{\infty}\leq\left\|\bar{x}-T^{t}(x)\right\|_{\infty}\qquad\forall t\in\mathbb{N},$$

yielding that $\{\|\bar{x}-T^{t}(x)\|_{\infty}\}_{t\in\mathbb{N}}$ is a decreasing sequence. Moreover, since $\bar{x}=\lim_{k}T^{k\bar{t}}(x)$, we have that the subsequence $\left\{\left\|\bar{x}-T^{k\bar{t}}(x)\right\|_{\infty}\right\}_{k\in\mathbb{N}}\subseteq\{\|\bar{x}-T^{t}(x)\|_{\infty}\}_{t\in\mathbb{N}}$ converges to 0. This implies that $\lim_{t}T^{t}(x)=\bar{x}$. Since x was arbitrarily chosen, the statement follows.

2. By Lemma 25 and since T is normalized and monotone, $\bar{A}(T)$ is nontrivial. Next, we consider a family of disjoint subsets $\left\{\hat{N}_{l}\right\}_{l=1}^{m+1}$ of N such that $\cup_{l=1}^{m+1}\hat{N}_{l}=N$ where $m\geq1$ and the first m sets are nonempty. We choose the first m elements of $\left\{\hat{N}_{l}\right\}_{l=1}^{m+1}$ to be the classes (the partition) of essential indexes of $\bar{A}(T)$ and we collect all the possible inessential indexes of $\bar{A}(T)$ in $\hat{N}_{m+1}$. If $l\in\{1,...,m\}$, then $\hat{N}_{l}$ is closed and strongly connected and $\bar{a}_{ij}=0$ for all $i\in\hat{N}_{l}$ and for all $j\in\hat{N}_{l}^{c}$. The set $\hat{N}_{m+1}$ might be empty. If $m=1$ and $\hat{N}_{m+1}=\emptyset$, then N is strongly connected under $\bar{A}(T)$. In this case, by Proposition 21, $\bar{A}(T)$ is aperiodic. Assume that either $m>1$ or $m=1$ and $\hat{N}_{m+1}\neq\emptyset$. By contradiction, assume that $\bar{A}(T)$ is not aperiodic. This implies that there exists a closed group M which is not aperiodic under $\bar{A}(T)$. It is immediate to see that there exists $l\in\{1,...,m\}$ such that $\hat{N}_{l}\subseteq M$. Since $\hat{N}_{l}$ has (simple) cycles and the simple cycles of $\hat{N}_{l}$ are simple cycles of M and M is not aperiodic, the greatest common divisor of the lengths of the cycles of $\hat{N}_{l}$ is greater than the one of the cycles of M and, in particular, ≥2. Set $\hat{N}_{l}=\{i_{1},...,i_{r}\}$. Clearly, $r\geq2$. We introduce two maps $P:\mathbb{R}^{r}\to\mathbb{R}^{n}$ and $\pi:\mathbb{R}^{n}\to\mathbb{R}^{r}$. The first is defined by $x=P(\tilde{x})$ where $x_{i}=\min_{h\in\{1,...,r\}}\tilde{x}_{h}$ if $i\notin\hat{N}_{l}$ and $x_{i_{h}}=\tilde{x}_{h}$ for all $h\in\{1,...,r\}$. The second one is defined by $\tilde{x}=\pi(x)$ where $\tilde{x}_{h}=x_{i_{h}}$ for all $h\in\{1,...,r\}$. It is immediate to check that $P(\pi(z))=z1_{\hat{N}_{l}}+\left(\min_{h\in\{1,...,r\}}z_{i_{h}}e\right)1_{\hat{N}_{l}^{c}}$ for all $z\in\mathbb{R}^{n}$. Note that $P\left(\tilde{B}\right)\subseteq B$ and $\pi(B)\subseteq\tilde{B}$ where $\tilde{B}=I^{r}$. Next, we define $S:\tilde{B}\to\tilde{B}$ by $S(\tilde{x})=\pi(T(P(\tilde{x})))$ for all $\tilde{x}\in\tilde{B}$. It is routine to check that S is a robust opinion aggregator. Moreover, by construction and since $\hat{N}_{l}$ is strongly connected and not aperiodic, we also have that the restricted set of agents $\tilde{N}=\{1,...,r\}$ is strongly

connected and not aperiodic under $\bar{A}(S)$. Note that $S^t(\tilde{x}) = \pi(T^t(P(\tilde{x})))$ for all $\tilde{x} \in \tilde{B}$. Indeed, by Lemma 25 and induction and since $\bar{a}_{ij} = 0$ for all $i \in \hat{N}_l$ and for all $j \in \hat{N}_l^c$, we have that for each $t \in \mathbb{N}$ and for each $\tilde{x} \in \tilde{B}$

$$\begin{aligned} S^{t+1}(\tilde{x}) &= \pi\left(T\left(P\left(\pi\left(T^t(P(\tilde{x}))\right)\right)\right)\right) \\ &= \pi\left(T\left(T^t(P(\tilde{x}))\,1_{\hat{N}_l} + \left(\min_{h\in\{1,\ldots,r\}} T^t_{i_h}(P(\tilde{x}))\,e\right)1_{\hat{N}_l^c}\right)\right) \\ &= \pi\left(T\left(T^t(P(\tilde{x}))\right)\right) = \pi\left(T^{t+1}(P(\tilde{x}))\right). \end{aligned}$$

Since T is convergent and π is continuous, this implies that S is convergent. By Proposition 21 and since S is a convergent robust opinion aggregator such that $\tilde{N}$ is strongly connected under $\bar{A}(S)$, this is a contradiction with $\tilde{N}$ not being aperiodic.■

Proof of Corollary 4. Since T is self-influential, it follows that each row of $\underline{A}(T)$ is not null, yielding that $\underline{A}(T)$ is nontrivial. Moreover, since there is a simple cycle of length 1 from i to i for all $i \in N$, each closed group is aperiodic. By Theorem 8, the statement follows. ■

In order to prove Proposition 15, we begin by making two simple observations about convergence and fixed points of the opinion aggregator T: i) convergence is always toward a fixed point of T; ii) simple properties on the network $\underline{A}(T)$ yield that those fixed points are constant vectors. We denote by $E(T)$ the set of fixed points/equilibria of T. Recall that D is the consensus subset, that is, $x \in D \subseteq B$ if and only if $x_i = x_j$ for all $i, j \in N$.

Proposition 22. *Let T be a robust opinion aggregator. If $\underline{A}(T)$ is nontrivial, has a unique strongly connected and closed group M, and M is aperiodic under $\underline{A}(T)$, then $E(T) = D$.*

Proof of Proposition 15. 1. Since $\underline{A}(T)$ is nontrivial, has a unique strongly connected and closed group M, and M is aperiodic under $\underline{A}(T)$, we have that any other closed group M' is a superset of M, yielding that M' is aperiodic under $\underline{A}(T)$. By Theorem 8 and Proposition 22 and since standard convergence implies Cesaro convergence and T is continuous, it is immediate to see that T is convergent and

$\bar{T}(x) = \lim_t T^t(x) \in E(T) = D$ for all $x \in B$, proving the statement.

2. Consider the same family of disjoint subsets $\left\{\hat{N}_l\right\}_{l=1}^{m+1}$ of N, as in the proof of point 2 of Theorem 8. Recall that if $l \in \{1, ..., m\}$, then $\hat{N}_l$ is closed and strongly connected and $\bar{a}_{ij} = 0$ for all $i \in \hat{N}_l$ and for all $j \in \hat{N}_l^c$. Recall also that $\hat{N}_{m+1}$ might be empty. By Theorem 8 and since T is convergent (to consensus), $\bar{A}(T)$ is aperiodic and nontrivial. By contradiction and since $\bar{A}(T)$ is nontrivial and each closed group is aperiodic under $\bar{A}(T)$, assume that T does not have a unique strongly connected and closed group. Since $\bar{A}(T)$ is nontrivial, this implies that there are at least two distinct strongly connected and closed groups and, in particular, $m \geq 2$. Since I has nonempty interior, consider $a, b \in I$ such that $a > b$. Consider a vector $x \in B$ such that $x_i = a$ for all $i \in \hat{N}_1$, $x_i = b$ for all $i \in \hat{N}_l$ and for all $l \in \{2, ..., m\}$. Since T is convergent, define $\bar{x} = \lim_t T^t(x)$. By Lemma 25 and induction and since $\bar{a}_{ij} = 0$ for all $i \in \hat{N}_l$, for all $j \in \hat{N}_l^c$, and for all $l \in \{1, ..., m\}$, we have that

$$T_i^t(x) = x_i \qquad \forall i \in \hat{N}_l, \forall l \in \{1, ..., m\}, \forall t \in \mathbb{N},$$

proving that $\bar{x}_i = x_i$ for all $i \in \hat{N}_l$ and for all $l \in \{1, ..., m\}$. Since $a \neq b$, we have that $\bar{x}$ is not a constant vector, a contradiction with convergence to consensus. ■

B.9 Appendix: vox populi, vox Dei?

All the missing proofs are in the Supplementary Appendix (see Section B.11.2).

Proof of Theorem 9. Given $n \in \mathbb{N}$, for notational convenience, we define $\hat{B} = \hat{I}^n$. We first make a few observations. Since the random variables $\{X_i(n)\}_{i \in N, n \in \mathbb{N}}$ are uniformly bounded and $\bar{T}_i(n)$ is continuous for all $i \in N$ and for all $n \in \mathbb{N}$, it follows that $\omega \mapsto \bar{T}_i(n)(X_1(n)(\omega), ..., X_n(n)(\omega))$ is integrable for all $i \in N$ and for all $n \in \mathbb{N}$.

Fix $n \in \mathbb{N}$ and $i \in N$. By Rademacher's Theorem and since $\bar{T}(n)$ is nonexpansive, this implies that $\bar{T}(n)$ is almost everywhere differentiable. Let $\mathcal{D}\left(\bar{T}(n)\right) \subseteq \hat{B}$ be the subset of $\hat{B}$ where $\bar{T}(n)$ is differentiable. Clearly, $\bar{T}_i(n)$ is differentiable on $\mathcal{D}\left(\bar{T}(n)\right)$

and, in particular, Clarke differentiable. Since $\bar{T}_i(n)$ is monotone and translation invariant, note that $\nabla \bar{T}_i(n)(x) \in \Delta_n$ for all $x \in \mathcal{D}\left(\bar{T}(n)\right)$. Consider $\bar{x} \in \hat{B}$. Recall that Clarke's differential is the set:

$$\partial \bar{T}_i(n)(\bar{x}) = \operatorname{co}\left\{p \in \Delta_n : p = \lim_k \nabla \bar{T}_i(n)\left(x^k\right) \text{ s.t. } x^k \to \bar{x} \text{ and } x^k \in \mathcal{D}\left(\bar{T}(n)\right)\right\}. \tag{B.35}$$

By Definition 25 and (B.35) and since i and n were arbitrarily chosen, note that

$$0 \leq \underline{s}_{ij}(T(n)) \leq p_j \leq \bar{s}_{ij}(T(n)) \quad \forall i, j \in N, \forall p \in \partial \bar{T}_i(n)(x), \forall x \in \hat{B}, \forall n \in \mathbb{N}. \tag{B.36}$$

1. We start the proof of point 1 with an ancillary claim.

Claim. For each $i, j \in N$ and for each $n \in \mathbb{N}$

$$\sup_{\left\{(x,t) \in \hat{B} \times \mathbb{R} : x + te^j \in \hat{B}\right\}} \left|\bar{T}_i(n)\left(x + te^j\right) - \bar{T}_i(n)(x)\right| \leq \ell \bar{s}_{ij}(T(n)).$$

Proof of the Claim. Fix $i \in N$ and $n \in \mathbb{N}$ and consider $j \in N$, $x \in \hat{B}$, and $t \in \mathbb{R}$ such that $x + te^j \in \hat{B}$. Define $y = x + te^j$. By Lebourg's Mean Value Theorem, we have that there exist $\lambda \in (0,1)$ and $\bar{p} \in \partial \bar{T}_i(n)(z)$ where $z = \lambda y + (1-\lambda)x \in \hat{B}$ such that

$$\bar{T}_i(n)\left(x + te^j\right) - \bar{T}_i(n)(x) = \bar{T}_i(n)(y) - \bar{T}_i(n)(x) = \sum_{l=1}^{n} \bar{p}_l (y_l - x_l).$$

By (B.36), this implies that

$$\left|\bar{T}_i(n)\left(x + te^j\right) - \bar{T}_i(n)(x)\right| = \left|\bar{p}_j (y_j - x_j)\right| = \bar{p}_j |y_j - x_j| \leq \ell \bar{p}_j \leq \ell \bar{s}_{ij}(T(n)).$$

Since x and t were arbitrarily chosen, it follows that

$$\sup_{\left\{(x,t) \in \hat{B} \times \mathbb{R} : x + te^j \in \hat{B}\right\}} \left|\bar{T}_i(n)\left(x + te^j\right) - \bar{T}_i(n)(x)\right| \leq \ell \bar{s}_{ij}(T(n)).$$

Since i, n, and j were also arbitrarily chosen, the statement follows. □

Consider now $n \in \mathbb{N}$ and $i \in N$. By McDiarmid's inequality as well as the previous

claim, we can conclude that for each $\delta > 0$

$$P\left(\left\{\omega \in \Omega : \left|\bar{T}_i(n)(X_1(n)(\omega),...,X_n(n)(\omega)) - \mathbb{E}\left(\bar{T}_i(n)(X_1(n),...,X_n(n))\right)\right|^2 \geq \delta\right\}\right)$$
$$= P\left(\left\{\omega \in \Omega : \left|\bar{T}_i(n)(X_1(n)(\omega),...,X_n(n)(\omega)) - \mathbb{E}\left(\bar{T}_i(n)(X_1(n),...,X_n(n))\right)\right| \geq \sqrt{\delta}\right\}\right)$$

$$\leq 2\exp\left(-\frac{2\delta}{\sum_{j=1}^{n}\left(\ell \bar{s}_{ij}(T(n))\right)^2}\right) = 2\exp\left(-\frac{2\delta}{\ell^2\sum_{j=1}^{n}\bar{s}_{ij}(T(n))^2}\right).$$

Next, since i and n were arbitrarily chosen, observe that

$$\begin{aligned}
&\mathrm{Var}\left(\bar{T}_i(n)(X_1(n),...,X_n(n))\right)\\
&= \mathbb{E}\left(\left(\bar{T}_i(n)(X_1(n),...,X_n(n)) - \mathbb{E}\left(\bar{T}_i(n)(X_1(n),...,X_n(n))\right)\right)^2\right)\\
&= \int_0^{\infty} P\left(\left\{\omega \in \Omega : \left(\bar{T}_i(n)(X_1(n)(\omega),...,X_n(n)(\omega)) - \mathbb{E}\left(\bar{T}_i(n)(X_1(n),...,X_n(n))\right)\right)^2 \geq t\right\}\right) dt\\
&= \int_0^{\ell^2} P\left(\left\{\omega \in \Omega : \left|\bar{T}_i(n)(X_1(n)(\omega),...,X_n(n)(\omega)) - \mathbb{E}\left(\bar{T}_i(n)(X_1(n),...,X_n(n))\right)\right|^2 \geq t\right\}\right) dt\\
&\leq \int_0^{\ell^2} 2\exp\left(-\frac{2t}{\ell^2\sum_{j=1}^{n}\bar{s}_{ij}(T(n))^2}\right) dt\\
&= \ell^2\left(\sum_{j=1}^{n}\bar{s}_{ij}(T(n))^2\right)\left[1 - \exp\left(-\frac{2}{\sum_{j=1}^{n}\bar{s}_{ij}(T(n))^2}\right)\right] \qquad \forall i \in N, \forall n \in \mathbb{N}.
\end{aligned}$$

If we consider $\iota \in \mathbb{N}$ and $n \geq \iota$, this implies that $\mathrm{Var}\left(\bar{T}_\iota(n)(X_1(n),...,X_n(n))\right) \to 0$, proving (B.16).

For the second statement of point 1, assume that $\{\varepsilon_i(n)\}_{i\in N, n\in\mathbb{N}}$ is symmetric and that $\{T(n)\}_{n\in\mathbb{N}}$ is odd. It is enough to show that $\bar{T}_i(n)$ is an unbiased estimator of μ for all $i \in N$ and for all $n \in \mathbb{N}$. By Theorem 7 as well as points 3 and 4 of Lemma 21 and since $I = \mathbb{R}$ and $T(n)$ is an odd robust opinion aggregator for all $n \in \mathbb{N}$, we have that $\bar{T}(n)$ is a well-defined odd robust opinion aggregator for all $n \in \mathbb{N}$. Since $\bar{T}(n)$ is odd for all $n \in \mathbb{N}$ and $\{\varepsilon_i(n)\}_{i\in N, n\in\mathbb{N}}$ is symmetric, this implies that for each $i \in N$ and for each $n \in \mathbb{N}$

$$\int_\Omega \bar{T}_i(n)(\varepsilon_1(n),...,\varepsilon_n(n))\,dP = \int_\Omega \bar{T}_i(n)(-\varepsilon_1(n),...,-\varepsilon_n(n))\,dP = -\int_\Omega \bar{T}_i(n)(\varepsilon_1(n),...,\varepsilon_n(n))\,dP.$$

It follows that $2\int_{\Omega}\bar{T}_i(n)(\varepsilon_1(n),...,\varepsilon_n(n))\,dP=0$ for all $i\in N$ and for all $n\in\mathbb{N}$. Since $\bar{T}(n)$ is translation invariant, we can conclude that for each $i\in N$ and for each $n\in\mathbb{N}$

$$\mathbb{E}\left(\bar{T}_i(n)(X_1(n),...,X_n(n))\right)=\int_{\Omega}\bar{T}_i(n)(X_1(n),...,X_n(n))\,dP$$
$$=\int_{\Omega}\bar{T}_i(n)(\mu+\varepsilon_1(n),...,\mu+\varepsilon_n(n))\,dP=\mu+\int_{\Omega}\bar{T}_i(n)(\varepsilon_1(n),...,\varepsilon_n(n))\,dP=\mu,$$

proving that $\bar{T}_i(n)$ is an unbiased estimator of μ and thus concluding the proof of point 1.

2. Fix $n\in\mathbb{N}$ and $i,j\in N$. Consider $x,y\in\hat{B}$ such that $x\geq y$. By Lebourg's Mean Value Theorem and (B.36), we have that there exist $\lambda\in(0,1)$ and $p\in\partial\bar{T}_i(n)(z)$ where $z=\lambda x+(1-\lambda)y\in\hat{B}$ such that $\bar{T}_i(n)(x)-\bar{T}_i(n)(y)=\sum_{l=1}^{n}p_l(x_l-y_l)\geq p_j(x_j-y_j)\geq\underline{s}_{ij}(T(n))(x_j-y_j)$. Since x and y were arbitrarily chosen, we have that

$$\bar{T}_i(n)(x)-\bar{T}_i(n)(y)\geq\underline{s}_{ij}(T(n))(x_j-y_j)\quad\forall x,y\in\hat{B}\text{ s.t. }x\geq y.\tag{B.37}$$

By definition and since $\bar{T}(n)$ is a robust opinion aggregator, we have that $\underline{s}_{ij}(T(n))\in[0,1]$. If $\underline{s}_{ij}(T(n))<1$, define $R_{ij}(n):\hat{B}\to\mathbb{R}$ by $R_{ij}(n)(x)=\left(\bar{T}_i(n)(x)-\underline{s}_{ij}(T(n))x_j\right)/\left(1-\underline{s}_{ij}(T($ for all $x\in\hat{B}$. By (B.37), it is immediate to see that $R_{ij}(n)$ is monotone and

$$\bar{T}_i(n)(x)=\underline{s}_{ij}(T(n))x_j+\left(1-\underline{s}_{ij}(T(n))\right)R_{ij}(n)(x)\qquad\forall x\in\hat{B}.\tag{B.38}$$

If $\underline{s}_{ij}(T(n))=1$, then $\bar{T}_i(n)(x)=x_j$ for all $x\in\hat{B}$ and we can choose $R_{ij}(n):\hat{B}\to\mathbb{R}$ to be any monotone functional and obtain (B.38). Since n, i, and j were arbitrarily chosen, it follows that (B.38) holds for all $i,j\in N$ and for all $n\in\mathbb{N}$.

By assumption, there exists $\iota\in\mathbb{N}$ such that $\alpha=\limsup_n\max_{j\in N}\underline{s}_{\iota j}(T(n))/2>0$. It follows that there exist a subsequence $\{T(n_m)\}_{m\in\mathbb{N}}$ and a sequence $\{j_m\}_{m\in\mathbb{N}}\subseteq\mathbb{N}$ such that $\underline{s}_{\iota j_m}(T(n_m))\geq\alpha$ and $j_m\leq n_m$ for all $m\in\mathbb{N}$. Fix $m\in\mathbb{N}$. By (B.38) and Harris' inequality and since $\{X_i(n_m)\}_{i\in N}$ is a collection of independent random

variables, we have that

$$\begin{aligned}
&\operatorname{Var}\left(\bar{T}_{\iota}\left(n_m\right)\left(X_1\left(n_m\right),...,X_{n_m}\left(n_m\right)\right)\right)\\
&=\left(1-\underline{s}_{\iota j_m}\left(T\left(n_m\right)\right)\right)^2\operatorname{Var}\left(R_{\iota j_m}\left(n_m\right)\left(X_1\left(n_m\right),...,X_{n_m}\left(n_m\right)\right)\right)+\underline{s}_{\iota j_m}\left(T\left(n_m\right)\right)^2\operatorname{Var}\left(X_{j_m}\left(n_m\right)\right)\\
&+2\left(1-\underline{s}_{\iota j_m}\left(T\left(n_m\right)\right)\right)\underline{s}_{\iota j_m}\left(T\left(n_m\right)\right)\operatorname{Cov}\left(R_{\iota j_m}\left(n_m\right)\left(X_1\left(n_m\right),...,X_{n_m}\left(n_m\right)\right),X_{j_m}\left(n_m\right)\right)\\
&\geq\alpha^2\operatorname{Var}\left(X_{j_m}\left(n_m\right)\right)=\alpha^2\operatorname{Var}\left(\varepsilon_{j_m}\left(n_m\right)\right)\geq\alpha^2\sigma^2>0.
\end{aligned}$$

Since m was arbitrarily chosen, we can conclude that $\{T(n)\}_{n\in\mathbb{N}}$ does not have vanishing variance. Moreover, since $\{X_i(n)\}_{i\in N,n\in\mathbb{N}}$ is an array of uniformly bounded random variables, so is the array $\left\{\bar{T}_i(n)\left(X_1(n),...,X_n(n)\right)\right\}_{i\in N,n\in\mathbb{N}}$. This implies that $\bar{T}_{\iota}(n)\left(X_1(n),...,X_n(n)\right)$ cannot converge in probability to a constant (otherwise, $\{T(n)\}_{n\in\mathbb{N}}$ would have vanishing variance), proving that $\{T(n)\}_{n\in\mathbb{N}}$ is not wise. ■

B.10 Appendix: discussion

All the missing proofs are in the Supplementary Appendix (see Section B.11.3). Given the profile of loss functions $\phi=(\phi_i)_{i=1}^n$, define $\mathbf{T}^{\phi}:B\rightrightarrows B$ as

$$\mathbf{T}^{\phi}(x)=\prod_{i=1}^{n}\operatorname{argmin}_{c\in\mathbb{R}}\phi_i(x-ce)\qquad\forall x\in B. \tag{B.39}$$

The next two ancillary lemmas are instrumental in showing that $\mathbf{T}^{\phi}$ is well defined and behaved.

Lemma 26. *Let ϕ be a profile of loss functions. If $\phi\in\Phi_R$, then for each $i\in N$ and $\tilde{z}\in\mathbb{R}^n$*

$$\tilde{z}\gg 0\implies\phi_i(\tilde{z})>\phi_i\left(\tilde{z}-\min_{j\in N}\tilde{z}_j e\right),$$

and

$$0\gg\tilde{z}\implies\phi_i(\tilde{z})>\phi_i\left(\tilde{z}-\max_{j\in N}\tilde{z}_j e\right).$$

Lemma 27. *Let ϕ be a profile of loss functions. If $\phi \in \Phi_R$, then for each $i \in N$ and for each $x \in \mathbb{R}^n$ the function $f_{i,x} : \mathbb{R} \to \mathbb{R}_+$, defined by $f_{i,x}(c) = \phi_i(x - ce)$ for all $c \in \mathbb{R}$, is continuous and convex. Moreover, if ϕ has strictly increasing shifts, then $f_{i,x}$ is strictly convex for all $i \in N$ and for all $x \in \mathbb{R}^n$.*

To prove (i) implies (ii) of Theorem 10, we prove a more general result, namely, that the solution correspondence (B.39) of problem (B.19), always admits a selection which is a robust opinion aggregator.

Proposition 23. *Let ϕ be a profile of loss functions. If $\phi \in \Phi_R$, then the correspondence $\mathbf{T}^\phi$ is well defined and admits a selection T^ϕ which is a robust opinion aggregator. Moreover, if ϕ has strictly increasing shifts, then $\mathbf{T}^\phi = T^\phi$ is single-valued and, in particular, is a robust opinion aggregator.*

Proof. Fix $i \in N$. We begin by considering the correspondence $\mathbf{T}_i^\phi : B \rightrightarrows I$ defined by $\mathbf{T}_i^\phi(x) = \operatorname{argmin}_{c\in\mathbb{R}} \phi_i(x - ce)$ for all $x \in B$. We next show that $\mathbf{T}_i^\phi$ is well defined, nonempty-, convex-, and compact-valued, and such that for each $x, y \in B$

$$x \geq y \implies \mathbf{T}_i^\phi(x) \geq_{\text{SSO}} \mathbf{T}_i^\phi(y) \tag{B.40}$$

where $\geq_{\text{SSO}}$ is the strong set order. Fix $x \in B$. We next show that

$$\forall d \notin \left[\min_{j\in N} x_j, \max_{j\in N} x_j\right], \exists c \in \left[\min_{j\in N} x_j, \max_{j\in N} x_j\right] \text{ s.t. } \phi_i(x - ce) < \phi_i(x - de). \tag{B.41}$$

Consider d as above. We have two cases either $d < \min_{j\in N} x_j$ or $d > \max_{j\in N} x_j$. In the first case, we have that $x - de \gg 0$, in the second case, we have that $0 \gg x - de$. By Lemma 26 and since $\phi \in \Phi_R$, if we set $\tilde{c} = \min_{j\in N} x_j - d$ (resp. $\max_{j\in N} x_j - d$), we obtain that $\phi_i(x - de) > \phi_i(x - de - \tilde{c}e) = \phi_i(x - ce)$ where $c = \min_{j\in N} x_j \in [\min_{j\in N} x_j, \max_{j\in N} x_j]$ (resp. $c = \max_{j\in N} x_j \in [\min_{j\in N} x_j, \max_{j\in N} x_j]$), proving (B.41). By (B.41), we can conclude that

$$\min_{c\in\mathbb{R}} \phi_i(x - ce) = \min_{c\in I} \phi_i(x - ce) = \min_{c\in\left[\min_{j\in N} x_j, \max_{j\in N} x_j\right]} \phi_i(x - ce) \tag{B.42}$$

as well as $\operatorname{argmin}_{c\in\mathbb{R}} \phi_i(x-ce) = \operatorname{argmin}_{c\in I} \phi_i(x-ce) = \operatorname{argmin}_{c\in\left[\min_{j\in N} x_j, \max_{j\in N} x_j\right]} \phi_i(x-ce)$. By Weierstrass' Theorem and since, by Lemma 27, the map $c \mapsto \phi_i(x-ce)$ is continuous and convex, it follows that the above minimization problems admit solution and each argmin is a compact and convex set. Since x was arbitrarily chosen, this implies that $\mathbf{T}_i^{\phi}$ is well defined, nonempty-, convex-, and compact-valued and, in particular,

$$\emptyset \neq \mathbf{T}_i^{\phi}(x) \subseteq \left[\min_{j\in N} x_j, \max_{j\in N} x_j\right] \subseteq I \quad \forall x \in B. \tag{B.43}$$

We next prove (B.40). In order to do so, we rewrite explicitly (B.42) as a problem of parametric optimization/monotone comparative statics. Next, define $f : I \times B \to \mathbb{R}$ by $f(c,x) = -\phi_i(x-ce)$ for all $(c,x) \in I \times B$. It is immediate to see that $\mathbf{T}_i^{\phi}(x) = \operatorname{argmax}_{c\in I} f(c,x)$ for all $x \in B$. We next show that f has increasing differences in (c,x). Consider $x, y \in B$ as well as $c, d \in I$ such that $c \geq d$ and $x \geq y$. Define $z = x - ce$, $v = y - ce$, and $h = c - d$. Note that $z \geq v$ and $h \in \mathbb{R}_+$. Since $\phi \in \Phi_R$, it follows that

$$\begin{aligned} f(c,x) - f(d,x) &= \phi_i(x-de) - \phi_i(x-ce) = \phi_i(z+he) - \phi_i(z) \\ &\geq \phi_i(v+he) - \phi_i(v) = \phi_i(y-de) - \phi_i(y-ce) = f(c,y) - f(d,y). \end{aligned}$$

This shows that f satisfies the property of increasing differences in (c,x). By Milgrom and Shannon (1994), $\mathbf{T}_i^{\phi}$ satisfies (B.40). We finally show that $\mathbf{T}_i^{\phi}$ is such that for each $x \in B$ and for each $k \in \mathbb{R}$ such that $x + ke \in B$

$$c^{\star} \in \mathbf{T}_i^{\phi}(x) \iff c^{\star} + k \in \mathbf{T}_i^{\phi}(x+ke). \tag{B.44}$$

Fix $x \in B$. Consider $k \in \mathbb{R}$ such that $x + ke \in B$. Consider $c^{\star} \in \mathbf{T}_i^{\phi}(x)$. By definition, it follows that $\phi_i(x - c^{\star}e) \leq \phi_i(x-ce)$ for all $c \in \mathbb{R}$. This implies that $\phi_i(x + ke - (c^{\star}+k)e) = \phi_i(x - c^{\star}e) \leq \phi_i(x - (d-k)e) = \phi_i(x + ke - de)$ for all $d \in \mathbb{R}$. By definition of $\mathbf{T}_i^{\phi}$, this implies that $c^{\star} + k \in \mathbf{T}_i^{\phi}(x+ke)$. Vice versa, if $c^{\star} + k \in \mathbf{T}_i^{\phi}(x+ke)$, then $\phi_i(x + ke - (c^{\star}+k)e) \leq \phi_i(x + ke - de)$ for all $d \in \mathbb{R}$, yielding that $\phi_i(x - c^{\star}e) = \phi_i(x + ke - (c^{\star}+k)e) \leq \phi_i(x-ce)$ for all $c \in \mathbb{R}$,

proving that $c^\star \in \mathbf{T}_i^\phi(x)$.

To sum up, since $i \in N$ was arbitrarily chosen, we proved that, for each $i \in N$, $\mathbf{T}_i^\phi$ is well defined, nonempty-, convex-, and compact-valued, and satisfies (B.40) as well as (B.44). Observe also that $\mathbf{T}^\phi : B \rightrightarrows B$ is the product correspondence $\mathbf{T}^\phi = \prod_{i=1}^n \mathbf{T}_i^\phi$. We are ready to show that $\mathbf{T}^\phi$ admits a selection T^ϕ which is a robust opinion aggregator. Define $T^\phi : B \to B$ to be such that $T_i^\phi(x) = \min \mathbf{T}_i^\phi(x)$ for all $x \in B$, and for all $i \in N$. Since $\mathbf{T}_i^\phi(x)$ is nonempty and compact for all $x \in B$ and for all $i \in N$, it follows that $T_i^\phi(x)$ is well defined and, in particular, $T_i^\phi(x) \in \mathbf{T}_i^\phi(x)$ for all $x \in B$ and for all $i \in N$, proving that T^ϕ is a selection of $\mathbf{T}^\phi$. By (B.43), it follows that $\mathbf{T}_i^\phi(ke) = \{k\}$ for all $k \in I$ and for all $i \in N$, proving that $T_i^\phi(ke) = k$ for all $k \in I$ and for all $i \in N$, that is, that T^ϕ is normalized. Next, consider $x, y \in B$ such that $x \geq y$. By (B.40), we have that $T_i^\phi(x) \geq T_i^\phi(y)$ for all $i \in N$, proving monotonicity of T_i^ϕ for all $i \in N$ and so of T^ϕ. Finally, consider $x \in B$ and $k \in \mathbb{R}$ such that $x + ke \in B$. By (B.44) and definition of $T_i^\phi(x)$ as well as $T_i^\phi(x + ke)$, we have that $T_i^\phi(x) \in \mathbf{T}_i^\phi(x)$ for all $i \in N$, yielding that $T_i^\phi(x) + k \in \mathbf{T}_i^\phi(x + ke)$ for all $i \in N$ and, in particular, $T_i^\phi(x) + k \geq T_i^\phi(x + ke)$ for all $i \in N$. This implies that $T_i^\phi(x + ke) = T_i^\phi(x) + k$ for all $i \in N$, proving translation invariance.[25]

Finally, by Lemma 27, if ϕ has strictly increasing shifts, then the map $c \mapsto \phi_i(x - ce)$ is strictly convex, yielding that each $\mathbf{T}_i^\phi$ is single-valued and so is $\mathbf{T}^\phi$. ■

Proof of Theorem 10. (i) implies (ii). By Proposition 23 and since $\phi \in \Phi_R$ and has strictly increasing shifts, the implication follows.

(ii) implies (i). Let $T : B \to B$ be a robust opinion aggregator. By point 1 of Lemma 19, there exists an extension from $\mathbb{R}^n$ to $\mathbb{R}^n$. With a small abuse of notation, we denote it by the same symbol T. Fix $i \in N$. Define $\phi_i^T : \mathbb{R}^n \to \mathbb{R}_+$ by $\phi_i^T(z) = (T_i(z))^2$ for all $z \in \mathbb{R}^n$. Next, consider $h \in \mathbb{R} \setminus \{0\}$. Since T is normalized, it follows that $\phi_i^T(he) = (T_i(he))^2 = h^2 > 0 = (T_i(0))^2 = \phi_i^T(0)$. Since i and h were

[25] Fix $i \in N$. By the previous part of the proof, for each $x \in B$ and for each $k \in \mathbb{R}$ such that $x + ke \in B$, we have that $T_i^\phi(x + ke) \leq T_i^\phi(x) + k$. Next, note that if $x \in B$ and $x + ke \in B$, then $(x + ke) - ke = x \in B$. It follows that $T_i^\phi(x) = T_i^\phi((x + ke) - ke) \leq T_i^\phi(x + ke) - k$, proving the opposite inequality.

arbitrarily chosen, this implies that $\phi = \left(\phi_i^T\right)_{i=1}^n$ is sensitive. Since T is translation invariant, we have that

$$\phi_i^T(z+he) = \left(T_i(z+he)\right)^2 = \left(T_i(z)+h\right)^2 = \left(T_i(z)\right)^2 + 2hT_i(z) + h^2 \quad \forall h \in \mathbb{R}, \forall z \in \mathbb{R}^n. \tag{B.45}$$

Consider $z, v \in \mathbb{R}^n$ and $h \in \mathbb{R}_{++}$. By (B.45) and since T is monotone, we can conclude that

$$z \geq v \implies \phi_i^T(z+he) - \phi_i^T(z) = 2hT_i(z) + h^2 \geq 2hT_i(v) + h^2 = \phi_i^T(v+he) - \phi_i^T(v).$$

Since i was arbitrarily chosen, it follows that $\phi = \left(\phi_i^T\right)_{i=1}^n$ has increasing shifts and, in particular, $\phi \in \Phi_R$. Next, consider $z, v \in \mathbb{R}^n$ such that $z \gg v$. Set $k = \min_{j \in N}(z_j - v_j)$. It follows that $k > 0$ and $z \geq v + ke$. Since T is monotone and translation invariant and $k > 0$, we can conclude that $T(z) \geq T(v+ke) = T(v) + ke \gg T(v)$. Since $z, v \in \mathbb{R}^n$ were arbitrarily chosen, it follows that $z \gg v \implies T(z) \gg T(v)$. By (B.45), this implies that if $z, v \in \mathbb{R}^n$ and $h \in \mathbb{R}_{++}$, then

$$z \gg v \implies \phi_i^T(z+he) - \phi_i^T(z) = 2hT_i(z) + h^2 > 2hT_i(v) + h^2 = \phi_i^T(v+he) - \phi_i^T(v).$$

Since i was arbitrarily chosen, it follows that $\phi = \left(\phi_i^T\right)_{i=1}^n$ has strictly increasing shifts. We next prove (B.21). By Proposition 23 and since $\phi = \left(\phi_i^T\right)_{i=1}^n \in \Phi_R$ has strictly increasing shifts, we have that $\mathbf{T}_i^\phi(x) = \operatorname{argmin}_{c \in \mathbb{R}} \phi_i^T(x - ce)$ is well defined and single-valued for all $x \in B$ and for all $i \in N$. Finally, fix $i \in N$ and $x \in B$. By (B.45), we have that $\phi_i^T(x-ce) = \left(T_i(x)\right)^2 - 2cT_i(x) + c^2$ for all $c \in \mathbb{R}$, which, as a function of c, is quadratic and minimized at $c = T_i(x)$, proving the statement. ■

Bibliography

[1] D. Acemoglu and P.D. Azar, Endogenous production networks, *Econometrica*, 88, 33-82, 2020.

[2] D. Acemoglu, V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi, The network origins of aggregate fluctuations, *Econometrica*, 80, 1977-2016, 2010.

[3] D. Acemoglu and A. Ozdaglar, Opinion dynamics and learning in social networks, *Dynamic Games and Applications*, 1, 3–49, 2011.

[4] D. Angeli and P. A. Bliman, Extension of a result by Moreau on stability of leaderless multi-agent systems, *Proceedings of the 44th IEEE Conference on Decision and Control*, 759–764, 2005.

[5] I. Arieli, Y. Babichenko, and S. Shlomov, Virtually additive learning, *Journal of Economic Theory*, 197, 105322, 2021.

[6] J. B. Baillon, R. E. Bruck, and S. Reich, On the asymptotic behavior of nonexpansive mappings and semigroups in Banach spaces, *Houston Journal of Mathematics*, 4, 1–9, 1978.

[7] C. Ballester, A. Calvó-Armengol, and Y. Zenou, Who's who in networks. Wanted: The key player, *Econometrica*, 74, 1403–1417, 2006.

[8] A. Banerjee, E. Breza, A. G. Chandrasekhar, and M. Mobius, Naive learning with uninformed agents, *American Economic Review*, 111, 3540–3574, 2021.

[9] A. Banerjee and D. Fudenberg, Word-of-mouth learning, *Games and Economic Behavior*, 46, 1–22, 2004.

[10] L. Beaman, A. BenYishay, J. Magruder, and A. M. Mobarak, Can network theory-based targeting increase technology adoption?, *American Economic Review*, 111, 1918–1943, 2021.

[11] P. Billingsley, *Probability and Measure*, 3rd ed., John Wiley & Sons, New York, 1995.

[12] S. Boucheron, G. Lugosi, and P. Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence*, Oxford University Press, Oxford, 2013.

[13] E. Breza, A. G. Chandrasekhar, B. Golub, and A. Parvathaneni, Networks in economic development, *Oxford Review of Economic Policy*, 35, 678–721, 2019.

[14] E. Breza, A. G. Chandrasekhar, and A. Tahbaz-Salehi, Seeing the forest for the trees? An investigation of network knowledge, *National Bureau of Economic Research*, 24359, 2018.

[15] F. E. Browder and W. V. Petryshyn, The solution by iteration of nonlinear functional equations in Banach spaces, *Bulletin of the American Mathematical Society*, 72, 571–575, 1966.

[16] R. E. Bruck, On the almost-convergence of iterates of a nonexpansive mapping in Hilbert space and the structure of the weak ω-limit set, *Israel Journal of Mathematics*, 29, 1–16, 1978.

[17] A. Calvó-Armengol, J. De Martí, and A. Prat, Communication and influence, *Theoretical Economics,* 10, 649–690, 2015.

[18] D. Centola and M. Macy, Complex contagions and the weakness of long ties, *American Journal of Sociology*, 113, 702–734, 2007.

[19] S. Cerreia-Vioglio, R. Corrao, and G. Lanzani, Robust opinion aggregation and its dynamics, IGIER Working Paper, 662, 2020.

[20] S. Cerreia-Vioglio, R. Corrao, and G. Lanzani, Adaptation, coordination, and inertia in network games, mimeo, 2022.

[21] S. Cerreia-Vioglio, R. Corrao, and G. Lanzani, (Un-)Common Preferences, Ambiguity, and Coordination, mimeo, 2023.

[22] A. G. Chandrasekhar, H. Larreguy, and J. P. Xandri, Testing models of social learning on networks: Evidence from two experiments, *Econometrica*, 88, 1–32, 2020.

[23] S. Chatterjee and E. Seneta, Towards consensus: Some convergence theorems on repeated averaging, *Journal of Applied Probability*, 14, 89–97, 1977.

[24] Y. Chen, J. Lü, and Z. Lin, Consensus of discrete-time multi-agent systems with transmission nonlinearity, *Automatica*, 49, 1768–1775, 2013.

[25] F. H. Clarke, *Optimization and Nonsmooth Analysis*, SIAM, Philadelphia, 1990.

[26] J. Cortés, Distributed algorithms for reaching consensus on general functions, *Automatica,* 44, 726–737, 2008.

[27] K. Dasaratha, B. Golub, and N. Hak, Learning from neighbors about a changing state, *The Review of Economic Studies*, forthcoming.

[28] M. H. DeGroot, Reaching a consensus, *Journal of the American Statistical Association*, 69, 118–121, 1974.

[29] P. M. DeMarzo, D. Vayanos, and J. Zwiebel, Persuasion bias, social influence, and unidimensional opinions, *Quarterly Journal of Economics*, 118, 909–968, 2003.

[30] J. Dieudonne, *Foundations of Modern Analysis*, Academic Press, New York, 1960.

[31] M. Edelstein and R. C. O'Brien, Nonexpansive mappings, asymptotic regularity and successive approximations, *Journal of the London Mathematical Society*, 17, 547–554, 1978.

[32] M. Elliott and B. Golub, A network approach to public goods, *Journal of Political Economy*, 127, 730–776, 2019.

[33] N. E. Friedkin and E. C. Johnsen, Social Influence and Opinions, *The Journal of Mathematical Sociology*, 15, 193–205, 1990.

[34] A. Galeotti, B. Golub, and S. Goyal, Targeting interventions in networks, *Econometrica*, 88, 2445–2471, 2020.

[35] A. Galeotti, S. Goyal, M. O. Jackson, F. Vega-Redondo, and L. Yariv, Network games, *Review of Economic Studies*, 77, 218–244, 2010.

[36] S. Galperti and J. Perego, Information systems, mimeo, 2020.

[37] F. Galton, Vox populi, *Nature*, 75, 450–451, 1907.

[38] P. Ghirardato, F. Maccheroni, and M. Marinacci, Differentiating ambiguity and ambiguity attitude, *Journal of Economic Theory*, 118, 133–173, 2004.

[39] B. Golub and M. O. Jackson, Naïve learning in social networks and the wisdom of crowds, *American Economic Journal: Microeconomics*, 2, 112–149, 2010.

[40] B. Golub and M. O. Jackson, How homophily affects the speed of learning and best-response dynamics, *Quarterly Journal of Economics*, 127, 1287–1338, 2012.

[41] B. Golub and S. Morris, Expectations, networks, and conventions, mimeo, 2018.

[42] B. Golub and E. Sadler, Learning in social networks, in *The Oxford Handbook of the Economics of Networks* (Y. Bramoullé, A. Galeotti, and B. Rogers, eds.), Oxford University Press, New York, 2016.

[43] M. S. Granovetter, The strength of weak ties, *American Journal of Sociology*, 78, 1360–1380, 1973.

[44] P. Holme and M. E. Newman, Nonequilibrium phase transition in the coevolution of networks and opinions, *Physical Review E,* 74, 2006.

[45] P. J. Huber, Robust estimation of a location parameter, *Annals of Mathematical Statistics*, 35, 73–101, 1964.

[46] A. Jadbabaie, P. Molavi, A. Sandroni, and A. Tahbaz-Salehi, Non-Bayesian social learning, *Games and Economic Behavior*, 76, 210–225, 2012.

[47] A. Tahbaz-Salehi, and A. Jadbabaie, A one-parameter family of distributed consensus algorithms with boundary: From shortest paths to mean hitting times, *Proceedings of the 45th IEEE Conference on Decision and Control*, 4664-4669, 2006.

[48] J. G. Kemeny and J. L. Snell, *Finite Markov Chains*, 2nd ed., Springer, New York, 1976.

[49] D. Kempe, J. Kleinberg, and E. Tardos, Maximizing the spread of influence through a social network, *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, 137–146, 2003.

[50] U. Krause, *Positive Dynamical Systems in Discrete Time: Theory, Models, and Applications*, de Gruyter, Berlin, 2015.

[51] U. Krengel, *Ergodic Theorems*, De Gruyter, Berlin, 1986.

[52] G. Levy and R. Razin, Information diffusion in networks with the Bayesian peer influence heuristic, *Games and Economic Behavior*, 109, 262–270, 2018.

[53] F. Maccheroni, M. Marinacci, and A. Rustichini, Ambiguity aversion, robustness, and the variational representation of preferences, *Econometrica*, 74, 1447–1498, 2006.

[54] M. Marinacci and L. Montrucchio, Introduction to the mathematics of ambiguity, *Uncertainty in Economic Theory*, (I. Gilboa, ed.), New York: Routledge, 2004.

[55] P. Milgrom and C. Shannon, Monotone comparative statics, *Econometrica*, 62, 157–180, 1994.

[56] P. Molavi, A. Tahbaz-Salehi, and A. Jadbabaie, A theory of non-Bayesian social learning, *Econometrica*, 86, 445–490, 2018.

[57] L. Moreau, Stability of multiagent systems with time-dependent communication links, *IEEE Transactions on automatic control,* 50, 169–182, 2005.

[58] S. Morris, Contagion, *The Review of Economic Studies*, 67, 57–78, 2000.

[59] E. Mossel, N. Olsman, and O. Tamuz, Efficient Bayesian learning in social networks with Gaussian estimators, *Proceedings of the 54th Annual Allerton Conference on Communication, Control, and Computing*, 425–432, 2016.

[60] M. Mueller-Frank, A general framework for rational learning in social networks, *Theoretical Economics*, 8, 1–40, 2013.

[61] M. Mueller-Frank, Manipulating opinions in social networks, mimeo, 2018.

[62] M. Mueller-Frank and C. Neri, A general analysis of boundedly rational learning in social networks, *Theoretical Economics*, 16, 317–357, 2021.

[63] D. Prelec, The probability weighting function, *Econometrica*, 66, 497–527, 1998.

[64] E. Sadler, Influence campaigns, *American Economic Journal: Microeconomics*, forthcoming.

[65] E. Seneta, *Non-negative Matrices and Markov Chains*, 2nd ed., Springer, New York, 1981.

[66] D. Schmeidler, Subjective probability and expected utility without additivity, *Econometrica*, 57, 571–587, 1989.

[67] H. L. Smith, *Monotone Dynamical Systems: An Introduction to the Theory of Competitive and Cooperative Systems*, American Mathematical Society, 2008.

B.11 Supplementary Appendix

In this section, we confine all the missing proofs. They appear in the order in which the corresponding statements appear in the text, unless they are new ancillary results.

B.11.1 Convergence

Proof of Lemma 19. 1. Since T is robust, we have that $T_i : B \to \mathbb{R}$ is monotone and translation invariant for all $i \in N$.[26] By Cerreia-Vioglio, Maccheroni, Marinacci, and Rustichini (2014), T_i is a niveloid for all $i \in N$. By Cerreia-Vioglio, Maccheroni, Marinacci, and Rustichini (2014), T_i admits an extension $S_i : \mathbb{R}^n \to \mathbb{R}$ which is a niveloid for all $i \in N$. By Cerreia-Vioglio, Maccheroni, Marinacci, and Rustichini (2014), S_i is monotone and translation invariant for all $i \in N$. Define $S : \mathbb{R}^n \to \mathbb{R}^n$ to be such that the i-th component of $S(x)$ is $S_i(x)$ for all $i \in N$ and for all $x \in \mathbb{R}^n$. It is immediate to see that S is monotone and translation invariant. Fix $k' \in I$. Since S is translation invariant and T is normalized, it follows that for each $k \in \mathbb{R}$

$$S(ke) = S(k'e + (k - k')e) = S(k'e) + (k - k')e = T(k'e) + (k - k')e = k'e + (k - k')e = ke,$$

proving that S is normalized and, in particular, that S is robust.

2. By induction, if T is normalized and monotone, then T^t is normalized and monotone for all $t \in \mathbb{N}$. Consider $x \in B$ and $t \in \mathbb{N}$. Define $k_\star = \min_{i \in N} x_i$ and $k^\star = \max_{i \in N} x_i$. Note that $\|x\|_\infty = \max\{|k_\star|, |k^\star|\}$, $k_\star, k^\star \in I$, and $k_\star e \leq x \leq k^\star e$. Since T^t is normalized and monotone, we have that

$$k_\star e = T^t(k_\star e) \leq T^t(x) \leq T^t(k^\star e) = k^\star e,$$

yielding that $|T^t(x)| \leq \max\{|k_\star|, |k^\star|\} e$ and $\|T^t(x)\|_\infty \leq \|x\|_\infty$. Since t and x were arbitrarily chosen, the statement follows. ■

Proof of Lemma 20. Since T is a robust opinion aggregator, T_i is normalized, monotone, and translation invariant for all $i \in N$. By Cerreia-Vioglio, Maccheroni, Marinacci, and Rustichini (2014), it follows that T_i is a niveloid for all $i \in N$. By Cerreia-Vioglio, Maccheroni, Marinacci, and Rustichini (2014), it follows that $|T_i(x) - T_i(y)| \leq \|x - y\|_\infty$ for all $x, y \in B$ and for all $i \in N$. This implies that

$$\|T(x) - T(y)\|_\infty = \max_{i \in N} |T_i(x) - T_i(y)| \leq \|x - y\|_\infty \qquad \forall x, y \in B,$$

proving that T is nonexpansive.

[26] With a small abuse of terminology, we use the same name for similar properties that pertain to functionals and operators.

By induction, we next show that T^t is nonexpansive for all $t \in \mathbb{N}$. Since we have shown that T is nonexpansive, T^t is nonexpansive for $t = 1$, proving the initial step. By the induction hypothesis, assume that T^t is nonexpansive, we have that for each $x, y \in B$

$$\left\|T^{t+1}(x) - T^{t+1}(y)\right\|_\infty = \left\|T\left(T^t(x)\right) - T\left(T^t(y)\right)\right\|_\infty \leq \left\|T^t(x) - T^t(y)\right\|_\infty \leq \|x - y\|,$$

proving the inductive step. The statement follows by induction. ■

Proof of Lemma 21. Let $x \in B$. Since T is a selfmap, we have that $\{T^t(x)\}_{t\in\mathbb{N}} \subseteq B$. Since B is convex, we have that $\frac{1}{\tau}\sum_{t=1}^{\tau} T^t(x) \in B$ for all $\tau \in \mathbb{N}$. Since x was arbitrarily chosen, this implies that $A_\tau : B \to B$, defined by $A_\tau(x) = \sum_{t=1}^{\tau} T^t(x)/\tau$ for all $x \in B$, is well defined for all $\tau \in \mathbb{N}$. Since B is closed, we have that $\bar{T}(x) = \lim_\tau A_\tau(x) = \lim_\tau \frac{1}{\tau}\sum_{t=1}^{\tau} T^t(x) \in B$ for all $x \in B$, proving that $\bar{T}$ is well defined. So one has

$$A_\tau(T(x)) = \frac{\tau+1}{\tau} A_{\tau+1}(x) - \frac{1}{\tau} T(x) \qquad \forall x \in B, \forall \tau \in \mathbb{N}.$$

This implies that

$$\bar{T}(T(x)) = \lim_\tau A_\tau(T(x)) = \lim_\tau \frac{\tau+1}{\tau} \lim_\tau A_{\tau+1}(x) - \lim_\tau \frac{1}{\tau} T(x) = \bar{T}(x) \qquad \forall x \in B,$$

proving that $\bar{T} \circ T = \bar{T}$.

1. By the same inductive argument contained in the proof of Lemma 20, we have that for each $t \in \mathbb{N}$ the map $T^t : B \to B$ is nonexpansive. Since the convex linear combination of nonexpansive maps is nonexpansive, the map $A_\tau : B \to B$ is nonexpansive for all $\tau \in \mathbb{N}$. We can conclude that for each $x, y \in B$

$$\left\|\bar{T}(x) - \bar{T}(y)\right\|_\infty = \left\|\lim_\tau A_\tau(x) - \lim_\tau A_\tau(y)\right\|_\infty = \lim_\tau \|A_\tau(x) - A_\tau(y)\|_\infty \leq \|x - y\|_\infty,$$

proving that $\bar{T}$ is nonexpansive. Continuity of $\bar{T}$ trivially follows.

2. By induction, we have that for each $t \in \mathbb{N}$ the map $T^t : B \to B$ is normalized and monotone. Since the convex linear combination of normalized and monotone operators is normalized and monotone, the map $A_\tau : B \to B$ is normalized and monotone for all $\tau \in \mathbb{N}$. We can conclude that $\bar{T}(ke) = \lim_\tau A_\tau(ke) = ke$ for all $k \in I$ as well as

$$x \geq y \implies \bar{T}(x) = \lim_\tau A_\tau(x) \geq \lim_\tau A_\tau(y) = \bar{T}(y),$$

proving that $\bar{T}$ is normalized and monotone.

3. Since T is robust, T is normalized, monotone, and translation invariant. By the previous point, $\bar{T}$ is normalized and monotone. By induction, we have that for each $t \in \mathbb{N}$ the map $T^t : B \to B$ is translation invariant. Since the convex linear combination of translation invariant operators is translation invariant, the map $A_\tau : B \to B$ is translation invariant for all $\tau \in \mathbb{N}$. We can conclude

that for each $x \in B$ and for each $k \in \mathbb{R}$ such that $x + ke \in B$

$$\bar{T}(x + ke) = \lim_{\tau} A_\tau(x + ke) = \lim_{\tau} [A_\tau(x) + ke] = \bar{T}(x) + ke,$$

proving that $\bar{T}$ is translation invariant and, in particular, robust.

4. By induction, we have that for each $t \in \mathbb{N}$ the map $T^t : B \to B$ is odd. Since the convex linear combination of odd maps is odd, the map $A_\tau : B \to B$ is odd for all $\tau \in \mathbb{N}$. We can conclude that

$$\bar{T}(-x) = \lim_{\tau} A_\tau(-x) = \lim_{\tau} [-A_\tau(x)] = -\bar{T}(x) \quad \forall x \in B,$$

proving that $\bar{T}$ is odd. ■

In order to prove Lemma 22, we are going to rely upon Lorentz's Theorem.

Theorem 11 (Lorentz)**.** *Let $\{x^t\}_{t\in\mathbb{N}} \subseteq \mathbb{R}^n$ be a bounded sequence. The following statements are equivalent:*

(i) There exists $\bar{x} \in \mathbb{R}^n$ such that

$$\forall \varepsilon > 0 \ \exists \bar{\tau} \in \mathbb{N} \ \forall m \in \mathbb{N} \ s.t. \ \left\| \frac{1}{\tau} \sum_{t=1}^{\tau} x^{m+t} - \bar{x} \right\|_\infty < \varepsilon \qquad \forall \tau \geq \bar{\tau}$$

and $\lim_t \left\| x^{t+1} - x^t \right\|_\infty = 0$;

(ii) $\lim_t x^t = \bar{x}$.

Proof of Lemma 22. By Theorem 7 and since T is robust, we have that if $\hat{B}$ is a bounded subset of B, then

$$\lim_{\tau} \left(\sup_{x \in \hat{B}} \left\| \frac{1}{\tau} \sum_{t=1}^{\tau} T^t(x) - \bar{T}(x) \right\|_\infty \right) = 0 \tag{B.46}$$

where $\bar{T} : B \to B$ is a robust opinion aggregator such that $\bar{T} \circ T = \bar{T}$. Since $\bar{T}(T(x)) = \bar{T}(x)$ for all $x \in B$, by induction, we have that $\bar{T}(T^m(x)) = \bar{T}(x)$ for all $m \in \mathbb{N}$ and for all $x \in B$.

(i) implies (ii). Fix $x \in B$. Define the sequence $x^t = T^t(x)$ for all $t \in \mathbb{N}$. By point 2 of Lemma 19, we have that $\{x^t\}_{t\in\mathbb{N}}$ is bounded. Set $\hat{B} = \{x^t\}_{t\in\mathbb{N}}$. Note that for each $\tau \in \mathbb{N}$ and for each $m \in \mathbb{N}$

$$\frac{1}{\tau} \sum_{t=1}^{\tau} x^{m+t} = \frac{1}{\tau} \sum_{t=1}^{\tau} T^{m+t}(x) = \frac{1}{\tau} \sum_{t=1}^{\tau} T^t(T^m(x)).$$

Since (B.46) holds, if we define $\bar{x} = \bar{T}(x)$, then we have that for each $m \in \mathbb{N}$

$$\lim_{\tau} \frac{1}{\tau} \sum_{t=1}^{\tau} x^{m+t} = \lim_{\tau} \frac{1}{\tau} \sum_{t=1}^{\tau} T^t(T^m(x)) = \bar{T}(T^m(x)) = \bar{T}(x) = \bar{x}.$$

It follows that

$$\sup_{m\in\mathbb{N}}\left\|\frac{1}{\tau}\sum_{t=1}^{\tau}x^{m+t}-\bar{x}\right\|_{\infty}=\sup_{m\in\mathbb{N}}\left\|\frac{1}{\tau}\sum_{t=1}^{\tau}T^{t}\left(T^{m}\left(x\right)\right)-\bar{T}\left(T^{m}\left(x\right)\right)\right\|_{\infty}\leq\sup_{x\in\hat{B}}\left\|\frac{1}{\tau}\sum_{t=1}^{\tau}T^{t}\left(x\right)-\bar{T}\left(x\right)\right\|_{\infty}.$$

Since (B.46) holds and T is asymptotically regular, we have that $\{x^t\}_{t\in\mathbb{N}}$ satisfies point (i) of Theorem 11. By Theorem 11, we have that $\lim_t T^t(x) = \lim_t x^t$ exists. Since x was arbitrarily chosen, the implication follows.

(ii) implies (i). Fix $x \in B$. Define $x^t = T^t(x)$ for all $t \in \mathbb{N}$. Since T is convergent, we have that $\{x^t\}_{t\in\mathbb{N}}$ converges and, in particular, is bounded. By Theorem 11, we have that $\lim_t \left\|T^{t+1}(x) - T^t(x)\right\|_\infty = \lim_t \left\|x^{t+1} - x^t\right\|_\infty = 0$. Since x was arbitrarily chosen, the implication follows. ■

Proof of Lemma 23. We first offer two definitions and make two observations. Define the diameter of $\{T^t(x) : x \in C \text{ and } t \in \mathbb{N}_0\}$ by $\bar{D}$.[27] Given $x \in B$, define $x^t = T^t(x)$ as well as $y^t = S(x^t)$ for all $t \in \mathbb{N}_0$. Since T is nonexpansive, recall that $\left\{\left\|x^t - x^{t-1}\right\|_\infty\right\}_{t\in\mathbb{N}}$ is a decreasing sequence for all $x \in B$. Note that this implies that $\|T(x) - x\|_\infty \geq \left\|T^{t+1}(x) - T^t(x)\right\|_\infty$ for all $t \in \mathbb{N}_0$ and for all $x \in B$, yielding that $k > \delta$.

By contradiction, assume that $\{T^t(x) : x \in C \text{ and } t \in \mathbb{N}_0\}$ is bounded. This implies that $\bar{D} < \infty$. Consider $M \in \mathbb{N}\backslash\{1\}$ and $P \in \mathbb{N}$ to be such that $M\delta > \bar{D} + \delta + 1$ and $\lfloor \frac{P}{M} \rfloor > \max\left\{1, \frac{k}{(1-\varepsilon)\varepsilon^M}\right\}$. By (B.28) and since $P \in \mathbb{N}$, there exists $x \in C$ such that $\left\|x^{P+1} - x^P\right\|_\infty = \left\|T^{P+1}(x) - T^P(x)\right\|_\infty \geq \delta$. Now, we list seven useful facts:

1. By (B.27) and since $\left\{\left\|x^t - x^{t-1}\right\|_\infty\right\}_{t\in\mathbb{N}}$ is a decreasing sequence, it follows that $k \geq \left\|x^{i+1} - x^i\right\|_\infty \geq \delta$ for all $i \in \{1, ..., P\}$.

2. By definition of $\{y^t\}_{t\in\mathbb{N}_0}$ and since S is nonexpansive, we have that $\left\|y^t - y^{t-1}\right\|_\infty \leq \left\|x^t - x^{t-1}\right\|_\infty$ for all $t \in \mathbb{N}$.

3. By definition of $\{x^t\}_{t\in\mathbb{N}_0}$ and since $T = \varepsilon J + (1-\varepsilon)S$, we have that $x^t = T\left(x^{t-1}\right) = \varepsilon J\left(x^{t-1}\right) + (1-\varepsilon)y^{t-1}$ for all $t \in \mathbb{N}$, that is,

$$y^{t-1} = \frac{1}{1-\varepsilon}x^t - \frac{\varepsilon}{1-\varepsilon}J\left(x^{t-1}\right) \qquad \forall t \in \mathbb{N}.$$

   By point 2, this yields that $\left\|\frac{1}{1-\varepsilon}\left(x^{t+1} - x^t\right) - \frac{\varepsilon}{1-\varepsilon}\left(J(x^t) - J\left(x^{t-1}\right)\right)\right\|_\infty = \left\|y^t - y^{t-1}\right\|_\infty \leq \left\|x^t - x^{t-1}\right\|_\infty$ for all $t \in \mathbb{N}$.

4. Let L be an integer in $\mathbb{N}$ such that

$$L > \frac{k}{(1-\varepsilon)\varepsilon^M}. \tag{B.47}$$

[27] Recall that the diameter of a subset $\hat{A}$ of B is the quantity $\sup\left\{\|x - y\|_\infty : x, y \in \hat{A}\right\}$.

Define $b_m = \delta + m\,(1-\varepsilon)\,\varepsilon^M$ for all $m \in \{0, ..., L\}$. It follows that the collection of intervals $\{[b_m, b_{m+1}]\}_{m=0}^{L-1}$ contains L elements whose union is a superset of $[\delta, k]$.

5. Note that $\varepsilon^{M-1}\frac{1-\varepsilon^i}{\varepsilon^i} = \varepsilon^{M-i-1} - \varepsilon^{M-1} \leq \varepsilon^{M-i-1}$ for all $i \in \{1, ..., M-1\}$. Since $\varepsilon \in (0,1)$, this implies that

$$(1-\varepsilon)\,\varepsilon^M \sum_{i=1}^{M-1} \frac{1-\varepsilon^i}{\varepsilon^i} \leq (1-\varepsilon)\,\varepsilon \sum_{i=1}^{M-1} \varepsilon^{M-i-1} = (1-\varepsilon)\,\varepsilon \sum_{i=0}^{M-2} \varepsilon^i \leq (1-\varepsilon)\,\varepsilon \frac{1}{1-\varepsilon} \leq \varepsilon < 1.$$

6. Let $t \in \mathbb{N}$, $j \in N$, and $b, \kappa, c \geq 0$. If $x_j^{t+1} - x_j^t \geq b - c$ and $\left\|x^t - x^{t-1}\right\|_\infty \leq b + \kappa$, then (by point 3): $\frac{b-c}{1-\varepsilon} - \frac{\varepsilon}{1-\varepsilon}\left(x_{k_l}^t - x_{k_l}^{t-1}\right) = \frac{b-c}{1-\varepsilon} - \frac{\varepsilon}{1-\varepsilon}\left(J_j\left(x^t\right) - J_j\left(x^{t-1}\right)\right) \leq b + \kappa$ where l is such that $j \in \hat{N}_l$. This yields that

$$x_{k_l}^t - x_{k_l}^{t-1} \geq b - \frac{c}{\varepsilon} - \frac{1-\varepsilon}{\varepsilon}\kappa. \tag{B.48}$$

7. Let $t \in \mathbb{N}$, $j \in N$, and $b, \kappa, c \geq 0$. If $x_j^t - x_j^{t+1} \geq b - c$ and $\left\|x^t - x^{t-1}\right\|_\infty \leq b + \kappa$, then (by point 3): $\frac{b-c}{1-\varepsilon} - \frac{\varepsilon}{1-\varepsilon}\left(x_{k_l}^{t-1} - x_{k_l}^t\right) = \frac{b-c}{1-\varepsilon} - \frac{\varepsilon}{1-\varepsilon}\left(J_j\left(x^{t-1}\right) - J_j\left(x^t\right)\right) \leq b + \kappa$ where l is such that $j \in \hat{N}_l$. This yields that

$$x_{k_l}^{t-1} - x_{k_l}^t \geq b - \frac{c}{\varepsilon} - \frac{1-\varepsilon}{\varepsilon}\kappa. \tag{B.49}$$

By definition of P, we have that $\lfloor P/M \rfloor$ satisfies (B.47). By point 4, there exists a collection of intervals $\{[b_m, b_{m+1}]\}_{m=0}^{\lfloor P/M \rfloor - 1}$ which covers $[\delta, k]$. By point 1, $[\delta, k]$ contains $\left\{\left\|x^{i+1} - x^i\right\|_\infty\right\}_{i=1}^P$. Since we have $\lfloor P/M \rfloor$ intervals and the first P elements (of the sequence $\left\{\left\|x^{t+1} - x^t\right\|_\infty\right\}_{t\in\mathbb{N}}$) belong to these intervals, we have that there exists one of them, $\hat{I} = [b_{\bar{m}}, b_{\bar{m}+1}]$, which contains at least M elements of $\left\{\left\|x^{i+1} - x^i\right\|_\infty\right\}_{i=1}^P$. Since $\left\{\left\|x^t - x^{t-1}\right\|_\infty\right\}_{t\in\mathbb{N}}$ is decreasing, we have that there exists $K \in \mathbb{N}_0$ such that $\left\|x^{K+i+1} - x^{K+i}\right\|_\infty \in \hat{I}$ for all $i \in \{1, ..., M\}$. This implies that there exists $j \in \{1, ..., n\}$ such that $\left|x_j^{K+M+1} - x_j^{K+M}\right| \geq b_{\bar{m}}$ and $||x^{K+M} - x^{K+M-1}||_\infty \leq b_{\bar{m}+1} = b_{\bar{m}} + (1-\varepsilon)\,\varepsilon^M$. We have two cases:

a. $x_j^{K+M+1} - x_j^{K+M} \geq b_{\bar{m}}$. Set $b = b_{\bar{m}}$, $c = 0$, and $\kappa = (1-\varepsilon)\,\varepsilon^M$. By (B.48), we can conclude that

$$x_{k_l}^{K+M} - x_{k_l}^{K+M-1} \geq b_{\bar{m}} - (1-\varepsilon)\,\varepsilon^M \frac{(1-\varepsilon)}{\varepsilon}. \tag{B.50}$$

By (finite) induction, we next prove that

$$x_{k_l}^{K+M+1-i} - x_{k_l}^{K+M-i} \geq b_{\bar{m}} - (1-\varepsilon)\,\varepsilon^M \frac{\left(1-\varepsilon^i\right)}{\varepsilon^i} \qquad \forall i \in \{1, ..., M-1\}. \tag{B.51}$$

By (B.50), the statement is true for $i = 1$. Next, we assume it is true for $i \in \{1, ..., M-1\}$ and prove it is still true for $i+1$ when $i+1 \in \{1, ..., M-1\}$. This implies that $i \leq M-2$.

Define $t = K + M - i$. By the induction hypothesis, we have that

$$x_{k_l}^{t+1} - x_{k_l}^{t} = x_{k_l}^{K+M+1-i} - x_{k_l}^{K+M-i} \geq b_{\bar{m}} - (1-\varepsilon)\,\varepsilon^M \frac{\left(1-\varepsilon^i\right)}{\varepsilon^i}.$$

Moreover, we also have that $\left\|x^t - x^{t-1}\right\|_\infty = \left\|x^{K+M-i} - x^{K+M-i-1}\right\|_\infty \leq b_{\bar{m}} + (1-\varepsilon)\,\varepsilon^M$. Set $b = b_{\bar{m}}$, $c = (1-\varepsilon)\,\varepsilon^M \frac{\left(1-\varepsilon^i\right)}{\varepsilon^i}$, and $\kappa = (1-\varepsilon)\,\varepsilon^M$. By (B.48), we can conclude that

$$\begin{aligned} x_{k_l}^{K+M+1-(i+1)} - x_{k_l}^{K+M-(i+1)} &= x_{k_l}^{K+M-i} - x_{k_l}^{K+M-i-1} = x_{k_l}^{t} - x_{k_l}^{t-1} \\ &\geq b_{\bar{m}} - (1-\varepsilon)\,\varepsilon^M \frac{\left(1-\varepsilon^i\right)}{\varepsilon^i}\frac{1}{\varepsilon} - \frac{1-\varepsilon}{\varepsilon}(1-\varepsilon)\,\varepsilon^M \\ &= b_{\bar{m}} - (1-\varepsilon)\,\varepsilon^M \frac{\left(1-\varepsilon^{i+1}\right)}{\varepsilon^{i+1}}, \end{aligned}$$

proving (B.51). By (B.51) and summation as well as point 5, this implies that

$$x_{k_l}^{K+M} - x_{k_l}^{K+1} \geq (M-1)\,b_{\bar{m}} - (1-\varepsilon)\,\varepsilon^M \sum_{i=1}^{M-1} \frac{1-\varepsilon^i}{\varepsilon^i} \geq (M-1)\,b_{\bar{m}} - 1,$$

that is, $\left\|x^{K+M} - x^{K+1}\right\|_\infty \geq x_{k_l}^{K+M} - x_{k_l}^{K+1} \geq (M-1)\,b_{\bar{m}} - 1$. Since $b_{\bar{m}} \geq \delta > 0$, we have that $(M-1)\,b_{\bar{m}} \geq (M-1)\,\delta > \bar{D} + 1$. We can conclude that $\bar{D} \geq \left\|x^{K+M} - x^{K+1}\right\|_\infty \geq (M-1)\,b_{\bar{m}} - 1 > \bar{D}$, a contradiction.

b. $x_j^{K+M} - x_j^{K+M+1} \geq b_{\bar{m}}$. Set $b = b_{\bar{m}}$, $c = 0$, and $\kappa = (1-\varepsilon)\,\varepsilon^M$. By (B.49), we can conclude that

$$x_{k_l}^{K+M-1} - x_{k_l}^{K+M} \geq b_{\bar{m}} - (1-\varepsilon)\,\varepsilon^M \frac{1-\varepsilon}{\varepsilon}. \tag{B.52}$$

By (finite) induction, we next prove that

$$x_{k_l}^{K+M-i} - x_{k_l}^{K+M+1-i} \geq b_{\bar{m}} - (1-\varepsilon)\,\varepsilon^M \frac{\left(1-\varepsilon^i\right)}{\varepsilon^i} \qquad \forall i \in \{1, ..., M-1\}. \tag{B.53}$$

By (B.52), the statement is true for $i = 1$. Next, we assume it is true for $i \in \{1, ..., M-1\}$ and prove it is still true for $i+1$ when $i+1 \in \{1, ..., M-1\}$. This implies that $i \leq M-2$. Define $t = K + M - i$. By the induction hypothesis, we have that

$$x_{k_l}^{t} - x_{k_l}^{t+1} = x_{k_l}^{K+M-i} - x_{k_l}^{K+M+1-i} \geq b_{\bar{m}} - (1-\varepsilon)\,\varepsilon^M \frac{\left(1-\varepsilon^i\right)}{\varepsilon^i}.$$

Moreover, we also have that $\left\|x^t - x^{t-1}\right\|_\infty = \left\|x^{K+M-i} - x^{K+M-i-1}\right\|_\infty \leq b_{\bar{m}} + (1-\varepsilon)\,\varepsilon^M$.

Set $b = b_{\bar{m}}$, $c = (1-\varepsilon)\,\varepsilon^M \frac{(1-\varepsilon^i)}{\varepsilon^i}$, and $\kappa = (1-\varepsilon)\,\varepsilon^M$. By (B.49), we can conclude that

$$\begin{aligned} x_{k_l}^{K+M-(i+1)} - x_{k_l}^{K+M+1-(i+1)} &= x_{k_l}^{K+M-i-1} - x_{k_l}^{K+M-i} = x_{k_l}^{t-1} - x_{k_l}^{t} \\ &\geq b_{\bar{m}} - (1-\varepsilon)\,\varepsilon^M \frac{\left(1-\varepsilon^i\right)}{\varepsilon^i}\frac{1}{\varepsilon} - \frac{1-\varepsilon}{\varepsilon}(1-\varepsilon)\,\varepsilon^M \\ &= b_{\bar{m}} - (1-\varepsilon)\,\varepsilon^M \frac{\left(1-\varepsilon^{i+1}\right)}{\varepsilon^{i+1}}, \end{aligned}$$

proving (B.53). By (B.53) and summation as well as point 5, this implies that

$$x_{k_l}^{K+1} - x_{k_l}^{K+M} \geq (M-1)\,b_{\bar{m}} - (1-\varepsilon)\,\varepsilon^M \sum_{i=1}^{M-1} \frac{1-\varepsilon^i}{\varepsilon^i} \geq (M-1)\,b_{\bar{m}} - 1,$$

that is, $\left\|x^{K+1} - x^{K+M}\right\|_\infty \geq x_{k_l}^{K+1} - x_{k_l}^{K+M} \geq (M-1)\,b_{\bar{m}} - 1$. Since $b_{\bar{m}} \geq \delta > 0$, we have that $(M-1)\,b_{\bar{m}} \geq (M-1)\,\delta > \bar{D} + 1$. We can conclude that $\bar{D} \geq \left\|x^{K+1} - x^{K+M}\right\|_\infty \geq (M-1)\,b_{\bar{m}} - 1 > \bar{D}$, a contradiction.

Points a and b prove the statement. ■

Proof of Lemma 25. Consider generic $x, y \in B$ and $l \in N$. Define $y^0 = y$. For each $t \in \{1, ..., n-1\}$ define $y^t \in B$ to be such that $y_i^t = x_i$ for all $i \leq t$ and $y_i^t = y_i$ for all $i \geq t+1$. Define $y^n = x$. Note that $y^j - y^{j-1} = (x_j - y_j)\,e^j$ for all $j \in \{1, ..., n\}$. We also have that

$$T_l(x) - T_l(y) = T_l(y^n) - T_l\left(y^0\right) = \sum_{j=1}^{n} \left[T_l\left(y^j\right) - T_l\left(y^{j-1}\right)\right]. \tag{B.54}$$

Since I has nonempty interior, we have that there exist $a, b \in I$ such that $a > b$. By contradiction, assume that $\bar{A}(T)$ is not nontrivial, that is, there exists $i \in N$ such that $\bar{a}_{ij} = 0$ for all $j \in N$, yielding that $T_i\left(z + he^j\right) = T_i(z)$ for all $h \in \mathbb{R}$ and for all $z \in B$ such that $z + he^j \in B$. Set $x = ae$ and $y = be$. By (B.54) and since T is normalized, it follows that $0 < a - b = T_i(ae) - T_i(be) = 0$, a contradiction, proving the first part of the statement. Next, consider $\bar{\imath} \in N$ and define $\bar{N}_{\bar{\imath}} = \{j \in N : \bar{a}_{\bar{\imath}j} = 1\}$. By assumption, we have that $\bar{N}_{\bar{\imath}} \subseteq C_{[r_{\bar{\imath}}]}$. Let x be as in (B.32) and $y = x^{[r_{\bar{\imath}}]}$. By definition of $\bar{A}(T)$, it is immediate to see that $\bar{a}_{\bar{\imath}j} = 0$ only if $T_{\bar{\imath}}\left(z + he^j\right) = T_{\bar{\imath}}(z)$ for all $h \in \mathbb{R}$ and for all $z \in B$ such that $z + he^j \in B$. Consider $j \in \{1, ..., n\}$. We have two cases: either $j \in \bar{N}_{\bar{\imath}}$ or $j \notin \bar{N}_{\bar{\imath}}$. In the first case, since $\bar{N}_{\bar{\imath}} \subseteq C_{[r_{\bar{\imath}}]}$, we have that $y^j - y^{j-1} = \left(x_j^{[r_{\bar{\imath}}]} - x_j^{[r_{\bar{\imath}}]}\right) e^j = 0$ and $T_{\bar{\imath}}\left(y^j\right) - T_{\bar{\imath}}\left(y^{j-1}\right) = 0$. In the second case, since $\bar{a}_{\bar{\imath}j} = 0$, we have that $T_{\bar{\imath}}\left(y^j\right) = T_{\bar{\imath}}\left(y^{j-1} + \left(x_j - x_j^{[r_{\bar{\imath}}]}\right) e^j\right) = T_{\bar{\imath}}\left(y^{j-1}\right)$, yielding that $T_{\bar{\imath}}\left(y^j\right) - T_{\bar{\imath}}\left(y^{j-1}\right) = 0$. By (B.54), it follows that $T_{\bar{\imath}}(x) - T_{\bar{\imath}}\left(x^{[r_{\bar{\imath}}]}\right) = 0$. ■

Proof of Proposition 22. By Proposition 19, since $\underline{A}(T)$ is nontrivial, there exist $W \in \mathcal{W}$ and $\varepsilon \in (0,1)$ such that

$$T(x) = \varepsilon W x + (1-\varepsilon)\,S(x) \qquad \forall x \in B \tag{B.55}$$

where $S : B \to B$ is a robust opinion aggregator. Moreover, W can be chosen to be such that $A(W) = \underline{A}(T)$. By induction and (B.55), we have that if $t \in \mathbb{N}$, then there exist $\gamma \in (0,1)$ and a

robust opinion aggregator $\tilde{S}: B \to B$ (which both depend on t) such that

$$T^t(x) = \gamma W^t x + (1-\gamma)\tilde{S}(x) \qquad \forall x \in B. \tag{B.56}$$

As usual, we denote the ij-th entry of W^t by $w_{ij}^{(t)}$. Since T is normalized, observe that $E(T) \supseteq D$. By induction, if $t \in \mathbb{N}$, then $D \subseteq E(T) \subseteq E(T^t)$. Since $A(W) = \underline{A}(T)$, it follows that $A(W)$ has a unique strongly connected and closed group M, and M is aperiodic under $A(W)$. By Jackson (2008), W is such that there exist $\bar{t} \in \mathbb{N}$ and $k \in N$ such that $w_{ik}^{(\bar{t})} > 0$ for all $i \in N$. Let $\tilde{S}$ denote the robust opinion aggregator for $\bar{t}$ in equation (B.56). We next show that $E\left(T^{\bar{t}}\right) = D$. By contradiction, assume that there exists $x \in B \backslash D$ such that $T^{\bar{t}}(x) = x$. Define $x_i = \min_{l \in N} x_l$ and $x_j = \max_{l \in N} x_l$. It follows that $x_j > x_i$ and $i \neq j$. We have two cases:

1. $x_k < x_j$. It follows that

$$\begin{aligned} 0 = \left\| T^{\bar{t}}(x) - x \right\|_\infty \geq \left| T_j^{\bar{t}}(x) - x_j \right| &= \left| \gamma \sum_{l=1}^n w_{jl}^{(\bar{t})} x_l + (1-\gamma)\tilde{S}_j(x) - x_j \right| \\ = \gamma \sum_{l=1}^n w_{jl}^{(\bar{t})}(x_j - x_l) + (1-\gamma)\left(x_j - \tilde{S}_j(x)\right) &\geq \gamma w_{jk}^{(\bar{t})}(x_j - x_k) > 0, \end{aligned}$$

   a contradiction.

2. $x_k > x_i$. It follows that

$$\begin{aligned} 0 = \left\| T^{\bar{t}}(x) - x \right\|_\infty \geq \left| T_i^{\bar{t}}(x) - x_i \right| &= \left| \gamma \sum_{l=1}^n w_{il}^{(\bar{t})} x_l + (1-\gamma)\tilde{S}_i(x) - x_i \right| \\ = \gamma \sum_{l=1}^n w_{il}^{(\bar{t})}(x_l - x_i) + (1-\gamma)\left(\tilde{S}_i(x) - x_i\right) &\geq \gamma w_{ik}^{(\bar{t})}(x_k - x_i) > 0, \end{aligned}$$

   a contradiction.

Cases 1 and 2 prove that $E\left(T^{\bar{t}}\right) = D$, and hence that $E(T) = D$. ■

Proof of Proposition 16. We omit the proof of point 2 which follows from well-known facts.[28]

1. Consider $\theta \in \mathbb{R} \backslash \{0\}$ and $\rho : \mathbb{R} \to \mathbb{R}_+$ defined by $\rho(\tilde{s}) = e^{\theta \tilde{s}} - \theta \tilde{s}$ for all $\tilde{s} \in \mathbb{R}$. It is easy to see that ρ is strictly convex and differentiable. Given $x \in B$ and $i \in N$, consider also the function $c \mapsto \phi_i^\theta(x - ce) = \sum_{j=1}^n w_{ij}\rho(x_j - c)$. Since ρ is strictly convex and differentiable, so is $c \mapsto \phi_i^\theta(x - ce)$. Given $x \in B$ and $i \in N$, this implies that the minimizer of the function $c \mapsto \phi_i^\theta(x - ce)$ is then uniquely pinned down by the first order conditions. Moreover, as we will immediately see, minimizing $c \mapsto \phi_i^\theta(x - ce)$ over I is equivalent to minimize it over $\mathbb{R}$. We compute

[28] The result for $\hat{\theta} = \infty$ is also known as Laplace's method. The case for $\hat{\theta} = -\infty$ is instead obtained from the previous one and by observing that $\theta x_j = -\theta(-x_j)$ and that $\theta \to -\infty$ yields $-\theta \to \infty$. The case of $\hat{\theta} = 0$ is a standard result in risk theory.

the first order conditions where $c^\star$ is the optimal value:

$$-\sum_{j=1}^{n} w_{ij}\left[\theta\exp\left(\theta\left(x_j-c^\star\right)\right)-\theta\right]=0 \implies \sum_{j=1}^{n} w_{ij}\exp\left(\theta x_j\right)=\exp\left(\theta c^\star\right) \implies c^\star=\frac{1}{\theta}\ln\left(\sum_{j=1}^{n} w_{ij}\exp\left(\theta x_j\right)\right)\in I.$$

Since i and x were arbitrarily chosen, equation (B.14) is satisfied. It is routine to show that T^θ is a robust opinion aggregator. As for the second part, fix $i,j\in N$. Observe that T_i^θ is continuously differentiable in the interior of B. Moreover, $\frac{\partial T_i^\theta}{\partial x_j}(x)>0$ for some $x\in\operatorname{int}B$ if and only if there exists $\varepsilon\in(0,1)$ such that $\frac{\partial T_i^\theta}{\partial x_j}(x)\geq\varepsilon$ for all $x\in\operatorname{int}B$ if and only if $w_{ij}>0$. By the Mean Value Theorem and since i and j were arbitrarily chosen, this implies that $\underline{A}\left(T^\theta\right)=\bar{A}\left(T^\theta\right)=A(W)$.

3. Let $S:\mathbb{R}^n\to\mathbb{R}^n_{++}$ be defined by $S_i(x)=\exp\left(\theta x_i\right)$ for all $i\in N$ and for all $x\in\mathbb{R}^n$. Define $\hat{T}:\mathbb{R}^n\to\mathbb{R}^n$ by $\hat{T}(x)=Wx$ for all $x\in\mathbb{R}^n$. We next show that

$$\left(T^\theta\right)^t=S^{-1}\hat{T}^tS \qquad \forall t\in\mathbb{N}. \tag{B.57}$$

By definition of T^θ, if $t=1$, then $T^\theta(x)=S^{-1}(WS(x))$ for all $x\in B$, yielding (B.57). Next, assume that (B.57) holds for t. We have that $\left(T^\theta\right)^{t+1}=T^\theta\left(T^\theta\right)^t=S^{-1}\hat{T}SS^{-1}\hat{T}^tS=S^{-1}\hat{T}^{t+1}S$, proving that (B.57) holds for $t+1$. By induction, (B.57) follows. Consider $x\in B$. By (B.15), it follows that $\lim_t\hat{T}^t(S(x))=\lim_t W^tS(x)=\left(\sum_{i=1}^n s_i\exp\left(\theta x_i\right)\right)e\in\mathbb{R}^n_{++}$. By (B.57) and since S^{-1} is continuous, we have that $\lim_t\left(T^\theta\right)^t(x)=\left(\frac{1}{\theta}\ln\left(\sum_{i=1}^n s_i\exp\left(\theta x_i\right)\right)\right)e=\bar{T}^\theta(x)$. Since x was arbitrarily chosen, the statement follows. ■

B.11.2 Vox populi, vox Dei?

To ease notation, we discuss the next ancillary result by dropping the n indexing. Let $\mathcal{W}_{\text{un}}$ denote the subset of $\mathcal{W}$ such that $W\in\mathcal{W}_{\text{un}}$ if and only if there exists an undirected and strongly connected graph with an $n\times n$ adjacency matrix A such that $w_{ij}=\frac{a_{ij}}{d_i}$ for all $i,j\in N$ where $d_i=\sum_{l=1}^n a_{il}$. It is well known that if $W\in\mathcal{W}_{\text{un}}$, then W is reversible and there exists a unique left Perron-Frobenius eigenvector $\bar{w}\in\Delta$, that is $\bar{w}^{\mathrm{T}}W=\bar{w}^{\mathrm{T}}$, and

$$\bar{w}_i=\frac{d_i}{\sum_{j=1}^n d_j} \qquad \forall i\in N.$$

In particular, note that

$$0\leq\bar{w}_k\leq\frac{1}{n}\frac{\max_{i\in N}d_i}{\min_{i\in N}d_i} \qquad \forall k\in N. \tag{B.58}$$

Finally, recall that if $W\in\mathcal{W}_{\text{un}}$ and $n\geq 2$, then the eigenvalues of W are real and, accounting for multiplicity, such that $1=\tilde{\lambda}_1\geq\tilde{\lambda}_2\geq\ldots\geq\tilde{\lambda}_n\geq-1$. We denote by λ_2 $\left(=\max_{i=2,\ldots,n}\left|\tilde{\lambda}_i\right|\right)$ the second largest eigenvalue in modulus (SLEM).

Lemma 28. *Let T be a robust opinion aggregator and $n\geq 2$. If there exist $\kappa\geq 1$ and $W\in\mathcal{W}_{\text{un}}$*

such that

$$\frac{\partial T_i}{\partial x_j}(x) \leq \kappa w_{ij} \qquad \forall x \in \mathcal{D}(T), \forall i,j \in N, \tag{B.59}$$

then

$$\bar{s}_{ij}(T) \leq \kappa^t \bar{w}_j + \sqrt{\frac{\max_{i\in N} d_i}{\min_{i\in N} d_i}} \kappa^t \lambda_2^t \quad \forall i,j \in N, \forall t \in \mathbb{N},$$

where $\lambda_2 \in \mathbb{R}_+$ is the SLEM of W.

Proof. Define $\hat{B} = \hat{I}^n$. Before starting, we introduce an useful object: the Clarke differential of T. By Rademacher's Theorem and since T is robust, T is Lipschitz continuous and, in particular, almost everywhere differentiable on $\mathbb{R}^n$. Recall that $\mathcal{D}(T)$ denotes the set of points of $\hat{B}$ where T is differentiable. We denote the Jacobian of T at $x \in \mathcal{D}(T)$ by $J_T(x)$. Since T is a robust opinion aggregator, we have that $J_T(x) \in \mathcal{W}$ for all $x \in \mathcal{D}(T)$. Finally, given $x \in \hat{B}$, we denote the Clarke differential of T at x by $\partial T(x)$ where

$$\partial T(x) = \operatorname{co}\left\{ W \in \mathcal{W} : W = \lim_k J_T\left(x^k\right) \text{ s.t. } x^k \to x \text{ and } x^k \in \mathcal{D}(T) \right\}.$$

By Theorem 7, recall that $\bar{T} \circ T = \bar{T}$, yielding that $\bar{T}_i \circ T = \bar{T}_i$ for all $i \in N$. By the Chain rule, we have that

$$\partial \bar{T}_i(x) \subseteq \operatorname{co}\left\{ \partial \bar{T}_i(T(x)) \partial T(x) \right\} \qquad \forall i \in N, \forall x \in \hat{B} \tag{B.60}$$

where $\partial \bar{T}_i(T(x)) \partial T(x)$ is the set of probability vectors $p \in \Delta$ such that $p^{\mathrm{T}} = q^{\mathrm{T}} \tilde{W}$ where $q \in \partial \bar{T}_i(T(x))$ and $\tilde{W} \in \partial T(x)$. By definition of $\partial T(x)$ and since T satisfies (B.59), we have that

$$\tilde{W} \leq \kappa W \qquad \forall \tilde{W} \in \partial T(x), \forall x \in \hat{B}. \tag{B.61}$$

We next prove by induction that for each $x \in \hat{B}$, for each $i \in N$, for each $p \in \partial \bar{T}_i(x)$, and for each $t \in \mathbb{N}$ there exists $q \in \Delta$ such that

$$p^{\mathrm{T}} \leq q^{\mathrm{T}} \left(\kappa^t W^t\right). \tag{B.62}$$

By (B.61), we have that $q^{\mathrm{T}} \tilde{W} \leq q^{\mathrm{T}}(\kappa W)$ for all $q \in \partial \bar{T}_i(T(x))$, for all $\tilde{W} \in \partial T(x)$, for all $x \in \hat{B}$, and for all $i \in N$. By (B.60) and since $\partial \bar{T}_i(T(x)) \subseteq \Delta$ for all $i \in N$, this implies that (B.62) holds for $t = 1$. Next, we assume that the statement holds for t and we show it holds for $t+1$. Consider $x \in \hat{B}$, $i \in N$, and $p \in \partial \bar{T}_i(x)$. By (B.60), we have that there exist $\left\{\tilde{q}^k\right\}_{k=1}^m \subseteq \partial \bar{T}_i(T(x))$, $\left\{\tilde{W}_k\right\}_{k=1}^m \subseteq \partial T(x)$, and $\{\alpha_k\}_{k=1}^m \subseteq [0,1]$ such that $\sum_{k=1}^m \alpha_k = 1$ and $p^{\mathrm{T}} = \sum_{k=1}^m \alpha_k \left(\tilde{q}^k\right)^{\mathrm{T}} \tilde{W}_k$. By inductive hypothesis and since $\left\{\tilde{q}^k\right\}_{k=1}^m \subseteq \partial \bar{T}_i(T(x))$ and $T(x) \in \hat{B}$, for each $k \in \{1,...,m\}$ we have that $\left(\tilde{q}^k\right)^{\mathrm{T}} \kappa W \leq \left(\hat{q}^k\right)^{\mathrm{T}} \left(\kappa^t W^t\right) \kappa W = \left(\hat{q}^k\right)^{\mathrm{T}} \left(\kappa^{t+1} W^{t+1}\right)$ for some $\hat{q}^k \in \Delta$. By (B.61), this yields that

$$p^{\mathrm{T}} = \sum_{k=1}^m \alpha_k \left(\tilde{q}^k\right)^{\mathrm{T}} \tilde{W}_k \leq \sum_{k=1}^m \alpha_k \left(\tilde{q}^k\right)^{\mathrm{T}} (\kappa W) \leq \left(\sum_{k=1}^m \alpha_k \left(\hat{q}^k\right)^{\mathrm{T}} \right) \left(\kappa^{t+1} W^{t+1}\right).$$

Since $\sum_{k=1}^m \alpha_k \hat{q}^k \in \Delta$ and x, i, as well as p were arbitrarily chosen, the inductive step follows. By induction, (B.62) holds.

By Bremaud (2017) and since $W \in \mathcal{W}_{\text{un}}$, we have that

$$\max_{i,j \in N} \left| w_{ij}^{(t)} - \bar{w}_j \right| \leq \sqrt{\frac{\max_{i \in N} d_i}{\min_{i \in N} d_i}} \lambda_2^t \quad \forall t \in \mathbb{N}.$$

Consider $\bar{x} \in \hat{B}$, $p \in \partial \bar{T}_i(\bar{x})$, $i \in N$, and $t \in \mathbb{N}$. By (B.62), this implies that $p^{\mathrm{T}} \leq q^{\mathrm{T}}(\kappa^t W^t) = \kappa^t q^{\mathrm{T}} W^t$ for some $q \in \Delta$, yielding that

$$\begin{aligned} p_j &\leq \kappa^t \sum_{i=1}^n q_i w_{ij}^{(t)} = \kappa^t \bar{w}_j + \kappa^t \sum_{i=1}^n q_i \left(w_{ij}^{(t)} - \bar{w}_j \right) \\ &\leq \kappa^t \bar{w}_j + \kappa^t \sum_{i=1}^n q_i \left| w_{ij}^{(t)} - \bar{w}_j \right| \leq \kappa^t \bar{w}_j + \kappa^t \sqrt{\frac{\max_{i \in N} d_i}{\min_{i \in N} d_i}} \lambda_2^t \qquad \forall j \in N. \end{aligned}$$

Since $\bar{x}$, p, and t were arbitrarily chosen, and $\nabla \bar{T}_i(x) \in \partial \bar{T}_i(x)$ for all $x \in \mathcal{D}\left(\bar{T}\right)$, we have that

$$\bar{s}_{ij}(T) = \sup_{x \in \mathcal{D}\left(\bar{T}\right)} \frac{\partial \bar{T}_i}{\partial x_j}(x) \leq \kappa^t \bar{w}_j + \kappa^t \sqrt{\frac{\max_{i \in N} d_i}{\min_{i \in N} d_i}} \lambda_2^t \quad \forall j \in N, \forall t \in \mathbb{N}.$$

Since i was arbitrarily chosen, the statement follows. ■

Proof of Proposition 17. 1. Fix $n \in \mathbb{N}$ and define $\hat{B} = \hat{I}^n$. Since $T(n)$ is a robust opinion aggregator, we have that $T(n)$ is Lipschitz continuous. By Rademacher's Theorem, this implies that $T(n)$ is almost everywhere differentiable on $\hat{B}$ and, in particular, Clarke differentiable. Since $T_j(n)$ is monotone and translation invariant for all $j \in N$, note that $\nabla T_j(n)(x) \in \Delta_n$ for all $x \in \mathcal{D}(T(n))$ and for all $j \in N$. Recall that the Clarke's differential is the set:

$$\partial T_j(n)(\bar{x}) = \operatorname{co}\left\{ p \in \Delta_n : p = \lim_k \nabla T_j(n)\left(x^k\right) \text{ s.t. } x^k \to \bar{x} \text{ and } x^k \in \mathcal{D}(T(n)) \right\} \qquad \forall \bar{x} \in \hat{B}, \forall j \in N. \tag{B.63}$$

By Theorem 7, recall that $\bar{T}(n) \circ T(n) = \bar{T}(n)$. Fix $\bar{x} \in \hat{B}$. Define by $\Pi_{j=1}^n \partial T_j(n)(\bar{x})$ the collection of all $n \times n$ square matrices whose j-th row is an element of $\partial T_j(n)(\bar{x})$. From the previous part of the proof, we have that $\Pi_{j=1}^n \partial T_j(n)(\bar{x}) \subseteq \mathcal{W}$. For each $i \in N$, define

$$\begin{aligned} &\partial \bar{T}_i(n)(T(n)(\bar{x}))\, \Pi_{j=1}^n \partial T_j(n)(\bar{x}) \\ &= \left\{ \tilde{w} \in \Delta_n : \exists p \in \partial \bar{T}_i(n)(T(n)(\bar{x})), \exists W \in \Pi_{j=1}^n \partial T_j(n)(\bar{x}) \text{ s.t. } p^{\mathrm{T}} W = \tilde{w}^{\mathrm{T}} \right\}. \end{aligned}$$

By the Chain Rule, we have that for each $i \in N$

$$\partial \bar{T}_i(n)(\bar{x}) \subseteq \operatorname{co}\left\{ \partial \bar{T}_i(n)(T(n)(\bar{x}))\, \Pi_{j=1}^n \partial T_j(n)(\bar{x}) \right\}. \tag{B.64}$$

By assumption, we have that for each $i,j\in N$

$$\sup_{x\in\mathcal{D}(T(n))}\frac{\partial T_i(n)}{\partial x_j}(x)\leq\frac{\kappa}{\bar{d}_i(n)}\leq\frac{\kappa}{\bar{d}_{\min}(n)}. \tag{B.65}$$

By (B.63) and (B.65), we have that $0\leq p_j\leq\frac{\kappa}{\bar{d}_{\min}(n)}$ for all $p\in\partial T_i(n)(\bar{x})$ and for all $i,j\in N$. By (B.64), $0\leq p_j\leq\frac{\kappa}{\bar{d}_{\min}(n)}$ for all $p\in\partial\bar{T}_i(n)(\bar{x})$ and for all $i,j\in N$. Finally, observe that if $x\in\mathcal{D}\left(\bar{T}(n)\right)$, we have that $\nabla\bar{T}_i(n)(x)\in\partial\bar{T}_i(n)(x)$ and, in particular, $\frac{\partial\bar{T}_i(n)}{\partial x_j}(x)\leq\frac{\kappa}{\bar{d}_{\min}(n)}$ for all $i,j\in N$. This yields that

$$\bar{s}_{ij}(T(n))=\sup_{x\in\mathcal{D}\left(\bar{T}(n)\right)}\frac{\partial\bar{T}_i(n)}{\partial x_j}(x)\leq\frac{\kappa}{\bar{d}_{\min}(n)}\quad\forall i,j\in N.$$

Therefore, since $\lim_n\frac{\sqrt{n}}{\bar{d}_{\min}(n)}=0$ and n was arbitrarily chosen, we have that for each $\iota\in\mathbb{N}$

$$\lim_n\sum_{j=1}^n\left(\bar{s}_{\iota j}(T(n))\right)^2\leq\lim_n\sum_{j=1}^n\left(\frac{\kappa}{\bar{d}_{\min}(n)}\right)^2=\lim_n\frac{n\kappa^2}{\left(\bar{d}_{\min}(n)\right)^2}=0.$$

By point 1 of Theorem 9, this implies point 1.

2. For each $n\in\mathbb{N}$ denote by $W(n)\in\mathcal{W}$ the stochastic matrix whose ij-th entry is $\bar{a}_{ij}(n)/\bar{d}_i(n)$. By assumption, each $W(n)$ is in $\mathcal{W}_{\text{un}}$ and has a unique left Perron-Frobenius eigenvector that we denote $\bar{w}(n)\in\Delta_n$. By assumption, it follows that there exists $\bar{\kappa}>1$ and $\varepsilon>0$ such that $\{T(n)\}_{n\in\mathbb{N}}$ is $\bar{\kappa}$-dominated and $\sup_{n\in\mathbb{N}}\lambda_2(n)<\frac{1}{\bar{\kappa}^{2+\varepsilon}}$. Set $\bar{m}=\sup_{n\in\mathbb{N}}\sqrt{\frac{\bar{d}_{\max}(n)}{\bar{d}_{\min}(n)}}\in\mathbb{R}_+$ and $t_n=\max\left\{1,\left\lfloor\log_{\bar{\kappa}^2}\left(\max_{k\in N}\bar{w}_k(n)\right)^{-\alpha}\right\rfloor\right\}$ for all $n\in\mathbb{N}$ where $\alpha=\frac{1+\delta}{1+\varepsilon}$ with $\delta\in(0,\varepsilon)$. Note that $\alpha\in(0,1)$ and $(1+\varepsilon)\alpha=1+\delta$. By (B.58), we have that $0\leq\max_{k\in N}\bar{w}_k(n)\leq\bar{m}^2/n$ for all $n\in\mathbb{N}$ and, in particular, $\lim_n\max_{k\in N}\bar{w}_k(n)=0$. By Lemma 28, recall that

$$0\leq\bar{s}_{ij}(T(n))\leq\bar{\kappa}^{t_n}\bar{w}_j(n)+\bar{m}\bar{\kappa}^{t_n}\lambda_2^{t_n}(n)\quad\forall i,j\in N,\forall n\in\mathbb{N}\backslash\{1\}.$$

It follows that

$$\bar{s}_{ij}(T(n))^2\leq\bar{\kappa}^{2t_n}\bar{w}_j(n)^2+2\bar{\kappa}^{t_n}\bar{w}_j(n)\bar{m}\bar{\kappa}^{t_n}\lambda_2^{t_n}(n)+\bar{m}^2\bar{\kappa}^{2t_n}\lambda_2^{2t_n}(n)\quad\forall i,j\in N,\forall n\in\mathbb{N}\backslash\{1\}$$

and

$$\sum_{j=1}^n\bar{s}_{ij}(T(n))^2\leq a_n+b_n+c_n\qquad\forall i\in N,\forall n\in\mathbb{N}\backslash\{1\} \tag{B.66}$$

where $a_n=\sum_{j=1}^n\bar{\kappa}^{2t_n}\bar{w}_j(n)^2$, $b_n=\sum_{j=1}^n2\bar{m}\bar{\kappa}^{2t_n}\lambda_2^{t_n}(n)\bar{w}_j(n)$, and $c_n=\sum_{j=1}^n\bar{m}^2\bar{\kappa}^{2t_n}\lambda_2^{2t_n}(n)$ for all $n\in\mathbb{N}\backslash\{1\}$. Note that these three sequences only depend on n and not on $i,j\in N$. We next show that $\lim_n a_n=\lim_n b_n=\lim_n c_n=0$. Since $\lim_n\max_{k\in N}\bar{w}_k(n)=0$ and $\bar{\kappa}>1$, observe that $\lim_n\left(\max_{k\in N}\bar{w}_k(n)\right)^{-\alpha}=\infty$ and $\lim_n\log_{\bar{\kappa}^2}\left(\max_{k\in N}\bar{w}_k(n)\right)^{-\alpha}=\infty$. This implies that $\lim_n t_n=\infty$. Moreover, there exists $\bar{n}\in\mathbb{N}\backslash\{1\}$ such that $\log_{\bar{\kappa}^2}\left(\max_{k\in N}\bar{w}_k(n)\right)^{-\alpha}-1\leq t_n=$

$\left\lfloor \log_{\bar{\kappa}^2} \left(\max_{k\in N} \bar{w}_k(n)\right)^{-\alpha} \right\rfloor \leq \log_{\bar{\kappa}^2} \left(\max_{k\in N} \bar{w}_k(n)\right)^{-\alpha}$ for all $n \geq \bar{n}$.

- Since $1-\alpha \in (0,1)$, $\bar{\kappa} > 1$, and $\lim_n \max_{k\in N} \bar{w}_k(n) = 0$, observe that for each $n \geq \bar{n}$

$$\begin{aligned} 0 \leq a_n &= \bar{\kappa}^{2t_n} \sum_{j=1}^{n} \bar{w}_j(n)^2 \leq \bar{\kappa}^{2t_n} \max_{k\in N} \bar{w}_k(n) \sum_{j=1}^{n} \bar{w}_j(n) \\ &= \bar{\kappa}^{2t_n} \max_{k\in N} \bar{w}_k(n) \leq \left(\bar{\kappa}^2\right)^{\log_{\bar{\kappa}^2}\left(\max_{k\in N} \bar{w}_k(n)\right)^{-\alpha}} \max_{k\in N} \bar{w}_k(n) = \left(\max_{k\in N} \bar{w}_k(n)\right)^{1-\alpha} \to 0 \text{ as } n \to \infty. \end{aligned}$$

- Since $\bar{\kappa} > 1$, we have that $0 \leq \sup_{n\in\mathbb{N}} \bar{\kappa}^2 \lambda_2(n) \leq \frac{1}{\bar{\kappa}^{\varepsilon}} < 1$. Since $t_n \in \mathbb{N}$ for all $n \in \mathbb{N}$ and $\lim_n t_n = \infty$, this implies that

$$0 \leq b_n = 2\bar{m}\bar{\kappa}^{2t_n} \lambda_2^{t_n}(n) \sum_{j=1}^{n} \bar{w}_j(n) = 2\bar{m}\left(\bar{\kappa}^2 \lambda_2(n)\right)^{t_n} \leq 2\bar{m}\left(\sup_{n\in\mathbb{N}} \bar{\kappa}^2 \lambda_2(n)\right)^{t_n} \to 0 \text{ as } n \to \infty.$$

- Since $\sup_{n\in\mathbb{N}} \lambda_2(n) \leq \frac{1}{\bar{\kappa}^{2+\varepsilon}}$, we have that $\sup_{n\in\mathbb{N}} \lambda_2^2(n) \leq \frac{1}{\bar{\kappa}^{4+2\varepsilon}}$, that is, $0 \leq \sup_{n\in\mathbb{N}} \bar{\kappa}^2 \lambda_2^2(n) \leq \frac{1}{\bar{\kappa}^{2+2\varepsilon}}$. Since $t_n \in \mathbb{N}$ for all $n \in \mathbb{N}$, this implies that $\left(\sup_{n\in\mathbb{N}} \bar{\kappa}^2 \lambda_2^2(n)\right)^{t_n} \leq \left(\frac{1}{\bar{\kappa}^{2+2\varepsilon}}\right)^{t_n}$ for all $n \in \mathbb{N}$.

  Since $(1+\varepsilon)\alpha = 1+\delta$ and $\delta > 0$, we obtain that for each $n \geq \bar{n}$

$$\begin{aligned} 0 \leq c_n &= \bar{m}^2 n \bar{\kappa}^{2t_n} \lambda_2^{2t_n}(n) = \bar{m}^2 n \left(\bar{\kappa}^2 \lambda_2^2(n)\right)^{t_n} \leq \bar{m}^2 n \left(\frac{1}{\bar{\kappa}^{2+2\varepsilon}}\right)^{t_n} = \bar{m}^2 n \left(\frac{1}{\bar{\kappa}^{2(1+\varepsilon)}}\right)^{t_n} \\ &\leq \bar{m}^2 n \left(\bar{\kappa}^2\right)^{-(1+\varepsilon)\left(\log_{\bar{\kappa}^2}\left(\max_{k\in N} \bar{w}_k(n)\right)^{-\alpha} - 1\right)} = \bar{m}^2 n \bar{\kappa}^{2(1+\varepsilon)} \left(\bar{\kappa}^2\right)^{-(1+\varepsilon)\log_{\bar{\kappa}^2}\left(\max_{k\in N} \bar{w}_k(n)\right)^{-\alpha}} \\ &= \bar{m}^2 \bar{\kappa}^{2(1+\varepsilon)} n \left(\max_{k\in N} \bar{w}_k(n)\right)^{(1+\varepsilon)\alpha} \leq \bar{m}^2 \bar{\kappa}^{2(1+\varepsilon)} n \left(\frac{\bar{m}^2}{n}\right)^{(1+\varepsilon)\alpha} \\ &= \bar{m}^{4+2\delta} \bar{\kappa}^{2(1+\varepsilon)} n^{-\delta} \to 0 \text{ as } n \to \infty. \end{aligned}$$

By (B.66), we have $\lim_n \sum_{j=1}^{n} s_{\iota j}(T(n))^2 = 0$ for all $\iota \in \mathbb{N}$. By point 1 of Theorem 9 and since $\{T(n)\}_{n\in\mathbb{N}}$ is a sequence of odd robust opinion aggregators and $\{\varepsilon_i(n)\}_{i\in N, n\in\mathbb{N}}$ is symmetric, point 2. of the statement follows. ■

B.11.3 Discussion

Proof of Lemma 26. Fix $i \in N$. Consider $\tilde{z} \in \mathbb{R}^n$ such that $\tilde{z} \gg 0$. Define $z = \tilde{z} - \min_{j\in N} \tilde{z}_j e$, $v = 0$, and $h = \min_{j\in N} \tilde{z}_j$. Note that $z \geq v$ as well as $h \in \mathbb{R}_{++}$. Since ϕ has increasing shifts and is sensitive, we obtain that

$$\phi_i(\tilde{z}) - \phi_i\left(\tilde{z} - \min_{j\in N} \tilde{z}_j e\right) = \phi_i(z + he) - \phi_i(z) \geq \phi_i(v + he) - \phi_i(v) = \phi_i\left(\min_{j\in N} \tilde{z}_j e\right) - \phi_i(0) > 0,$$

proving the first inequality. A symmetric argument yields the second inequality. ■

Proof of Lemma 27. Fix $i \in N$ and $x \in \mathbb{R}^n$. Define $g_{i,x} : \mathbb{R} \to \mathbb{R}_+$ by $g_{i,x}(c) = \phi_i(x + ce)$ for all $c \in \mathbb{R}$. Consider $c_1, c_2 \in \mathbb{R}$ such that $c_1 > c_2$ and $h > 0$. Since $\phi \in \Phi_R$ and $x + c_1 e \geq x + c_2 e$, it follows that

$$\begin{aligned} g_{i,x}(c_1 + h) - g_{i,x}(c_1) &= \phi_i((x + c_1 e) + he) - \phi_i(x + c_1 e) \\ &\geq \phi_i((x + c_2 e) + he) - \phi_i(x + c_2 e) = g_{i,x}(c_2 + h) - g_{i,x}(c_2). \end{aligned}$$

It follows that $g_{i,x}$ is midconvex. Next, fix $c \in \mathbb{R}$ and $c' \in (c-1, c+1)$. Set $c_1 = 2c - c'$, $c_2 = c - 1$, and $h = c' - (c-1)$. Since $c_1 > c_2$, $h > 0$, and $\phi_i \geq 0$, we have that

$$g_{i,x}(c') - g_{i,x}(c-1) \leq g_{i,x}(c+1) - g_{i,x}(2c - c') \implies 0 \leq g_{i,x}(c') \leq g_{i,x}(c-1) + g_{i,x}(c+1).$$

Since c' was arbitrarily chosen, we have that $g_{i,x}$ is bounded on $(c-1, c+1)$. It follows that $g_{i,x}$ is continuous and convex. Finally, observe that $f_{i,x} = g_{i,x} \circ h$ where $h(c) = -c$ for all $c \in \mathbb{R}$, yielding that $f_{i,x}$ is convex and continuous being the composition of a convex and continuous function with an affine and continuous function. Next, assume that ϕ has also strictly increasing shifts and, in particular, has increasing shifts. By the previous part of the proof, $g_{i,x}$ is convex. By contradiction, assume that $g_{i,x}$ is not strictly convex. This implies that there exists an interval $[d_2, d_1]$, with $d_2 < d_1$, where $g_{i,x}$ is affine. Define $c_1 = \frac{1}{2}d_1 + \frac{1}{2}d_2$, $c_2 = d_2$, and $h = (d_1 - d_2)/2$. Note that $c_1 > c_2$ and $h > 0$. Since ϕ has strictly increasing shifts, by the same computations of the previous part of the proof, we have that

$$\begin{aligned} g_{i,x}(d_1) - g_{i,x}\left(\frac{1}{2}d_1 + \frac{1}{2}d_2\right) &= g_{i,x}(c_1 + h) - g_{i,x}(c_1) \\ &> g_{i,x}(c_2 + h) - g_{i,x}(c_2) = g_{i,x}\left(\frac{1}{2}d_1 + \frac{1}{2}d_2\right) - g_{i,x}(d_2), \end{aligned}$$

yielding that $g_{i,x}\left(\frac{1}{2}d_1 + \frac{1}{2}d_2\right) < \frac{1}{2}g_{i,x}(d_1) + \frac{1}{2}g_{i,x}(d_2)$, a contradiction with affinity. Since $g_{i,x}$ is strictly convex, so is $f_{i,x} = g_{i,x} \circ h$. ■

Proof of Proposition 18. Before starting, we make few observations about strong convexity. Since each ρ_i is strongly convex and twice continuously differentiable, we have that for each $i \in N$ there exists $\alpha_i > 0$ such that $\rho_i''(s) \geq \alpha_i$ for all $s \in \mathbb{R}$. Moreover, we have that for each $i \in N$

$$(\rho_i'(s_1) - \rho_i'(s_2))(s_1 - s_2) \geq \alpha_i (s_1 - s_2)^2 \quad \forall s_1, s_2 \in \mathbb{R}. \tag{B.67}$$

Finally, since each ρ_i is twice continuously differentiable and I is compact, for each $i \in N$ we have that there exists $L_i > 0$ such that

$$|\rho_i'(s_1) - \rho_i'(s_2)| \leq L_i |s_1 - s_2| \quad \forall s_1, s_2 \in [\min I - \max I, \max I - \min I]. \tag{B.68}$$

Recall that $\phi_i : \mathbb{R}^n \to \mathbb{R}_+$ is defined by $\phi_i(z) = \sum_{j=1}^n w_{ij}\rho_i(z_j)$ for all $z \in \mathbb{R}^n$ and for all $i \in N$. By assumption, $\phi \in \Phi_A \subseteq \Phi_R$. Since $\rho_i'' \geq \alpha_i > 0$ for all $i \in N$, this implies that ρ_i is strictly convex for

all $i \in N$. Standard computations yield that ϕ has strictly increasing shifts. By Proposition 23, we have that $\mathbf{T}^{\phi} = T^{\phi}$ is single-valued and a robust opinion aggregator from B to B. Moreover, $T_i^{\phi}(x)$ is the unique solution of

$$\min_{c \in \mathbb{R}} \phi_i (x - ce) = \min_{c \in I} \phi_i (x - ce) \qquad \forall i \in N, \forall x \in B. \tag{B.69}$$

Fix $i \in N$. Since ρ_i is differentiable and convex, so is the map $c \mapsto \phi_i (x - ce)$ for all $x \in B$. The solution of (B.69) is then given by the first order condition $\sum_{j=1}^{n} w_{ij} \rho_i' \left(x_j - T_i^{\phi}(x) \right) = 0$ for all $x \in B$. Consider $x \in B$, $h > 0$, and $l \in N$ such that $x + he^l \in B$. We have that

$$\sum_{j=1}^{n} w_{ij} \rho_i' \left(x_j - T_i^{\phi}(x) \right) = 0 \text{ and } \sum_{j=1}^{n} w_{ij} \rho_i' \left(x_j + he_j^l - T_i^{\phi}\left(x + he^l \right) \right) = 0. \tag{B.70}$$

Note that if $w_{il} = 0$, then $\sum_{j=1}^{n} w_{ij} \rho_i' \left(x_j + he_j^l - c \right) = \sum_{j=1}^{n} w_{ij} \rho_i' (x_j - c)$ for all $c \in \mathbb{R}$, proving that $T_i^{\phi}\left(x + he^l \right) = T_i^{\phi}(x)$. Since x and h were arbitrarily chosen, we have that $w_{il} = 0$ implies $\bar{a}_{il} = 0$. In particular, since i and l were arbitrarily chosen, we have that $A(W) \geq \bar{A}\left(T^{\phi} \right)$.

Next, assume that $w_{il} > 0$. By (B.68), (B.70), and (B.67) and since T^{ϕ} is monotone and $h > 0$, we can conclude that

$$\begin{aligned} L_i \left(T_i^{\phi}\left(x + he^l \right) - T_i^{\phi}(x) \right) &\geq \sum_{j=1}^{n} w_{ij} \rho_i' \left(x_j + he_j^l - T_i^{\phi}(x) \right) - \sum_{j=1}^{n} w_{ij} \rho_i' \left(x_j + he_j^l - T_i^{\phi}\left(x + he^l \right) \right) \\ &= \sum_{j=1}^{n} w_{ij} \rho_i' \left(x_j + he_j^l - T_i^{\phi}(x) \right) - \sum_{j=1}^{n} w_{ij} \rho_i' \left(x_j - T_i^{\phi}(x) \right) \\ &= w_{il} \left[\rho_i' \left(x_l + h - T_i^{\phi}(x) \right) - \rho_i' \left(x_l - T_i^{\phi}(x) \right) \right] \geq w_{il} \alpha_i h, \end{aligned}$$

proving that $T_i^{\phi}\left(x + he^l \right) - T_i^{\phi}(x) \geq \varepsilon_{il} h$ where $\varepsilon_{il} = L_i^{-1} w_{il} \alpha_i / 2 \in (0, 1)$. Since x and h were arbitrarily chosen, we have that $w_{il} > 0$ implies $\underline{a}_{il} = 1$. In particular, since i and l were arbitrarily chosen, we have that $\underline{A}\left(T^{\phi} \right) \geq A(W)$. Since $\bar{A}\left(T^{\phi} \right) \geq \underline{A}\left(T^{\phi} \right)$, we can conclude that $A(W) = \bar{A}\left(T^{\phi} \right) = \underline{A}\left(T^{\phi} \right)$, proving the statement. ■

Bibliography

[1] C. D. Aliprantis and K. C. Border, *Infinite Dimensional Analysis*, 3rd ed., Springer-Verlag, Berlin, 2006.

[2] P. Brémaud, *Discrete Probability Models and Methods*, Springer, New York, 2017.

[3] F. H. Clarke, *Optimization and Nonsmooth Analysis*, SIAM, Philadelphia, 1990.

[4] S. Cerreia-Vioglio, F. Maccheroni, and M. Marinacci, A characterization of probabilities with full support and the Laplace method, *Journal of Optimization Theory and Applications*, 181, 470–478, 2019.

[5] S. Cerreia-Vioglio, F. Maccheroni, M. Marinacci, and A. Rustichini, Niveloids and their extensions: Risk measures on small domains, *Journal of Mathematical Analysis and Applications*, 413, 343–360, 2014.

[6] M. O. Jackson, *Social and Economic Networks*, Princeton University Press, Princeton, 2008.

[7] G. G. Lorentz, A contribution to the theory of divergent sequences, *Acta Mathematica*, 80, 167–190, 1948.

[8] A. W. Roberts and D. E. Varberg, *Convex Functions*, Academic Press, New York, 1973.