Available online at www.sciencedirect.com

ScienceDirect

Journal of Economic Theory 200 (2022) 105399

JOURNAL OF Economic Theory

www.elsevier.com/locate/jet

Bayesian social aggregation with accumulating evidence [☆]

Marcus Pivato

THEMA, CY Cergy Paris Université, France

Received 20 November 2020; final version received 16 November 2021; accepted 5 December 2021
Available online 9 December 2021

This paper is dedicated to the memory of Philippe Mongin (1950-2020)

Abstract

How should we aggregate the ex ante preferences of Bayesian agents with heterogeneous beliefs? Suppose the state of the world is described by a random process that unfolds over time. Different agents have different beliefs about the probabilistic laws governing this process. As new information is revealed over time by the process, agents update their beliefs and preferences via Bayes rule. Consider a Pareto principle that applies only to preferences which remain stable in the long run under these updates. I show that this "eventual Pareto" principle implies that the social planner must be a utilitarian. But it does not impose any relationship between the beliefs of the individuals and those of the planner, except for a weak compatibility condition.

JEL classification: D70; D81

Keywords: Subjective expected utility; Utilitarian; Ex ante Pareto; Stochastic process

[☆] This research supported by Labex MME-DII (ANR11-LBX-0023-01) and CHOp (ANR-17-CE26-0003). I thank Jean Baccelli, Franz Dietrich, Marc Fleurbaey, Brian Hill, Xiangyu Qu, John Weymark, and Stéphane Zuber for helpful comments. I am especially grateful to two anonymous reviewers, the Associate Editor and Lead Editor for their very careful attention to this manuscript and many perceptive remarks.

E-mail address: marcus-pivato@cyu.fr.

https://doi.org/10.1016/j.jet.2021.105399
0022-0531/© 2021 Elsevier Inc. All rights reserved.

1. Introduction

In a watershed 1955 paper, Harsanyi considered social decisions in the presence of risk. He showed that if all individuals and the social planner are expected utility maximizers, and the planner's ex ante preferences satisfy the Pareto property with respect to the individual ex ante preferences, then the planner is a "utilitarian", in the sense that the social utility function is a weighted average of the individual utility functions. While the connection between this formal result and philosophical utilitarianism can be debated (Weymark, 1991), there can be no doubt that it has been highly influential in welfare economics.

Harsanyi formulated his result in the von Neumann-Morgenstern framework, where risks are described by objective probabilities. But around the same time, Savage (1954) developed a theory of decision-making under uncertainty, where each agent maximizes expected utility relative to an idiosyncratic probability distribution, interpreted as her "subjective beliefs". This raised a question: does Harsanyi's result remains valid when the individuals and the planner are Savage-style subjective expected utility maximizers, perhaps with different beliefs? In an influential paper, Mongin (1995) answered this question in the negative: in the Savage framework, the planner can satisfy the ex ante Pareto axiom if and only if all agents have the *same* beliefs. Since such homogeneity of beliefs is empirically implausible and perhaps even normatively undesirable, Mongin interpreted this result as an impossibility theorem.

This seemed to deal a fatal blow to Harsanyi's project of founding utilitarianism in the theory of rational decisions. But in 2004, Gilboa et al. rescued Harsanyi by weakening the ex ante Pareto axiom so that it applied *only* to preferences between acts that depend upon events about whose probabilities all agents agreed. They showed that this restricted ex ante Pareto axiom was compatible with belief heterogeneity, but still strong enough to imply *not only* that the social planner is a utilitarian, but *also* that her subjective beliefs are a weighted average of the subjective beliefs of the individuals.

As already noted by Mongin (1995, 1997), when individuals have heterogeneous beliefs, a naïve application of the Pareto axiom to their ex ante preferences may lead to cases of *spurious unanimity*, where there is disagreement in both the utility functions and the beliefs of the individuals, but these disagreements "cancel out", to create apparent agreement in their ex ante preferences over acts. The key insight of Gilboa et al. (2004) was to restrict the Pareto axiom to *exclude* such cases of spurious unanimity, by applying it only when the relevant underlying beliefs are in agreement. Gilboa et al.'s landmark result became the point of reference for all subsequent literature on social decisions under uncertainty (Chambers and Hayashi, 2006, 2014; Alon and Gayer, 2016; Danan et al., 2016; Billot and Vergopoulos, 2016; Zuber, 2016; Qu, 2017; Desai et al., 2018; Sprumont, 2018, 2019; Hayashi and Lombardi, 2019; Ceron and Vergopoulos, 2019; Brandl, 2021; Dietrich, 2021; Billot and Qu, 2021).[1] Most of these papers either employ non-SEU preferences or follow Gilboa et al. (2004) in weakening ex ante Pareto, so as to avoid cases of spurious unanimity while still axiomatizing a simultaneous aggregation of utilities and beliefs. To distinguish mutually beneficial financial trades from mere "betting", Gayer et al. (2014) and Gilboa et al. (2014) consider weak ex ante Pareto conditions that are still stronger than the one proposed by Gilboa et al. (2004). But Mongin and Pivato (2020, §6) recently argued that Gilboa et al.'s restricted Pareto axiom is actually *too strong*. The linear pooling of beliefs which they derive from this axiom might not be an asset, but a liability, because there

[1] See Mongin and Pivato (2016) or Fleurbaey (2018) for reviews of this literature. See also §4.8.

are situations where linear pooling of beliefs is not desirable —in particular, it is not compatible with Bayesian updating under the arrival of new information. More fundamentally, Gilboa et al.'s restricted Pareto axiom is still vulnerable to a sort of "spurious unanimity" in *beliefs themselves*: different individuals may assign the same probability to an event, but for different and incompatible reasons. For example: starting from the same prior, they might Bayes-update on different private information, but coincidentally end up assigning the same posterior probability to some event, even though their combined information would yield a *different* conditional probability for this event. Mongin and Pivato refer to this as *complementary ignorance*.

These cases of spurious unanimity and complementary ignorance suggest that ex ante Pareto is a mistake, even in weakened form. Perhaps we should jettison it entirely, and fall back on a purely ex post approach. But in some situations, this is not possible. In a temporally extended decision problem with an infinite planning horizon, there is a progressive resolution of uncertainty over time, but this process never terminates in a state of final certainty. The ex post outcome on Monday evening becomes the ex ante situation on Tuesday morning, and the ex post of Tuesday evening becomes the ex ante of Wednesday morning, *ad infinitum*. So a purely ex post approach may be unavailable.

Even in decisions without an explicitly intertemporal element, we may never obtain total knowledge of the state of nature. Risse (2001, 2003) and Hild et al. (2003, 2008) consider decisions under uncertainty with a non-atomic Boolean algebra of events, so that any event can always be split up into smaller events, encoding more precise information. They construct an example where ever-more-precise information can cause agents to reverse their preferences, and then reverse them again, repeatedly, forever. Their conclusion is that, for practical purposes, there is no such thing as ex post.[2]

Meanwhile, in social decisions where all agents share the same probabilistic beliefs, and it is plausible that these beliefs are *correct*, the ex ante Pareto axiom is unproblematic, and even normatively compelling. Consider an insurable risk with a publicly known and well-established loss distribution (e.g., based on extensive actuarial data). If Alice is more risk-averse than Bob, then there are insurance contracts that Bob is willing to sell and Alice is willing to buy. The ex ante Pareto axiom explains why society should endorse such transactions. In this case, the axiom asserts a kind of *nonpaternalism*: it says that if rational agents with correct beliefs can negotiate a mutually beneficial risk-pooling arrangement, then we should approve. A repudiation of ex ante Pareto would make it difficult to explain why insurance markets are socially valuable and should be facilitated, whereas markets for quack medicines or bets on sports events are not. So ex ante Pareto should not be *entirely* rejected, but rather, restricted to cases where it is "appropriate" because agents agree for the "right reasons". This was precisely the justification given by Gilboa et al. (2004) for their restricted Pareto axiom. But as I noted earlier, their particular restriction is not appropriate in decision environments with changing information.

However, there is another important issue, which has not received sufficient attention in the literature on Bayesian social aggregation. Humans are fallible. Not only are their beliefs susceptible to future revision in light of new information, but their expected-utility calculations themselves

[2] All four papers use the same example. But Risse (2001) also states a theorem showing that such examples are generic. The example is formulated in the Bolker-Jeffrey SEU model, where there is no distinction between outcomes and states of nature, and "acts" are just *subsets* of the state/outcome space. To obtain a similar example in the Savage framework, one needs a *sequence* of Savage models, each with its own state space and outcome space, and a rule identifying each "outcome" in model n with an *act* in model $n+1$. The stochastic process framework of the present paper provides one natural way to do this.

could be inaccurate, because of misspecifications in their probabilistic beliefs or an imperfect understanding of the causal relationship between actions and consequences. Such fallibility is especially relevant in complex, multi-period stochastic decision problems. So an individual's preference for one policy over another is more credible if it is *robust*, in the sense that it has a margin of error. Likewise, a unanimous preference in a society is more persuasive if it is a unanimously *robust* preference.

The present paper is a reaction to these concerns. I will consider a model of decision-making under uncertainty in which agents steadily receive more information over time, and update their beliefs and their preferences accordingly. Many contemporary social decision problems have this structure. Three obvious examples are anthropogenic climate change, emerging pandemics, and macroeconomic crises. These are all complex, poorly understood phenomena, unfolding over time. Different agents may have different beliefs about how these phenomena will evolve in the future, either because they assign different values to parameters in their scientific models, or because they use entirely different models. Due to different beliefs and different utility functions, different agents may have different preferences over policies. As time passes and new empirical data arrives (e.g. about weather patterns, infection and mortality rates, GDP trends, etc.), the agents may update their beliefs. They may revise their estimates of model parameters, or even discard certain models altogether in the face of new evidence. Thus, their policy preferences may change over time.

In particular, there might initially be unanimous consensus amongst the individuals that policy A is better than policy B, but this consensus might crumble as the individuals learn new facts about the world. Thus, in retrospect, it would have been a mistake to apply the ex ante Pareto axiom to this ephemeral consensus. At the same time, the individuals might gradually converge on a unanimous consensus that policy C is better than policy D. If this new consensus *persists* over the long term, then it may be a suitable target for the ex ante Pareto axiom.

As earlier noted, the individuals might never obtain *complete* knowledge about the underlying phenomenon. Thus, they might never converge to *perfect* agreement in their beliefs. But there might still be enough belief-convergence to support an enduring consensus that C is better than D. Is such an enduring consensus a sufficient foundation for a Paretian social preference for C over D? Not necessarily, because of the fallibility of individual preferences, discussed above. This enduring consensus would be more compelling if it was built from *enduringly robust* preferences, each having a margin of error.

A unanimous, enduring, and robust preference for C over D provides a cogent Paretian justification for a social preference for C over D. But does it justify an *enduring and robust* social preference for C over D? In light of individual fallibility, perhaps not. A more conservative Pareto principle would simply require a social planner to not *directly contradict* the individuals' robust, enduring consensus for C over D by developing a robust, enduring social preference for D over C instead.

In view of these considerations, I will restrict the Pareto axiom to cases where the individuals not only unanimously prefer one act to another, but these preferences are *robust*, and this unanimity *persists* as the individuals acquire more and more information. This Pareto axiom prohibits the social planner's robust, enduring preferences from directly opposing a robust, enduring consensus of the individuals. I will show that this weak, asymptotic form of the Pareto axiom is necessary and sufficient for the social planner to be a utilitarian. But it does *not* imply that social beliefs are an aggregate of individual beliefs. (I argue that this should be seen as a strength, rather than a weakness; see §4.7.)

To obtain this utilitarian conclusion, I require only a weak compatibility between agents' beliefs. The agents may have heterogeneous beliefs, but there must be some probability distribution (perhaps not representing anyone's beliefs) which is *absolutely continuous* with respect to the beliefs of all agents (including the social planner). Roughly, this means that while agents can disagree about probabilities, there is some agreement about which events are *impossible* (i.e. have probability zero) or *almost-certain* (i.e. have probability one): an event deemed impossible by one agent cannot be deemed almost-certain by another.

The paper is organized as follows. Section 2 introduces some tools from probability theory. Section 3 contains the framework and main result. Section 4 contains further interpretive remarks and conceptual discussions. Appendix A contains the proof of the main result, while Appendix B contains the proofs of other statements made in the paper.

2. Preliminaries

Throughout this paper, let $\mathcal{S}$ be a countable set (i.e. either finite or denumerably infinite). Let $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$, and let $\mathcal{S}^{\mathbb{N}}$ be the set of all $\mathbb{N}$-indexed, infinite sequences $\mathbf{s} = (s_t)_{t=0}^{\infty}$ of elements drawn from $\mathcal{S}$. Endow $\mathcal{S}^{\mathbb{N}}$ with the product sigma-algebra. An *event* is a measurable subset of $\mathcal{S}^{\mathbb{N}}$. Let $\Delta(\mathcal{S}^{\mathbb{N}})$ be the set of all (countably additive) probability measures on this sigma algebra. A measure $\rho \in \Delta(\mathcal{S}^{\mathbb{N}})$ is called a *stochastic process*.

Elements of $\mathcal{S}$ are called *instantaneous states*, and elements of $\mathcal{S}^{\mathbb{N}}$ are called *histories*.[3] For any history $\mathbf{s} \in \mathcal{S}^{\mathbb{N}}$, write $\mathbf{s} = (s_t)_{t=0}^{\infty}$. For any $T \in \mathbb{N}$, let $\mathbf{s}_{[0..T]} := (s_t)_{t=0}^{T}$ (an element of $\mathcal{S}^{[0..T]}$, describing all the information revealed up to and including time T) and let $\mathbf{s}_{(T..\infty)} := (s_t)_{t=T+1}^{\infty}$ (an element of $\mathcal{S}^{(T..\infty)}$, describing all the information that will be revealed after time T).[4] For any $\mathbf{q} \in \mathcal{S}^{[0..T]}$ and $\mathbf{r} \in \mathcal{S}^{(T..\infty)}$, define $(\mathbf{q}, \mathbf{r})$ to be the unique $\mathbf{s} \in \mathcal{S}^{\mathbb{N}}$ such that $\mathbf{s}_{[0..T]} = \mathbf{q}$ and $\mathbf{s}_{(T..\infty)} = \mathbf{r}$. We will interpret a history $\mathbf{s}$ as a flow of information revealed over time, with s_t being the information revealed to all agents at time t. For example, in a macroeconomic decision problem, s_t could be a vector of inflation data, employment data, and other economic indicators observed at time t. In the context of a pandemic, s_t could be a vector of geographically localized rates of infection, transmission, morbidity, mortality, vaccination and other epidemiological data observed at time t. In the context of anthropogenic climate change, s_t could be a vector of meteorological, atmospheric, glaciological and oceanographic data observed at time t. (See §4.4 for further discussion.)

Conditional probabilities Let $T \in \mathbb{N}$. For any $\mathbf{q} \in \mathcal{S}^{[0..T]}$, let $[\mathbf{q}] := \{\mathbf{s} \in \mathcal{S}^{\mathbb{N}};\ \mathbf{s}_{[0..T]} = \mathbf{q}\}$. For any $\mathcal{B} \subseteq \mathcal{S}^{(T..\infty)}$, define $\{\mathbf{q}\} \times \mathcal{B} := \{\mathbf{s} \in \mathcal{S}^{\mathbb{N}};\ \mathbf{s}_{[0..T]} = \mathbf{q}$ and $\mathbf{s}_{(T..\infty)} \in \mathcal{B}\}$. If $\rho[\mathbf{q}] \neq 0$, then we define $\rho_{\mathbf{q},T} \in \Delta(\mathcal{S}^{(T..\infty)})$ as follows: for any event $\mathcal{B} \subseteq \mathcal{S}^{(T..\infty)}$,

$$\rho_{\mathbf{q},T}(\mathcal{B}) \quad := \quad \frac{\rho\left(\{\mathbf{q}\} \times \mathcal{B}\right)}{\rho[\mathbf{q}]}. \tag{1}$$

If $\mathbf{s} \in \mathcal{S}^{\mathbb{N}}$ is a random history drawn from ρ, then $\rho_{\mathbf{q},T}(\mathcal{B})$ is the *conditional probability* that $\mathbf{s}_{(T..\infty)} \in \mathcal{B}$, given $\mathbf{s}_{[0..T]} = \mathbf{q}$. For any $\mathbf{s} \in \mathcal{S}^{\mathbb{N}}$ and $T \in \mathbb{N}$, let $\rho_{\mathbf{s},T} := \rho_{\mathbf{q},T}$, where $\mathbf{q} = \mathbf{s}_{[0..T]}$.

[3] There is unfortunately a slight terminological incompatibility between the jargon of decision theory and that of stochastic processes. Elements of $\mathcal{S}^{\mathbb{N}}$ ("histories") will play the role of *states of nature*, in Savage's terminology. Elements of $\mathcal{S}$ ("states") are best seen as *signals* about these states of nature.

[4] In this paper, for any $N, M \in \mathbb{N}$, the notation "$[N..M]$" denotes $\{N, N+1, \ldots, M\}$, while "$[N..M)$" denotes $\{N, \ldots, M-1\}$ and "$(N..M]$" denotes $\{N+1, \ldots, M\}$. The notation "$(N..\infty)$" is defined similarly.

Coalescent processes For any $t \in \mathbb{N}$, and any $\mathbf{s} \in \mathcal{S}^{\mathbb{N}}$, define ${}^{\vec{t}}\mathbf{s} \in \mathcal{S}^{(t..\infty)}$ by setting ${}^{\vec{t}}s_n := s_{n-t-1}$ for all $n \in (t..\infty)$. This defines a bijection $\mathcal{S}^{\mathbb{N}} \ni \mathbf{s} \mapsto {}^{\vec{t}}\mathbf{s} \in \mathcal{S}^{(t..\infty)}$, which is measurable with respect to the product sigma-algebras on $\mathcal{S}^{\mathbb{N}}$ and $\mathcal{S}^{(t..\infty)}$. For any measurable subset $\mathcal{B} \subseteq \mathcal{S}^{\mathbb{N}}$, define ${}^{\vec{t}}\mathcal{B} := \{{}^{\vec{t}}\mathbf{b};\ \mathbf{b} \in \mathcal{B}\}$; this is a measurable subset of $\mathcal{S}^{(t..\infty)}$. Heuristically, if $\mathcal{B}$ describes a possible future event seen from the perspective of time 0 (e.g. "It will rain in three hours"), then ${}^{\vec{t}}\mathcal{B}$ describes the same event *as seen from time* $t+1$ (e.g. "It will rain three hours after time $t+1$").

Let $(\mathcal{A}, d)$ be a metric space, let $\mathcal{B}, \mathcal{C} \subseteq \mathcal{A}$, and let $\epsilon > 0$. Say that $\mathcal{B}$ is *ϵ-dense* in $\mathcal{C}$ if, for all $c \in \mathcal{C}$ there exists $b \in \mathcal{B}$ with $d(b, c) < \epsilon$. A set is *totally bounded* if, for any $\epsilon > 0$, it has a finite, ϵ-dense subset. (A metric space is compact if and only if it is complete and totally bounded.) I will now introduce a condition on stochastic processes which *roughly* requires the set of conditional probability distributions over the future to be "totally bounded".

Let $\eta \in \Delta(\mathcal{S}^{\mathbb{N}})$ be a stochastic process. For any $\mathbf{s} \in \mathcal{S}^{\mathbb{N}}$ and measurable $\mathcal{B} \subseteq \mathcal{S}^{\mathbb{N}}$, let $\mathcal{C}_{\mathbf{s},\mathcal{B}} \subseteq [0, 1]$ be the set of cluster points of the sequence $\{\eta_{\mathbf{s},t}({}^{\vec{t}}\mathcal{B})\}_{t=1}^{\infty}$. In other words, for any $c \in [0, 1]$, we have $c \in \mathcal{C}_{\mathbf{s},\mathcal{B}}$ if and only if there is an increasing sequence $t_1 < t_2 < t_3 < \cdots$ in $\mathbb{N}$ with $\lim_{n\to\infty} \eta_{\mathbf{s},t_n}({}^{\vec{t_n}}\mathcal{B}) = c$. Let us say that η is *coalescent* if for any $\epsilon > 0$, there is an event $\mathcal{F} \subseteq \mathcal{S}^{\mathbb{N}}$ with $\eta(\mathcal{F}) > 0$, and a finite set $\{\mu_1, \ldots, \mu_N\} \subset \Delta(\mathcal{S}^{\mathbb{N}})$ of nonatomic[5] measures such that, for all $\mathbf{s} \in \mathcal{F}$ and $\mathcal{B} \subseteq \mathcal{S}^{\mathbb{N}}$, the set $\{\mu_1(\mathcal{B}), \ldots, \mu_N(\mathcal{B})\}$ is ϵ-dense in $\mathcal{C}_{\mathbf{s},\mathcal{B}}$.

Coalescence is stronger than the (obvious) statement that the Cartesian product space $\prod_{\mathbf{s}\in\mathcal{F}} \prod_{\mathcal{B}\subseteq\mathcal{S}^{\mathbb{N}}} \mathcal{C}_{\mathbf{s},\mathcal{B}}$ is compact in the Tychonoff topology.[6] Heuristically, it means that for all $\mathbf{s} \in \mathcal{F}$ and $\mathcal{B} \subseteq \mathcal{S}^{\mathbb{N}}$, and any large enough $t \in \mathbb{N}$, the value $\eta_{\mathbf{s},t}({}^{\vec{t}}\mathcal{B})$ can be well-approximated by $\mu_n(\mathcal{B})$ for some $n \in [1..N]$.[7] Let us say η is *fully coalescent* if $\mathcal{F} = \mathcal{S}^{\mathbb{N}}$ for all $\epsilon > 0$. Here are some examples of coalescent processes. (Nonobvious proofs are in Appendix B.)

(i) *i.i.d. processes.* Suppose η describes an $\mathbb{N}$-indexed sequence of independent, identically distributed $\mathcal{S}$-valued random variables. Then $\mathcal{C}_{\mathbf{s},\mathcal{B}} = \{\eta(\mathcal{B})\}$ for all $\mathbf{s} \in \mathcal{S}^{\mathbb{N}}$ and $\mathcal{B} \subseteq \mathcal{S}^{\mathbb{N}}$. Thus, η is fully coalescent.

(ii) *Exchangeable processes and other mixtures.* Suppose $\eta = q\,\eta_1 + (1-q)\,\eta_2$ for some $q \in (0, 1]$ and some $\eta_1, \eta_2 \in \Delta(\mathcal{S}^{\mathbb{N}})$ with disjoint support. If η_1 is coalescent, then it is easily verified that η is also coalescent. In particular, if η is an exchangeable stochastic process, then de Finetti's Theorem says η is a mixture of i.i.d. processes. If one of these i.i.d. processes has nonzero mass in the mixture (in particular, if η is a mixture of a countable collection of i.i.d. processes), then η is coalescent.

(iii) *Markov chains.* Suppose $\mathcal{S}$ is finite, and η is a nonatomic Markov chain with transition probability matrix $\mathbf{P}$.[8] For any $s \in \mathcal{S}$, let μ_s be the Markov chain generated by $\mathbf{P}$ starting from state s.[9] For any $t \in \mathbb{N}$, we have $\eta_{\mathbf{s},t} = \mu_{s_t}$. So $\mathcal{C}_{\mathbf{s},\mathcal{B}} \subseteq \{\mu_s(\mathcal{B})\}_{s\in\mathcal{S}}$ for any $\mathbf{s} \in \mathcal{S}^{\mathbb{N}}$ and any event $\mathcal{B} \subseteq \mathcal{S}^{\mathbb{N}}$. Thus, η is fully coalescent. (For any ϵ, use the finite set $\{\mu_s\}_{s\in\mathcal{S}}$.)

[5] A measure μ is *nonatomic* if there is no $\mathbf{q} \in \mathcal{S}^{\mathbb{N}}$ such that $\mu[\{\mathbf{q}\}] > 0$.

[6] This space is not metrizable, so its compactness cannot be expressed in terms of total boundedness.

[7] This does *not* mean $\eta_{\mathbf{s},t}$ itself can be well-approximated by μ_n, because n might depend on $\mathcal{B}$.

[8] That is: $\mathbf{P} = (p_{r,s})_{r,s\in\mathcal{S}}$ is an $\mathcal{S} \times \mathcal{S}$ matrix of non-negative real numbers such that for all $r \in \mathcal{S}$, the "row vector" $\mathbf{p}_r = (p_{r,s})_{s\in\mathcal{S}}$ is a probability vector (i.e. $\sum_{s\in\mathcal{S}} p_{r,s} = 1$).

[9] i.e., for any $T \in \mathbb{N}$ and $\mathbf{r} \in \mathcal{S}^{[0..T]}$, $\mu_s[\mathbf{r}] = 0$ if $r_0 \neq s$, and $\mu_s[\mathbf{r}] = p_{s,r_1} \cdot p_{r_1,r_2} \cdots p_{r_{T-1},r_T}$ if $r_0 = s$.

(iv) *Hidden Markov chains.* Let $\mathcal{S}$ and $\mathcal{R}$ be countable sets, and let $\phi : \mathcal{S} \longrightarrow \mathcal{R}$. Define $\Phi : \mathcal{S}^{\mathbb{N}} \longrightarrow \mathcal{R}^{\mathbb{N}}$ by $\Phi(s_0, s_1, \ldots) = (\phi(s_0), \phi(s_1), \ldots)$. For any $\eta \in \Delta(\mathcal{S}^{\mathbb{N}})$, let $\Phi(\eta) \in \Delta(\mathcal{R}^{\mathbb{N}})$ be the push-forward of η through Φ.[10] If η is a Markov chain, then $\Phi(\eta)$ is a *hidden Markov chain*. Any nonatomic, finite-state hidden Markov chain is fully coalescent.

More generally, for any $\mathbf{s} \in \mathcal{S}$, any event $\mathcal{B} \subseteq \mathcal{S}^{\mathbb{N}}$, and any $t \in \mathbb{N}$, let $\psi_t^{\mathcal{B}}(\mathbf{s}) := \inf_{c \in \mathcal{C}_{\mathbf{s},\mathcal{B}}} \left| \eta_{\mathbf{s},t}({}^{\vec{t}}\mathcal{B}) - c \right|$. Then $\lim_{t\to\infty} \psi_t^{\mathcal{B}}(\mathbf{s}) = 0$, by the definition of $\mathcal{C}_{\mathbf{s},\mathcal{B}}$. Let us say that a stochastic process η is *uniformly coalescent* if η is coalescent and furthermore the sequence of functions $\{\psi_t^{\mathcal{B}}\}_{t=1}^{\infty}$ converges *uniformly* to zero, for every event $\mathcal{B} \subseteq \mathcal{S}^{\mathbb{N}}$. (For example, i.i.d. processes and Markov chains are uniformly coalescent; indeed, in these cases, $\psi_t^{\mathcal{B}} = 0$ for all t and $\mathcal{B}$.) If η is fully and uniformly coalescent, then so is $\Phi(\eta)$.

(v) *Quasimarkovian processes.* Let $\mathcal{S}^* := \bigcup_{N=1}^{\infty} \mathcal{S}^N$. For any $N < M \leqslant L \leqslant \infty$, and any $\mathbf{s} \in \mathcal{S}^{[0..L]}$, let $\mathbf{s}_{(N..M]} := (s_{N+1}, s_{N+2}, \ldots, s_{M-1}, s_M)$; treat this as an element of $\mathcal{S}^{M-N}$ —and hence an element of $\mathcal{S}^*$ —in the obvious way. Say that a stochastic process η is *quasimarkovian* if there is a function $\mu : \mathcal{S}^* \longrightarrow \Delta(\mathcal{S}^{\mathbb{N}})$ with the following property: for any $\epsilon > 0$, there exist $M \in \mathbb{N}$ and an event $\mathcal{F} \subseteq \mathcal{S}^{\mathbb{N}}$ with $\eta(\mathcal{F}) > 0$, such that

$$\limsup_{t \geqslant M,\ t\to\infty} \left| \eta_{\mathbf{s},t}({}^{\vec{t}}\mathcal{B}) - \mu\left(\mathbf{s}_{(t-M\,..t]}\right)(\mathcal{B}) \right| \leqslant \epsilon, \quad \text{for all } \mathbf{s} \in \mathcal{F} \text{ and all events } \mathcal{B} \subseteq \mathcal{S}^{\mathbb{N}}. \tag{2}$$

In other words, for any $\mathbf{s} \in \mathcal{F}$ and event $\mathcal{B} \subseteq \mathcal{S}^{\mathbb{N}}$, there is some $T_{\mathbf{s},\mathcal{B}} \geqslant M$ such that $\left| \eta_{\mathbf{s},t}({}^{\vec{t}}\mathcal{B}) - \mu(\mathbf{s}_{(t-M\,..t]})(\mathcal{B}) \right| \leqslant \epsilon$ for all $t \geqslant T_{\mathbf{s},\mathcal{B}}$. Thus, at time t, given information about the "recent past" $(t-M\,..\,t]$, we can use μ to estimate the conditional probability of ${}^{\vec{t}}\mathcal{B}$ with ϵ-precision, *without* any information about what happened before time $t-M$. I will refer to μ as a *Markov function*. For instance, any Markov chain is quasimarkovian.[11] If $\mathcal{S}$ is finite, η is quasimarkovian, and $\mu(\mathbf{s})$ is nonatomic for all $\mathbf{s} \in \mathcal{S}^*$, then η is coalescent.

3. Framework and main result

Let $\mathcal{X}$ be a measurable space; let us refer to elements of $\mathcal{X}$ as *outcomes*. Let $\mathcal{A}$ be the set of all measurable functions from $\mathcal{S}^{\mathbb{N}}$ to $\mathcal{X}$ which take a finite number of distinct values; let us call these functions *acts*.[12] An element of $\mathcal{A}$ can be interpreted as a public policy (e.g. a fiscal stimulus plan, a carbon tax system, a vaccination program). Elements of $\mathcal{X}$ can be seen as possible *long-term consequences* of these policies. (See §4.5 for further discussion.)

Let $\succeq$ be a preference order on $\mathcal{A}$. Say that $\succeq$ has an *SEU representation* if there is probability measure ρ in $\Delta(\mathcal{S}^{\mathbb{N}})$ and a bounded, measurable function $u : \mathcal{X} \longrightarrow \mathbb{R}$ such that,

$$\text{for all } \alpha, \beta \in \mathcal{A}, \quad \left(\alpha \succeq \beta\right) \iff \left(\int_{\mathcal{S}^{\mathbb{N}}} u \circ \alpha \, \mathrm{d}\rho \geqslant \int_{\mathcal{S}^{\mathbb{N}}} u \circ \beta \, \mathrm{d}\rho \right). \tag{3}$$

[10] Formally: for any event $\mathcal{W} \subseteq \mathcal{R}^{\mathbb{N}}$, we define $\Phi(\eta)[\mathcal{W}] := \eta\left[\Phi^{-1}(\mathcal{W})\right]$.

[11] In fact, for a Markov chain, one can reduce μ to a function $\mu : \mathcal{S} \longrightarrow \Delta(\mathcal{S}^{\mathbb{N}})$, and inequality (2) is satisfied with $\epsilon = 0$, $M = 1$ and $\mathcal{F} = \mathcal{S}^{\mathbb{N}}$, without taking the limsup as $t \to \infty$.

[12] The main theorem is also true (with exactly the same proof) if we instead define $\mathcal{A}$ to be the set of *all* measurable functions from $\mathcal{S}^{\mathbb{N}}$ to $\mathcal{X}$. Restricting $\mathcal{A}$ to finitely-valued acts broadens the scope of the result: it emphasizes that the proof does not *require* preferences to be defined on any larger domain of acts.

Robust conditional preferences As already noted, for any $t \in \mathbb{N}$ there is an isomorphism between $\mathcal{S}^{\mathbb{N}}$ and $\mathcal{S}^{(t..\infty)}$. Thus, any function $\alpha : \mathcal{S}^{\mathbb{N}} \longrightarrow \mathcal{X}$ can be transformed into a function $\vec{{}^t\alpha} : \mathcal{S}^{(t..\infty)} \longrightarrow \mathcal{X}$ by defining $\vec{{}^t\alpha}(\vec{{}^t\mathbf{s}}) := \alpha(\mathbf{s})$ for all $\mathbf{s} \in \mathcal{S}^{\mathbb{N}}$. If α represents an action that one could execute at time zero, then $\vec{{}^t\alpha}$ represents executing the action α at time $t+1$.

For any $\mathbf{s} \in \mathcal{S}^{\mathbb{N}}$ and $t \in \mathbb{N}$, let $\succeq_{\mathbf{s},t}$ be the preference order on $\mathcal{A}$ defined as follows:

$$\text{For all } \alpha, \beta \in \mathcal{A}, \quad \left(\alpha \succeq_{\mathbf{s},t} \beta\right) \iff \left(\int_{\mathcal{S}^{(t..\infty)}} u \circ \vec{{}^t\alpha} \, \mathrm{d}\rho_{\mathbf{s},t} \geqslant \int_{\mathcal{S}^{(t..\infty)}} u \circ \vec{{}^t\beta} \, \mathrm{d}\rho_{\mathbf{s},t} \right). \tag{4}$$

If $\succeq$ has the SEU representation (3), then $\succeq_{\mathbf{s},t}$ is the *conditional preferences* that the agent would have for the same acts *starting at time* $t+1$, once she has already observed history $\mathbf{s}$ up until time t, and updated her beliefs to conditional probabilities via formula (1).

In fact, we will work with a "robust" version of the conditional preferences defined by formula (4). For any $\epsilon > 0$, let ${}_{\epsilon}{\succ}_{\mathbf{s},t}$ be the partial order on $\mathcal{A}$ defined as follows:

$$\text{For all } \alpha, \beta \in \mathcal{A}, \quad \left(\alpha \; {}_{\epsilon}{\succ}_{\mathbf{s},t} \; \beta\right) \iff \left(\int_{\mathcal{S}^{(t..\infty)}} u \circ \vec{{}^t\alpha} \, \mathrm{d}\rho_{\mathbf{s},t} \; > \; \epsilon + \int_{\mathcal{S}^{(t..\infty)}} u \circ \vec{{}^t\beta} \, \mathrm{d}\rho_{\mathbf{s},t} \right). \tag{5}$$

This means that the agent's conditional preference for α over β is "ϵ-robust", in the sense that there is an ϵ-sized margin of error in the expected utility advantage of α over β. As explained in Section 1, this guards against small errors in the initial specification of ρ, in the calculation of the updated beliefs $\rho_{\mathbf{s},t}$, or in the calculation of $u \circ \alpha$ and $u \circ \beta$.

Eventual preferences Let $\mathcal{H} \subseteq \mathcal{S}^{\mathbb{N}}$ be a measurable set with $\rho(\mathcal{H}) > 0$. Let us define a partial order $\succ_{\mathcal{H}}$ on $\mathcal{A}$ as follows: for any $\alpha, \beta \in \mathcal{A}$,

$$\left(\alpha \succ_{\mathcal{H}} \beta\right) \iff \left(\begin{array}{l} \text{There exists } \epsilon > 0 \text{ such that, for all } \mathbf{s} \in \mathcal{H}, \text{ there} \\ \text{is some } T_{\mathbf{s}} \in \mathbb{N} \text{ such that } \alpha \; {}_{\epsilon}{\succ}_{\mathbf{s},t} \; \beta \text{ for all } t \geqslant T_{\mathbf{s}} \end{array} \right). \tag{6}$$

Thus, if the agent observes any history in $\mathcal{H}$ for long enough, then she eventually develops an ϵ-robust conditional preference for α over β, which *persists* from that time onwards. Clearly, $\succ_{\mathcal{H}}$ is transitive and reflexive. But it is not complete; for many $\alpha, \beta \in \mathcal{A}$, it may be that neither $\alpha \succ_{\mathcal{H}} \beta$ nor $\alpha \prec_{\mathcal{H}} \beta$. (See §4.6 for further discussion.)

Statement (6) might seem hard to satisfy, and hard to verify even if it is satisfied. But the next example gives an easily checked condition that implies (6).

Example 1. Suppose ρ is quasimarkovian, with Markov function $\mu : \mathcal{S}^* \longrightarrow \Delta(\mathcal{S}^{\mathbb{N}})$. Let $\alpha, \beta \in \mathcal{A}$, and suppose there exist $\epsilon' > 0$ and $N \in \mathbb{N}$ such that for all $M > N$ and $\mathbf{s} \in \mathcal{S}^M$,[13]

$$\int_{\mathcal{S}^{\mathbb{N}}} u \circ \alpha \, \mathrm{d}\mu_{\mathbf{s}} \quad > \quad \epsilon' + \int_{\mathcal{S}^{\mathbb{N}}} u \circ \beta \, \mathrm{d}\mu_{\mathbf{s}}. \tag{7}$$

Then there is an event $\mathcal{H} \subseteq \mathcal{S}^{\mathbb{N}}$ with $\rho(\mathcal{H}) > 0$ such that $\alpha \succ_{\mathcal{H}} \beta$. (See Appendix B.) ◇

[13] For clarity in inequality (7), I write $\mu(\mathbf{s})$ as $\mu_{\mathbf{s}}$.

If $\mathcal{X}$ is a metric space, then one can reformulate definitions (5) and (6) in terms of stability under small perturbations of α in the uniform metric on $\mathcal{A}$, without any mention of expected utility. But there is insufficient space to discuss this in detail here.

Utilitarianism and weak utilitarianism For the rest of this paper, let $\mathcal{I}$ be a finite set of individuals, and let $\{\succeq^i\}_{i\in\mathcal{I}}$ be a set of preference orders on $\mathcal{A}$. Let $\succeq$ be another preference order on $\mathcal{A}$ (representing a social planner). Suppose that $\{\succeq^i\}_{i\in\mathcal{I}}$ and $\succeq$ have SEU representations (3) determined by probabilistic beliefs $\{\rho^i\}_{i\in\mathcal{I}}$ and ρ^0, utility functions $\{u^i\}_{i\in\mathcal{I}}$, and an ex post social welfare function W. Let us say that the SWF W is *weakly utilitarian* if there exist constants $c^i \geqslant 0$ for all $i \in \mathcal{I}$, and a constant $b \in \mathbb{R}$ such that

$$W \quad = \quad b + \sum_{i\in\mathcal{I}} c^i\, u^i. \tag{8}$$

If W is not a constant, then $c^i > 0$ for at least some $i \in \mathcal{I}$. But the definition still allows the possibility that $c^j = 0$ for some other $j \in \mathcal{I}$; in other words, the preferences of some individuals might be ignored. If $c^i > 0$ for *all* $i \in \mathcal{I}$, then let us say that W is *utilitarian*.

Minimal agreement and independent prospects The utility functions $\{u^i\}_{i\in\mathcal{I}}$ satisfy *Minimal Agreement* if there exist $\mu_1, \mu_2 \in \Delta(\mathcal{X})$ such that $\int_{\mathcal{X}} u^i \,\mathrm{d}\mu_1 > \int_{\mathcal{X}} u^i \,\mathrm{d}\mu_2$ for all $i \in \mathcal{I}$. In other words, there is some pair of “objective lotteries” over outcomes, for which all individuals have the same strict preference. This condition or its variations are ubiquitous in the literature on Bayesian social aggregation (see e.g. Danan et al. 2016).[14]

The utility functions $\{u^i\}_{i\in\mathcal{I}}$ satisfy *Independent Prospects* if for all $i \in \mathcal{I}$, there exist $x, y \in \mathcal{X}$ such that $u_i(x) > u_i(y)$ whereas $u_j(x) = u_j(y)$ for all $j \in \mathcal{I} \setminus \{i\}$. This is also a common condition (see e.g. Weymark 1991; Mongin 1998; Danan et al. 2016; Zuber 2016).

Riskless Pareto An act α is *riskless* if it is a constant function. Let us say that $\succeq$ satisfies the Riskless Pareto[15] axiom with respect to $\{\succeq^i\}_{i\in\mathcal{I}}$ if, for any riskless $\alpha, \beta \in \mathcal{A}$,

- If $\alpha \succeq^i \beta$ for all $i \in \mathcal{I}$, then $\alpha \succeq \beta$.
- If, in addition, $\alpha \succ^i \beta$ for some $i \in \mathcal{I}$, then $\alpha \succ \beta$.

Suppose $\{u^i\}_{i\in\mathcal{I}}$ satisfy Independent Prospects. Then it is easy to see that W is utilitarian if and only if it is weakly utilitarian and $\succeq$ satisfies Riskless Pareto with respect to $\{\succeq^i\}_{i\in\mathcal{I}}$. Therefore, the main focus of this article will be on establishing *weak* utilitarianism.

Unanimously non-null sets Let $\mathcal{J} := \mathcal{I} \sqcup \{0\}$. Let $\mathcal{H} \subseteq \mathcal{S}^{\mathbb{N}}$ be an event. It may happen that all agents agree that $\mathcal{H}$ has positive probability, but this agreement is “spurious”, because $\mathcal{H}$ is a disjoint union of several subsets, every one of which is deemed to be null by at least one agent. To rule out such a scenario, let say that $\mathcal{H}$ is *unanimously non-null* if there is a measure $\eta \in \Delta(\mathcal{S}^{\mathbb{N}})$ such that $\eta(\mathcal{H}) > 0$ and $\eta \ll \rho^j$ for all $j \in \mathcal{J}$ (hence, $\rho^j(\mathcal{H}) > 0$ for all $j \in \mathcal{J}$). Heuristically, η is

[14] Minimal Agreement is logically weaker than *Minimal Agreement on Consequences* (MAC), another common condition in the literature, which posits outcomes $x, y \in \mathcal{X}$ such that $u^i(x) > u^i(y)$ for all $i \in \mathcal{I}$ (see e.g. Mongin 1995, 1998; Alon and Gayer 2016).

[15] This is often called *ex post Pareto*. But in light of the remarks in Section 1, I eschew the term ex post.

a "weak consensus belief", whereby the agents can all agree that $\mathcal{H}$ is non-null, and furthermore ensure that this is not a "spurious" agreement.[16] We do not assume that $\{\rho^j\}_{j\in\mathcal{J}}$ are mutually absolutely continuous. But if they were, then it would be sufficient to require that $\rho^j(\mathcal{H}) > 0$ for some (hence, all) $j \in \mathcal{J}$.

Eventual Pareto The social preference $\succeq$ satisfies the Eventual Pareto axiom with respect to $\{\succeq^i\}_{i\in\mathcal{I}}$ if, for any unanimously non-null event $\mathcal{H} \subseteq \mathcal{S}^{\mathbb{N}}$, and any $\alpha, \beta \in \mathcal{A}$

$$\left(\alpha \succ^i_{\mathcal{H}} \beta \text{ for all } i \in \mathcal{I}\right) \Longrightarrow \left(\alpha \not\prec_{\mathcal{H}} \beta\right). \tag{9}$$

In other words, if all the individuals eventually agree on a robust conditional preference for α over β after observing any history in $\mathcal{H}$ for long enough, then the social planner cannot directly *oppose* this long-term consensus by developing a robust conditional preference for β over α after observing any history in $\mathcal{H}$ for long enough. This weak axiom does *not* require $\alpha \succ_{\mathcal{H}} \beta$. So the social planner does *not* need to eventually develop a conditional preference for α over β (even a non-robust one), after observing even *some* histories in $\mathcal{H}$ for long enough —it is enough that she does not develop the opposite preference.

Clearly, the Eventual Pareto axiom is binding only insofar as there exist α, β, and unanimously non-null $\mathcal{H}$ such that $\alpha \succ^i_{\mathcal{H}} \beta$ for all $i \in \mathcal{I}$. For example, if $\{\rho^j\}_{j\in\mathcal{J}}$ have disjoint support, then there are *no* unanimously non-null subsets; in this case, Eventual Pareto is vacuously satisfied. In the main result, at least one unanimously non-null event exists by the hypothesis of *concordance*.

Concordant preferences Let us say the beliefs $\{\rho^j\}_{j\in\mathcal{J}}$ are *concordant* if there is a nonatomic coalescent stochastic process η on $\mathcal{S}^{\mathbb{N}}$ that is absolutely continuous with respect to ρ^j for all $j \in \mathcal{J}$. This is a weak form of agreement between the beliefs of different agents.

Example 2. Let $\{\eta_n\}_{n=1}^N$ be a set of stochastic processes on $\mathcal{S}^{\mathbb{N}}$, and suppose η_1 is nonatomic and coalescent (e.g. a nonatomic Markov chain). For all $j \in \mathcal{J}$, suppose $\rho^j = \sum_{n=1}^N c_n^j \, \eta_n$ for some positive constants $\{c_n^j\}_{n=1}^N$ with $\sum_{n=1}^N c_n^j = 1$. Then the collection $\{\rho^j\}_{j\in\mathcal{J}}$ is concordant. To see this, note that $\eta_1 \ll \rho^j$ for all $j \in \mathcal{J}$ (because $c_1^j > 0$).

Heuristically, this collection of beliefs describes a system whose stochastic evolution is poorly understood. One of the processes $\{\eta_n\}_{n=1}^N$ is the *correct* model of the system, but we do not know which one it is. The coefficients $\{c_n^j\}_{n=1}^N$ describe the subjective probabilities that agent j (initially) assign to the different models.

In particular, such a collection of beliefs could arise through *deliberation*, as follows. Suppose $\{\eta_n\}_{n\in\mathcal{J}}$ are the original beliefs of the agents before deliberation, while $\{\rho^j\}_{j\in\mathcal{J}}$ are their beliefs after deliberation. During deliberation, agents learn about each other's beliefs. Each agent j might remain confident in her own beliefs, while acknowledging the possibility that she could be wrong and someone else might be correct. She can represent this by setting $\rho^j = \sum_{n\in\mathcal{J}} c_n^j \, \eta_n$, with $c_n^j > 0$ for all $n \in \mathcal{J}$. (Presumably c_j^j would be close to 1.) Suppose there is some $n \in \mathcal{J}$ such that η_n is coalescent (or η_n is a convex combination of measures, *one* of which is coalescent). Then $\{\rho^j\}_{j\in\mathcal{J}}$ is concordant. ◇

[16] See Proposition B.1 in Appendix B for a precise formulation of this statement.

We now come to the main result.

Theorem. *Let $\succeq$ and $\{\succeq^i\}_{i\in\mathcal{I}}$ be preference orders on $\mathcal{A}$ with SEU representations given by utility functions W and $\{u^i\}_{i\in\mathcal{I}}$ and a concordant collection of probability measures ρ^0 and $\{\rho^i\}_{i\in\mathcal{I}}$, and suppose $\{u^i\}_{i\in\mathcal{I}}$ satisfy Minimal Agreement. Then $\succeq$ satisfies* Eventual Pareto *with respect to $\{\succeq^i\}_{i\in\mathcal{I}}$ if and only if W is weakly utilitarian.*

4. Discussion

This section discusses the interpretation of the theorem, some conceptual issues, and some key elements in the proof. (A sketch of the proof also appears in Appendix A.)

4.1. Spurious unanimity vs. asymptotic agreement

The theorem might seem surprising in light of Mongin's (1995) impossibility theorem. Couldn't the individuals converge to a "spuriously unanimous" preference for one act over another in the long run, despite maintaining different conditional beliefs? In such a scenario, Eventual Pareto would behave like the ex ante Pareto axiom from which Mongin derived a contradiction.

But in fact, this does *not* occur, because of concordance. If the agents' beliefs are concordant, then in the long run, their conditional beliefs must become very similar. To illustrate this heuristically, let us reconsider Example 2, but now suppose that $\{\eta_n\}_{n=1}^N$ are ergodic Markov chains.[17] Suppose η_n is generated by the transition probability matrix $\mathbf{P}_n$. For any $n \in [1..N]$ and $s \in \mathcal{S}$, let η_n^s denote the (nonstationary) Markov chain generated by $\mathbf{P}_n$ starting from state s (see Footnote 9). For any $\mathbf{s} \in \mathcal{S}^{\mathbb{N}}$ and $t \in \mathbb{N}$, note that $\eta_{n;\mathbf{s},t} = \eta_n^{s_t}$. From this, for any $j \in \mathcal{J}$, it can be verified that

$$\rho_{\mathbf{s},t}^j \quad = \quad \sum_{n=1}^N c_{n;\mathbf{s},t}^j \, \eta_n^{s_t}, \tag{10}$$

for some positive coefficients $\{c_{n;\mathbf{s},t}^j\}_{n=1}^N$ summing to one. Intuitively, $\{\eta_n\}_{n=1}^N$ are the different "hypotheses" considered by agent j, and as she learns more about $\mathbf{s}$, she adjusts the probabilities $\{c_{n;\mathbf{s},t}^j\}_{n=1}^N$ that she assigns to these hypotheses. Now suppose $\mathbf{s}$ is a random history drawn from one of the processes $\{\eta_n\}_{n=1}^N$, but j doesn't know which one. By observing a long enough initial segment of $\mathbf{s}$, she can, with very high probability, determine which of the ergodic processes $\{\eta_n\}_{n=1}^N$ generated $\mathbf{s}$.[18] Thus, if $\mathbf{s}$ is drawn from η_m, then $\lim_{t\to\infty} c_{m;\mathbf{s},t}^j = 1$ while $\lim_{t\to\infty} c_{n;\mathbf{s},t}^j = 0$ for all $n \neq m$. Thus, formula (10) yields $\lim_{t\to\infty} \left\| \rho_{\mathbf{s},t}^j - \eta_m^{s_t} \right\| = 0$. This holds for all $j \in \mathcal{J}$.

[17] A stationary Markov chain defined by a matrix $\mathbf{P}$ is *ergodic* if there is some $T \in \mathbb{N}$ such that all entries of $\mathbf{P}^T$ are nonzero. Thus, any state can be reached from any other state any time after T steps, with positive probability. In particular, any stationary Markov chain with full support is ergodic.

[18] This is a consequence of the Birkhoff Ergodic Theorem: if $\mathbf{s}$ is randomly generated by the ergodic process η_n, then for any $\mathbf{r} \in \mathcal{S}^2$, we have $\eta_n[\mathbf{r}] = \lim_{T\to\infty} \frac{1}{T} \#\{t \in [0..T) \, ; \, (s_t, s_{t+1}) = \mathbf{r}\}$, with η_n-probability 1 (see e.g. Petersen 1989, Thm 2.2.3, p. 30). So agent j can obtain an arbitrarily accurate estimate of η_n by looking at a sufficiently long initial segment of $\mathbf{s}$.

So although the agents might never *exactly* agree, the disagreements between their beliefs will become arbitrarily small in the long run. The formal statement of this is a celebrated result of Blackwell and Dubins (1962), which appears in Appendix A as Lemma A.1. So in the long run, it is not possible to sustain the sort of spurious unanimity which underlies Mongin's impossibility theorem.[19]

4.2. Public vs. private information

This argument assumes that all agents update their beliefs only with information from a common information source. This is crucial. To see what can go wrong otherwise, suppose that $\mathcal{S} = \mathcal{Q} \times \mathcal{R}$ for some finite nonsingleton sets $\mathcal{Q}$ and $\mathcal{R}$, so that $\mathcal{S}^{\mathbb{N}} \cong \mathcal{Q}^{\mathbb{N}} \times \mathcal{R}^{\mathbb{N}}$. Thus, any history in $\mathcal{S}^{\mathbb{N}}$ can be written as an ordered pair $(\mathbf{q}, \mathbf{r})$, where $\mathbf{q} \in \mathcal{Q}^{\mathbb{N}}$ and $\mathbf{r} \in \mathcal{R}^{\mathbb{N}}$. Suppose that there are two types of agents: *Type Q* and *Type R*. Type Q agents only observe $\mathbf{q}$, while type R agents only observe $\mathbf{r}$.

For any $\rho_Q \in \Delta(\mathcal{Q}^{\mathbb{N}})$ and $\rho_R \in \Delta(\mathcal{R}^{\mathbb{N}})$, there is a product measure $\rho_Q \otimes \rho_R \in \Delta(\mathcal{S}^{\mathbb{N}})$. Suppose all agents have beliefs of this kind. So each agent's belief has two independent components: a belief about the process on $\mathcal{Q}^{\mathbb{N}}$, and a belief about the process on $\mathcal{R}^{\mathbb{N}}$. As explained in §4.1, under certain assumptions, all Q-type agents will eventually converge to approximately the same beliefs about the $\mathcal{Q}^{\mathbb{N}}$-process. But their beliefs about the $\mathcal{R}^{\mathbb{N}}$-process need not ever converge. For R-type agents, the reverse is true. Thus, even in the long run, the agents can have very different beliefs, so spurious unanimity remains possible.

This does not mean that the main result of this paper is undermined by the presence of *any* private information. The main result just needs *some* common information source that is shared by all agents, and is independent of their private information sources. To see this, suppose that $\mathcal{S} = \mathcal{P} \times \mathcal{Q} \times \mathcal{R}$ for some finite nonsingleton sets $\mathcal{P}$, $\mathcal{Q}$ and $\mathcal{R}$, so that $\mathcal{S}^{\mathbb{N}} \cong \mathcal{P}^{\mathbb{N}} \times \mathcal{Q}^{\mathbb{N}} \times \mathcal{R}^{\mathbb{N}}$ and any history in $\mathcal{S}^{\mathbb{N}}$ can be written as a triple $(\mathbf{p}, \mathbf{q}, \mathbf{r})$, with $\mathbf{p} \in \mathcal{P}^{\mathbb{N}}$, $\mathbf{q} \in \mathcal{Q}^{\mathbb{N}}$ and $\mathbf{r} \in \mathcal{R}^{\mathbb{N}}$. Suppose that type Q agents only observe $(\mathbf{p}, \mathbf{q})$, and type R agents only observe $(\mathbf{p}, \mathbf{r})$. Thus, $\mathbf{q}$ and $\mathbf{r}$ are two sources of private information, while $\mathbf{p}$ is a public information source. Suppose that each agent's belief is a product measure $\rho_P \otimes \rho_Q \otimes \rho_R$, for some $\rho_P \in \Delta(\mathcal{P}^{\mathbb{N}})$, $\rho_Q \in \Delta(\mathcal{Q}^{\mathbb{N}})$ and $\rho_R \in \Delta(\mathcal{R}^{\mathbb{N}})$. Then all agents will eventually converge to approximately the same beliefs about the $\mathcal{P}^{\mathbb{N}}$-process. Thus, if we restrict Eventual Pareto to acts that depend only on $\mathcal{P}^{\mathbb{N}}$, then the main result of this paper implies that W is weakly utilitarian.

Also, the agents' initial beliefs $\{\rho^j\}_{j \in \mathcal{J}}$ might themselves be the result of updating on different private information, which was observed "primordially", prior to time zero. What is important is that their beliefs *after* time zero are updated only on the common information source. The main result of this paper remains valid even in the presence of heterogeneous primordial private information, as long as the initial beliefs are concordant.

4.3. Concordance

Concordance does not require $\{\rho^j\}_{j \in \mathcal{J}}$ themselves to be nonatomic or mutually absolutely continuous. It just says there is some coalescent process η that is "minimally compatible" with the beliefs of all agents. As explained in §4.1, this plays a key role in the proof, by inducing agents to η-almost-surely converge in their beliefs in the long term, thereby extinguishing spurious

[19] This argument works for any ergodic processes. I focused here on Markov chains only for simplicity.

unanimity. For this, we need some $\eta \in \Delta(\mathcal{S}^{\mathbb{N}})$ such that $\eta \ll \rho^j$ for all $j \in \mathcal{J}$. Blackwell and Dubins (1962) showed that the existence of such an η is sufficient for convergence of beliefs, while Kalai and Lehrer (1994) showed it is necessary.[20] Under what circumstances does such an η exist?

Recall that two probability measures ρ^1 and ρ^2 are *singular* if there exist disjoint measurable sets $\mathcal{B}_1, \mathcal{B}_2 \subseteq \mathcal{S}^{\mathbb{N}}$ with $\mathcal{S}^{\mathbb{N}} = \mathcal{B}_1 \sqcup \mathcal{B}_2$ such that $\rho^1(\mathcal{B}_1) = 0$ and $\rho^2(\mathcal{B}_2) = 0$. More generally, let us say that a collection $\{\rho^j\}_{j \in \mathcal{J}}$ is *singular* if there is a partition of $\mathcal{S}^{\mathbb{N}}$ into disjoint measurable sets $\{\mathcal{B}_j\}_{j \in \mathcal{J}}$ such that $\rho^j(\mathcal{B}_j) = 0$ for all $j \in \mathcal{J}$. So no set in this partition is deemed non-null by all agents. Under such conditions, we would not expect even the minimal level of agreement needed for concordance —or even for nonvacuity of Eventual Pareto itself. And indeed, there exists $\eta \in \Delta(\mathcal{S}^{\mathbb{N}})$ such that $\eta \ll \rho^j$ for all $j \in \mathcal{J}$ if and only if the collection $\{\rho^j\}_{j \in \mathcal{J}}$ is *not* singular (see Proposition B.1 in Appendix B). It is a separate question whether η is coalescent. But as explained in Section 2, coalescence is a fairly mild property, satisfied by many important families of stochastic processes.

4.4. The interpretation of stochastic processes

As explained at the start of Section 2, a stochastic process represents "information revealed over time". This could be construed in two ways.[21] In one interpretation, the world evolves over time, and each history $\mathbf{s}$ in $\mathcal{S}^{\mathbb{N}}$ describes one possible path that its future evolution could take; thus, s_t describes the observable state of the world at time t. (The examples in Section 2 concerned evolving economies, pandemics, and climate systems.) In another interpretation, the state of the world is fixed and already determined at time zero. But the agents' *information* about this state gradually expands over time. In this case, s_t describes the new information that the agents discover at time t. For example, geological surveys progressively uncover new mineral reserves in the Earth's crust. Analyses of ice cores and marine sediments reveal paleoclimatological data, informing our models of future anthropogenic climate change. Genomics gradually reveals genes linked with certain diseases, yielding new possibilities for diagnosis and treatment. Metagenomics can even discover entirely new species. More abstractly, the growth of human scientific knowledge can be seen as a process of this type.

In the first interpretation, a stochastic process describes *evolution*, while in the second it describes *discovery*. The model in this paper is compatible with both interpretations. Most of the concrete examples that appeared in Sections 2 and 3 seem to fit more naturally with the first interpretation. But the first interpretation itself can be seen as a special case of the second interpretation: watching a system evolve over time is just discovering what the system will do next.[22] Also, as explained in the discussion of equation (10) in §4.1, Bayesian updating from observations of an evolving system can be seen as a form of scientific hypothesis testing, with eventual convergence on the correct hypothesis.

[20] However, Diaconis and Freedman (1986), Kalai and Lehrer (1994) and Lehrer and Smorodinsky (1996a) have demonstrated weaker forms of belief-convergence without absolute continuity.

[21] I thank a referee for emphasizing the importance of this distinction.

[22] This might seem to assume *deterministic* evolution. But as a purely mathematical observation, it is equally true for systems whose evolution is genuinely random. In fact, the distinction between "deterministic" and "genuinely random" is less clear-cut than it appears; see e.g. List and Pivato (2015).

4.5. The meaning of acts and the realization of outcomes

Notwithstanding the remarks of §4.4, this paper is *not* a model of intertemporal choice. This is for two reasons. First, the elements of $\mathbb{N}$ represent consecutive moments in time when new information is revealed —but these moments might not be *equally spaced* in time. Second, in the model, an "act" is a function $\alpha : \mathcal{S}^{\mathbb{N}} \longrightarrow \mathcal{X}$, which transforms any history in $\mathcal{S}^{\mathbb{N}}$ into a single outcome in $\mathcal{X}$ — not a time-indexed consumption stream. (See also footnote 29 in §4.8.)

However, this raises a question. For any $\mathbf{s}$ in $\mathcal{S}^{\mathbb{N}}$, the act α yields an outcome $\alpha(\mathbf{s})$ in $\mathcal{X}$. But the history $\mathbf{s}$ is never fully revealed; at time t, only the initial segment $(s_0, \ldots, s_t)$ has been revealed. This suggests that $\alpha(\mathbf{s})$ is never actually realized until "the end of time", which makes it difficult to see how mortal agents could have preferences over acts at all.[23]

However, the outcome $\alpha(\mathbf{s})$ will already be known in the far (but *finite*) future. So for practical purposes, we do not need to wait until "the end of time". To see this, endow the countable set $\mathcal{S}$ with the discrete topology. Then the Tychonoff topology on $\mathcal{S}^{\mathbb{N}}$ makes it a totally disconnected Polish space (Aliprantis and Border, 2006, §3.13-3.14). The product sigma algebra on $\mathcal{S}^{\mathbb{N}}$ is the Borel sigma algebra induced by this topology. Let ρ be a probability measure on $\mathcal{S}^{\mathbb{N}}$ (e.g. the beliefs of some agent), and let α be an act; so $\alpha : \mathcal{S}^{\mathbb{N}} \longrightarrow \mathcal{Y}$ is a measurable function, for some finite subset $\mathcal{Y} \subseteq \mathcal{X}$. Endow $\mathcal{Y}$ with the discrete topology. For any $\epsilon > 0$, Lusin's Theorem yields a compact subset $\mathcal{K}_\epsilon \subseteq \mathcal{S}^{\mathbb{N}}$ with $\rho(\mathcal{K}_\epsilon) > 1 - \epsilon$, such that the restriction of α to $\mathcal{K}_\epsilon$ is continuous. This implies that there is some $T_\epsilon \in \mathbb{N}$ and $\widetilde{\alpha}_\epsilon : \mathcal{S}^{[0..T_\epsilon]} \longrightarrow \mathcal{Y}$ such that $\alpha(\mathbf{s}) = \widetilde{\alpha}_\epsilon(s_0, \ldots, s_{T_\epsilon})$ for all $\mathbf{s} \in \mathcal{K}_\epsilon$. In other words, with arbitrarily high probability, the outcome of α on any history is determined by a finite initial segment of that history. So an agent is almost certain to learn the outcome of any act after a *finite* amount of time, even though she will never learn the entire history.[24]

For a concrete illustration of this surprising claim, let $\mathcal{S} := \{H, T\}$, and suppose that $\rho \in \Delta(\mathcal{S}^{\mathbb{N}})$ is a coin-flipping process, and we are interested in whether we will first see six "heads" in a row, or six "tails" in a row. Let $x, y \in \mathcal{X}$, and define $\alpha : \mathcal{S}^{\mathbb{N}} \longrightarrow \mathcal{X}$ as follows: for any $\mathbf{s} \in \mathcal{S}^{\mathbb{N}}$, $\alpha(\mathbf{s}) = x$ if the block (H,H,H,H,H,H) appears in $\mathbf{s}$ earlier in time than the block (T,T,T,T,T,T), whereas $\alpha(\mathbf{s}) = y$ otherwise. For a random $\mathbf{s}$, the time until the first appearance of (H,H,H,H,H,H) or (T,T,T,T,T,T) is a random variable, *finite* but *unbounded*. The *expected* time is 32. With very high probability, it is less than 1000. But for any $t \in \mathbb{N}$, there is a small but non-zero probability that we must wait until time t to learn the outcome of α. Nevertheless, with probability 1, we *will* learn this outcome after a finite time. While simplistic, this example is in fact entirely typical of the general case.

4.6. The significance of eventual preferences

An agent's conditional preference order at time t concerns acts that could be executed at time $t + 1$; we are interested in the evolution of these preferences as $t \to \infty$. This does not mean

[23] I thank a referee for raising this issue. Its precise nature depends on our interpretation of the stochastic process. In the "evolution" interpretation, the outcome $\alpha(\mathbf{s})$ is not even *determined* until the end of time. In the "discovery" interpretation, the outcome $\alpha(\mathbf{s})$ is already determined at time zero; we just might not *learn* this outcome until the end of time. Either interpretation poses the same problem for decision-making.

[24] This argument actually does not require acts to be finitely valued. If $\mathcal{X}$ itself has a topology, then Lusin's theorem says that with arbitrarily high probability, the agent can estimate the outcome of α with *arbitrarily high precision* by observing a long enough segment of the history. If her utility function $u : \mathcal{X} \longrightarrow \mathbb{R}$ is continuous, she can thereby estimate the utility that α yields in that history.

that agents never make choices. It just means that they can make such choices after waiting an arbitrarily long time, and acquiring an arbitrarily large amount of information. The preference order $\succeq_{\mathbf{s},t}$ in formula (4) answers the question: "If you *had* to commit to a choice at time t, having observed the initial part of the history $\mathbf{s}$, then how would you choose?"

Eventual Pareto is formulated in terms of conditional preferences that are eventually stable under this process of gradual information acquisition. In other words, the axiom concerns the relationship between individual and social preferences *in the long run*. But as Keynes said, "in the long run, we are all dead." We must make decisions today. So what is the relevance of hypothetical social preferences which only apply in the long run?

However, this paper investigates a normative question. To answer such a question, it is perfectly appropriate to consider hypothetical preferences, such as those which obtain in the long run. Furthermore, the main result says that contemplation of these "long run" preferences actually has immediate policy relevance. It says: the only way for social preferences to be consistent with individual preferences "in the long run" (according to Eventual Pareto) is for the SWF to be utilitarian —something with immediate policy implications.

4.7. *Collective beliefs*

Notable by its absence in the main result is any rule for aggregating individual beliefs into a collective belief. This is in contrast to the classic theorem of Gilboa et al. (2004), which characterizes a combination of utilitarianism and linear pooling of beliefs, or the more recent and equally elegant result of Dietrich (2021), which characterizes a combination of utilitarianism and *geometric* pooling of beliefs.

However, this lack of a belief-aggregation rule is intentional and justified. The construction of an ex post social welfare function and the construction of a collective belief are two fundamentally different kinds of problems. The former is an ethical question, while the latter is a doxastic one. There is no reason that these two questions should be answered at the same time, by the same theorem, using the same axioms, or even with the same data. In some cases, it might be reasonable to suppose that there are "objectively correct" probabilities, which the collective belief should track as closely as possible. In other cases, probabilistic beliefs might be "purely subjective", in which case belief aggregation is more analogous to the aggregation of risk attitudes or discount rates.[25] But in either case, the doxastic question is disparate from the ethical one. We might expect an answer to the ethical question that holds under very general conditions, but accept that the answer to the doxastic one will be much more *ad hoc* and situation-specific. Depending on the circumstances, it might be more appropriate to form collective beliefs through linear pooling, geometric pooling, multiplicative pooling, or some other pooling rule (see Genest and Zidek 1986, Clemen and Winkler 2007 and Dietrich and List 2016 for surveys). In some cases, it might be better to adopt a "supra-Bayesian" approach, where the beliefs of the agents are treated as *data*, which a Bayesian social observer uses to update her own beliefs (Morris, 1974). In other cases, it might be best to form collective beliefs using a betting market (Hanson, 2013), or with a voting rule, as in the Condorcet Jury Theorem and its generalizations (see e.g. Pivato 2013, 2017). It might even be appropriate to totally ignore the individuals' beliefs, and instead form social beliefs by consulting an expert committee, totally disjoint from the individuals whose welfare is at stake. Finally, it may sometimes be best to form social beliefs through an im-

[25] I thank a referee for emphasizing this point.

personal algorithm, machine learning or some other statistical analysis. In light of these myriad alternatives, it may be unwise to commit to a particular doxastic procedure at the moment when we answer the ethical question.

But it is important not to overstate the dissociation between individual and collective beliefs, given the hypothesis of concordance. As explained in §4.1, concordance implies that all agents (including the social planner) will eventually converge to approximately the same belief. So in the long run, it doesn't matter exactly how (or even whether) the social beliefs were originally obtained from individual beliefs. On the other hand, concordance requires some minimal compatibility between social and individual beliefs at time zero —for example, they cannot have disjoint support (§4.3). Such compatibility is more plausible if social beliefs were obtained from individual beliefs through some reasonable aggregation procedure. In contrast, if social beliefs were completely divorced from the individual beliefs (e.g. obtained via machine learning), then concordance might be less likely.

Furthermore, while in some decision problems there is a natural way to obtain social beliefs, in other problems there is not. In some contexts, it might be appropriate to apply some belief aggregation rule, but not clear which rule to apply; then a single theorem which specified how to aggregate both utilities *and* beliefs would be quite attractive. The stochastic process setting of this paper demands dynamic rationality (i.e. Bayes-updating of beliefs) from all agents. Dietrich (2021) makes dynamical rationality a linchpin of his approach to Bayesian social aggregation, and derives a geometric pooling rule for beliefs. Is it possible to adapt his result to the stochastic process framework of the present paper?

4.8. The purely ex ante approach

The distinction between ethical and doxastic questions in §4.7 assumes that the utility functions $\{u^i\}_{i\in\mathcal{I}}$ really gauge *well-being*, while the probability measures $\{\rho^j\}_{j\in\mathcal{J}}$ really reflect the agents' *beliefs* about the state of nature. But according to a minimal, behaviourist interpretation of SEU, these objects simply provide a convenient mathematical *representation* of agents' choice behaviour, and have no real psychological significance.[26] This interpretation undermines both the ethical relevance of ex post utilitarian formulae like (8) and the epistemological significance of the belief aggregation rules characterized by Gilboa et al. (2004) and Dietrich (2021).[27]

Furthermore, the issue of belief aggregation (§4.7) only arises in the first place because we assumed that the social planner, like the individuals, is an SEU maximizer. This assumption makes sense if the "social planner" is a real agent (e.g. a benevolent government). But in some contexts, the "social planner" is just a loose metaphor for public policy; then there is no reason to ascribe Bayesian rationality to this "planner". Mongin's (1995) impossibility theorem shows that social Bayesian rationality is incompatible with the ex ante Pareto axiom, posing a dilemma. This paper, like most of the literature, seizes one horn of this dilemma, weakening the ex ante Pareto axiom so as to preserve social rationality. But one may instead seize the other horn, weakening

[26] For instance, if we allow the possibility that agents have *state-dependent* utility functions, then there is no reason to interpret $\{\rho^j\}_{j\in\mathcal{J}}$ as their beliefs; see Baccelli (2017) for a good discussion of this problem.

[27] This is not a purely theoretical question. In discussions of financial market regulation, Posner and Weyl (2013), Gayer et al. (2014), Gilboa et al. (2014), Brunnermeier et al. (2014), and Blume et al. (2018) have argued that one can identify "purely speculative" trades by criteria of "spurious unanimity", and perhaps subject them to more stringent policy scrutiny. But on the basis of the considerations in this paragraph (among others), Duffie (2014) has strongly disputed these arguments.

social rationality to save ex ante Pareto.[28] For examples, see Mongin (1998, Prop. 6), Chambers and Hayashi (2006, Thm. 1), Mongin and Pivato (2020, Thm. 1), Desai et al. (2018, Thm. 4),[29] and Sprumont (2018, 2019).

However, a choice between ex ante Pareto and social SEU *à la* Savage is only forced upon us because of the heterogeneity of agents' beliefs. As explained in §4.1, concordance implies that all agents will eventually converge to approximate agreement. So in the long term, the ex ante and ex post roads may lead to the same destination. Nevertheless, in the *short* term, a dilemma remains. The ex ante approach resolves this by answering the ethical question while obviating the doxastic one. In contrast, the present paper brackets the doxastic question, leaving it to be solved later, by other means.

4.9. Normative relevance

One might also argue that there is no longer any *need* to answer the ethical question. After all, didn't Harsanyi (1955) already give a convincing argument for utilitarianism in cases where risks are quantified with known, objective probabilities? Of course, many social decisions do *not* involve objective probabilities. But even in a social decision with radical uncertainty, for which we have only subjective beliefs (e.g. climate change), one could "augment" the decision problem with some independent source of objective risk (e.g. a fair coin toss). One could then first apply Harsanyi (1955) to agents' preferences over this *objective* risk to fix the ex post SWF as a weighted sum of individual utility functions, and then form social preferences with respect to *subjective* uncertainties using this ex post SWF. To put it another way: one could replace Bayesian social aggregation in a Savage framework with Bayesian social aggregation in an Anscombe-Aumann framework, where Harsanyi's result has some grip.[30] What, then, is the value of another result which simply recapitulates Harsanyi's classic answer to the ethical question?

If things were this simple, then the results of Gilboa et al. (2004) and Dietrich (2021) would face a similar criticism, since both linear pooling and geometric pooling have attractive axiomatic characterizations dating from the 1980s. So both the ethical question and the doxastic one already have well-established solutions in the literature, rendering any new axiomatic characterizations somewhat otiose. But things are *not* this simple, for several reasons. First, "augmentation" with an objective risk is only feasible when the social planner can actively intervene in the decision problem (e.g. by providing a fair coin to toss). The "augmentation" argument is less convincing when we wish to normatively evaluate policies concerning phenomena in which the social planner *cannot* directly intervene (e.g. involving large market institutions, or the far future). Second, proponents of a "subjectivist" or "personalist" account of probability (e.g. de Finetti, Savage)

[28] The first horn of the dilemma is often called the *ex post* approach, while the second is the *ex ante* approach. Raiffa (1968, Ch. 8, §13) calls them the *Group Bayesian* and *Paretian* approaches, while Sprumont (2018) calls them the *Savage* and *Pareto* approaches.

[29] Like the present paper, Desai et al. consider Bayesian social aggregation in a stochastic process. But whereas I consider acts which deliver a single outcome in the far future (cf. §4.5), Desai et al. consider acts described by partially observable Markov decision processes, which generate a history-contingent consumption stream, similar to the model of Kreps and Porteus (1978). In other words, their paper has a model of *learning while acting*. In contrast, the present paper has a model of *learning, then acting*.

[30] Mongin and Pivato (2020, Theorem 2) propose a similar solution, but one where the distinction between "subjective" and "objective" probabilities —and indeed, the fact that the agents have SEU representations at all —emerges endogenously from the representation, rather than being stipulated in advance.

do not believe objective risks even *exist*, so they would say such an augmentation is not even possible. Therefore, rather than augmenting the decision problem with an "artificial" source of objective randomness amenable to a Harsanyi-type argument, it would be better to find a "naturally occurring" phenomenon that is *already present* in the original decision problem, about which all agents will eventually and persistently agree, and apply a Harsanyi-type argument to this phenomenon. That is the strategy of the present paper.

Finally, Harsanyi's original proof depends critically on the ex ante Pareto axiom. This axiom seemed to Harsanyi and his contemporaries to be innocent, even normatively compelling —but Mongin's (1995) result can be seen as its *reductio ad absurdum*. Nevertheless, as I argued in Section 1, ex ante Pareto cannot be *entirely* rejected, because there are settings where it is still plausible or even indispensable, such as insurance markets (see also footnote 27). One could restrict the axiom to apply only to risks with purely *objective* probabilities; as suggested above, this would block Mongin's impossibility theorem, while still leaving room for Harsanyi's social aggregation theorem. But such an etiolated version of ex ante Pareto might be inadequate for a normative analysis of risk-sharing institutions, where objective probabilities are not always available. This was one motivation for the less restrictive Pareto conditions of Gilboa et al. (2004), Gayer et al. (2014) and Gilboa et al. (2014) —but as explained in Section 1, they run into trouble in environments with changing information. Also, an "objective-only" Pareto axiom may be incapable of coping with situations where subjective uncertainty is never entirely resolved, such as those discussed by Risse (2001, 2003) and Hild et al. (2003, 2008). This raises the question: is there a Pareto axiom applicable to environments with perennial and purely subjective uncertainty and changing information, strong enough to support utilitarian conclusions, but weak enough to avoid Mongin-style impossibilities? This paper answers that question.

Appendix A. Proof of the main result

The main result is proved by contradiction. Suppose that W is *not* weakly utilitarian. Then it is not contained in the convex cone spanned by $\{u^i\}_{i\in\mathcal{I}}$ in the Banach space of measurable real-valued functions on $\mathcal{X}$. Thus, the Separating Hyperplane Theorem yields a linear functional separating W from this cone. Using a Riesz-type representation theorem and the Jordan Decomposition Theorem we can convert this functional into two probability measures ν_1 and ν_2 on $\mathcal{X}$ that manifest a strict violation of a "Pareto" type property in terms of the expected-utility preferences defined by $\{u^i\}_{i\in\mathcal{I}}$ and W on probability measures over $\mathcal{X}$ (Lemma A.2). Since this is a *strict* violation, any measures on $\mathcal{X}$ sufficiently close to ν_1 and ν_2 will also strictly violate Pareto. In particular, we can define a partition $\mathfrak{Y}$ on $\mathcal{X}$ such that any pair of probability measures which assign approximately the same weight to all elements of $\mathfrak{Y}$ as do ν_1 and ν_2 will strictly violate Pareto (Claim 1 in the proof).

The goal now is to construct acts α_1 and α_2 that induce probability measures on $\mathcal{X}$ close to ν_1 and ν_2 in this sense. If all agents had the *same* beliefs, this would be easy —but they don't. However, as explained in §4.1, concordance implies that, after updating their beliefs on a long enough initial history, all agents will converge to *approximately* the same beliefs (Lemma A.1). This will happen on a set $\mathcal{G}$ of η-measure 1, where η is the "concordance" measure. This will allow us to construct α_1 and α_2 with the desired properties.

But there is a complication. The Eventual Pareto axiom (9) is formulated in terms of agents' eventual preferences (6), the asymptotically stable part of their robust conditional preferences (5). But their conditional preferences change with every time-step, as they update their beliefs with new information. So although at any moment in time in the far future, all agents will have roughly the *same* conditional beliefs (cf. previous paragraph), these conditional beliefs are a moving target. So α_1 and α_2 must track this moving target.

Now, η is coalescent, and all agents asymptotically have beliefs very close to η. So the conditional beliefs of all agents at any time t in the far future can be "approximated" by a finite collection of measures $\{\mu_1, \ldots, \mu_M\}$. This is only true on a set $\mathcal{F}$ of positive η-measure. But $\eta(\mathcal{G}) = 1$, so if we define $\mathcal{H} = \mathcal{F} \cap \mathcal{G}$, then $\eta(\mathcal{H}) > 0$ also. So $\mathcal{H}$ is unanimously non-null, hence a suitable site for application of the Eventual Pareto axiom. The Dubins-Spanier Theorem yields a measurable partition $\mathfrak{P}^1$ of $\mathcal{S}^{\mathbb{N}}$ such that:

- There is a bijective correspondence between the atoms of $\mathfrak{P}^1$ and the atoms of $\mathfrak{Y}$.
- *All* of the measures $\{\mu_1, \ldots, \mu_M\}$ assign to each atom of $\mathfrak{P}^1$ the same probability that ν_1 assigns to the corresponding atom of $\mathfrak{Y}$ (cf. eqn. (A24)).

We can likewise construct a partition $\mathfrak{P}^2$ of $\mathcal{S}^{\mathbb{N}}$ which "duplicates" the values of ν_2 on $\mathfrak{Y}$. Now define the act α^1 (respectively, α^2) to map each element of $\mathfrak{P}^1$ (resp. $\mathfrak{P}^2$) to the corresponding element of $\mathfrak{Y}$. Thus, at any time in the far future, along any history in $\mathcal{H}$, the subjective expected utilities assigned by the agents to α^1 and α^2 are very well-approximated by the expected values of $\{u^i\}_{i\in\mathcal{I}}$ and W with respect to ν_1 and ν_2 (cf. (A29) and (A30)). But by construction, ν_1 and ν_2 manifest a strict violation of Pareto; this implies that α^1 and α^2 manifest a violation of Pareto at all times in the far future along any history in $\mathcal{H}$, which leads to a violation of Eventual Pareto; hence a contradiction.

To proceed with the proof, we will need two lemmas. Let $t \in \mathbb{N}$. For any $\mu, \nu \in \Delta(\mathcal{S}^{(t..\infty)})$, the *total variation norm distance* between μ and ν is defined:

$$\|\mu - \nu\| \quad := \quad \sup_{\substack{\mathcal{B}\subseteq\mathcal{S}^{(t..\infty)} \\ \text{measurable}}} |\mu(\mathcal{B}) - \nu(\mathcal{B})|\,. \tag{A1}$$

For any measurable function $\phi : \mathcal{S}^{(t..\infty)} \longrightarrow \mathbb{R}$, it is easily verified that

$$\left| \int_{\mathcal{S}^{(t..\infty)}} \phi \, \mathrm{d}\mu - \int_{\mathcal{S}^{(t..\infty)}} \phi \, \mathrm{d}\nu \right| \quad \leqslant \quad \|\phi\|_\infty \cdot \|\mu - \nu\|. \tag{A2}$$

Lemma A.1. (Blackwell and Dubins, 1962) *Let $\mathcal{S}$ be countable, let ρ and η be two stochastic processes on $\mathcal{S}^{\mathbb{N}}$, and suppose that η is absolutely continuous with respect to ρ. Then there is a subset $\mathcal{G} \subseteq \mathcal{S}^{\mathbb{N}}$ such that $\eta(\mathcal{G}) = 1$, and such that for all $\mathbf{s} \in \mathcal{G}$, $\lim_{t\to\infty} \left\| \rho_{\mathbf{s},t} - \eta_{\mathbf{s},t} \right\| = 0$.*

In fact, the Blackwell-Dubins Theorem applies when $\mathcal{S}$ is *any* measurable space. The key requirement is that ρ and η be *predictive* stochastic processes. Roughly speaking, this means that for any $T \in \mathbb{N}$ there is a function $\rho_T : \mathcal{S}^{[0..T]} \longrightarrow \Delta(\mathcal{S}^{(T..\infty)})$ such that for any $\mathbf{q} \in \mathcal{S}^{[0..T]}$, $\rho_T(\mathbf{q})$ plays the role of the *conditional probability* given $\mathbf{q}$. If $\mathcal{S}$ is countable, then *all* stochastic processes on $\mathcal{S}^{\mathbb{N}}$ are predictive: define $\rho_T(\mathbf{q})$ using formula (1). Diaconis and Freedman (1986, 1990), Schervish and Seidenfeld (1990), Kalai and Lehrer (1994) and Lehrer and Smorodinsky (1996a) proved enhancements and variations of the Blackwell-Dubins Theorem; see Lehrer and Smorodinsky (1996b) for a survey of this literature. Interestingly, Miller and Sanchirico (1999) provided an alternative proof of the Blackwell-Dubins Theorem which specifically relies on its role in asymptotically eliminating “spurious unanimity” in zero-sum bets between two players. But they did not connect this to Bayesian social aggregation.

Let $\mathcal{U}$ be the Banach space of bounded, measurable, real-valued functions on $\mathcal{X}$, endowed with the norm $\|\cdot\|$ defined by $\|u\| := \sup_{x\in\mathcal{X}} |u(x)|$ for all $u \in \mathcal{U}$. Recall that $\mathcal{J} := \mathcal{I} \sqcup \{0\}$.

Lemma A.2. *Let $\{u^j\}_{j\in\mathcal{J}} \subset \mathcal{U}$, and suppose there is some $z \in \mathcal{X}$ such that $u^j(z) = 0$ for all $j \in \mathcal{J}$. Suppose that $\{u^i\}_{i\in\mathcal{I}}$ satisfy Minimal Agreement. Let $\mathcal{C}$ be the closed, convex cone in $\mathcal{U}$ spanned by $\{u^i\}_{i\in\mathcal{I}}$ and 0. If u^0 is not in $\mathcal{C}$, then there exist finitely additive probability measures ν_1 and ν_2 on $\mathcal{X}$ such that*

$$\int_{\mathcal{X}} u^0 \,\mathrm{d}\nu_1 < \int_{\mathcal{X}} u^0 \,\mathrm{d}\nu_2, \quad \textit{while} \quad \int_{\mathcal{X}} u^i \,\mathrm{d}\nu_1 > \int_{\mathcal{X}} u^i \,\mathrm{d}\nu_2 \quad \textit{for all } i \in \mathcal{I}. \tag{A3}$$

Proof. $\mathcal{C}$ is a closed subset of $\mathcal{U}$. So if $u^0 \notin \mathcal{C}$, then the Separating Hyperplane Theorem yields a continuous linear functional $\phi : \mathcal{U} \longrightarrow \mathbb{R}$ and some constant $R \in \mathbb{R}$ such that

$$\phi(u^0) < R < \phi(c), \qquad \text{for all } c \in \mathcal{C}. \tag{A4}$$

(see e.g. Dunford and Schwartz 1958, Theorem V.2.10, page 417, or Conway 1990, Theorem IV.3.13, p. 111). In particular, since $0 \in \mathcal{C}$, this means that $R < \phi(0) = 0$. Furthermore, we must have $\phi(c) \geqslant 0$ for all $c \in \mathcal{C}$, because if $\phi(c) < 0$ for some $c \in \mathcal{C}$, then $\phi(r\,c) < R$ for sufficiently large $r > 0$, contradicting the fact that $r\,c$ is also in $\mathcal{C}$ (because $\mathcal{C}$ is a cone). In particular, $\phi(u^i) \geqslant 0$ for all $i \in \mathcal{I}$. Now, $\{u^i\}_{i\in\mathcal{I}}$ satisfy Minimal Agreement, so there exist $\mu_1, \mu_2 \in \Delta(\mathcal{X})$ such that $\int_{\mathcal{X}} u^i \,\mathrm{d}\mu_1 > \int_{\mathcal{X}} u^i \,\mathrm{d}\mu_2$ for all $i \in \mathcal{I}$. Define $\psi : \mathcal{U} \longrightarrow \mathbb{R}$ by setting $\psi(u) := \int_{\mathcal{X}} u \,\mathrm{d}\mu_1 - \int_{\mathcal{X}} u \,\mathrm{d}\mu_2$ for all $u \in \mathcal{U}$. Then ψ is a linear functional, and $\psi(u^i) > 0$ for all $i \in \mathcal{I}$. Let $\phi' := \phi + \epsilon\,\psi$, for some small $\epsilon > 0$. Then $\phi'(u^i) > 0$ for all $i \in \mathcal{I}$. If ϵ is sufficiently small, then we still have $\phi'(u^0) < 0$, because $R < 0$ in inequality (A4). Thus, we get a continuous linear functional $\phi' : \mathcal{U} \longrightarrow \mathbb{R}$ such that

$$\phi'(u^0) < 0 < \phi'(u^i), \qquad \text{for all } i \in \mathcal{I}. \tag{A5}$$

The dual space of $\mathcal{U}$ is the space of finitely additive, signed measures on $\mathcal{X}$ (Dunford and Schwartz, 1958, Theorem IV.5.1, p. 258). So there is a finitely additive, signed measure λ on $\mathcal{X}$ such that $\phi'(u) = \int_{\mathcal{X}} u \,\mathrm{d}\lambda$ for all $u \in \mathcal{U}$. So we can rewrite inequality (A5) as:

$$\int_{\mathcal{X}} u^0 \,\mathrm{d}\lambda < 0 < \int_{\mathcal{X}} u^i \,\mathrm{d}\lambda, \qquad \text{for all } i \in \mathcal{I}. \tag{A6}$$

Let δ_z be “point mass” at z —that is, the finitely additive probability measure on $\mathcal{X}$ such that, for all measurable $\mathcal{B} \subseteq \mathcal{X}$, $\delta_z(\mathcal{B}) := 1$ if $z \in \mathcal{B}$, while $\delta_z(\mathcal{B}) := 0$ if $z \notin \mathcal{B}$. Let $L := \lambda(\mathcal{X})$, and define $\lambda' := \lambda - L\,\delta_z$. Then $\lambda'(\mathcal{X}) = 0$. Note that $\int_{\mathcal{X}} u^0 \,\mathrm{d}\lambda' = \int_{\mathcal{X}} u^0 \,\mathrm{d}\lambda$ because $\int_{\mathcal{X}} u^0 \,\mathrm{d}\delta_z = 0$. Likewise, $\int_{\mathcal{X}} u^i \,\mathrm{d}\lambda' = \int_{\mathcal{X}} u^i \,\mathrm{d}\lambda$ for all $i \in \mathcal{I}$. Thus, (A6) yields

$$\int_{\mathcal{X}} u^0 \,\mathrm{d}\lambda' < 0 < \int_{\mathcal{X}} u^i \,\mathrm{d}\lambda', \qquad \text{for all } i \in \mathcal{I}. \tag{A7}$$

The Jordan Decomposition Theorem yields unique positive, finitely additive measures ν_1' and ν_2' on $\mathcal{X}$ such that $\lambda' = \nu_1' - \nu_2'$ (see Dunford and Schwartz 1958, Theorem III.1.8, p. 98, or Bhaskara Rao and Bhaskara Rao 1983, Theorem 2.5.3, p. 53). Furthermore, $\nu_1'(\mathcal{X}) = \nu_2'(\mathcal{X})$, because $\lambda'(\mathcal{X}) = 0$ by construction. Let $H := \nu_1'(\mathcal{X}) = \nu_2'(\mathcal{X})$, and let $\nu_1 := \nu_1'/H$ and $\nu_2 := \nu_2'/H$. Then ν_1 and ν_2 are probability measures, and inequality (A7) yields the inequalities (A3). □

Proof of the Theorem. "$\Longleftarrow$" (by contradiction) Suppose W is weakly utilitarian, but $\succeq$ violates Eventual Pareto. Thus, there exists a measure $\eta \in \Delta(\mathcal{S}^{\mathbb{N}})$ with $\eta \ll \rho^j$ for all $j \in \mathcal{J}$, a measurable subset $\mathcal{H} \subseteq \mathcal{S}^{\mathbb{N}}$ with $\eta(\mathcal{H}) > 0$, and acts $\alpha, \beta \in \mathcal{A}$ violating statement (9) —i.e. such that $\alpha \succ^i_{\mathcal{H}} \beta$ for all $i \in \mathcal{I}$, but $\alpha \prec_{\mathcal{H}} \beta$. For all $j \in \mathcal{J}$, formula (6) yields some $\epsilon^j > 0$, and for all $\mathbf{s} \in \mathcal{H}$, it yields some $T^j_{\mathbf{s}} \in \mathbb{N}$ such that if $j \in \mathcal{I}$, then $\alpha \ {}_{\epsilon^j}\!\succ^j_{\mathbf{s},t} \ \beta$ for all $t \geqslant T^j_{\mathbf{s}}$, whereas $\alpha \ {}_{\epsilon^0}\!\prec_{\mathbf{s},t} \ \beta$ for all $t \geqslant T^0_{\mathbf{s}}$. Let $\epsilon := \min_{j\in\mathcal{J}} \epsilon^j$; then $\epsilon > 0$. For all $\mathbf{s} \in \mathcal{H}$, let $T_{\mathbf{s}} := \max_{j\in\mathcal{J}} T^j_{\mathbf{s}}$; then $T_{\mathbf{s}} \in \mathbb{N}$, and for all $t \geqslant T_{\mathbf{s}}$ and all $i \in \mathcal{I}$ we have $\alpha \ {}_{\epsilon}\!\succ^i_{\mathbf{s},t} \ \beta$, whereas $\alpha \ {}_{\epsilon}\!\prec_{\mathbf{s},t} \ \beta$. Thus, formula (5) yields

$$\int\limits_{\mathcal{S}^{(t..\infty)}} u^i \circ {}^{t}\vec{\alpha} \, \mathrm{d}\rho^i_{\mathbf{s},t} - \int\limits_{\mathcal{S}^{(t..\infty)}} u^i \circ {}^{t}\vec{\beta} \, \mathrm{d}\rho^i_{\mathbf{s},t} > \epsilon \quad \text{for all } i \in \mathcal{I} \text{ and } t \geqslant T_{\mathbf{s}}, \tag{A8}$$

$$\text{while} \int\limits_{\mathcal{S}^{(t..\infty)}} W \circ {}^{t}\vec{\alpha} \, \mathrm{d}\rho^0_{\mathbf{s},t} - \int\limits_{\mathcal{S}^{(t..\infty)}} W \circ {}^{t}\vec{\beta} \, \mathrm{d}\rho^0_{\mathbf{s},t} < -\epsilon, \quad \text{for all } t \geqslant T_{\mathbf{s}}. \tag{A9}$$

For all $j \in \mathcal{J}$, Lemma A.1 yields a measurable $\mathcal{G}_j \subseteq \mathcal{S}^{\mathbb{N}}$ with $\eta(\mathcal{G}_j) = 1$ such that $\lim\limits_{t\to\infty} \left\| \rho^j_{\mathbf{s},t} - \eta_{\mathbf{s},t} \right\| = 0$ for all $\mathbf{s} \in \mathcal{G}_j$. Let $\mathcal{G} := \bigcap_{j\in\mathcal{J}} \mathcal{G}_j$. Then $\eta(\mathcal{G}) = 1$, and for all $j \in \mathcal{J}$ and all $\mathbf{s} \in \mathcal{G}$, we have $\lim\limits_{t\to\infty} \left\| \rho^j_{\mathbf{s},t} - \eta_{\mathbf{s},t} \right\| = 0$. Since $\|u^j\|_\infty < \infty$, inequality (A2) yields

$$\lim_{t\to\infty} \left| \int\limits_{\mathcal{S}^{(t..\infty)}} u^j \circ {}^{t}\vec{\alpha} \, \mathrm{d}\rho^j_{\mathbf{s},t} - \int\limits_{\mathcal{S}^{(t..\infty)}} u^j \circ {}^{t}\vec{\alpha} \, \mathrm{d}\eta_{\mathbf{s},t} \right| = 0 \tag{A10}$$

$$\text{and} \quad \lim_{t\to\infty} \left| \int\limits_{\mathcal{S}^{(t..\infty)}} u^j \circ {}^{t}\vec{\beta} \, \mathrm{d}\rho^j_{\mathbf{s},t} - \int\limits_{\mathcal{S}^{(t..\infty)}} u^j \circ {}^{t}\vec{\beta} \, \mathrm{d}\eta_{\mathbf{s},t} \right| = 0. \tag{A11}$$

Now, $\mathcal{G} \cap \mathcal{H} \neq \emptyset$, because $\eta(\mathcal{G}) = 1$ and $\eta(\mathcal{H}) > 0$. For any $\mathbf{s} \in \mathcal{G} \cap \mathcal{H}$, we can combine inequalities (A8) and (A9) with equations (A10) and (A11) to obtain

$$\liminf_{t\to\infty} \left(\int\limits_{\mathcal{S}^{(t..\infty)}} u^i \circ {}^{t}\vec{\alpha} \, \mathrm{d}\eta_{\mathbf{s},t} - \int\limits_{\mathcal{S}^{(t..\infty)}} u^i \circ {}^{t}\vec{\beta} \, \mathrm{d}\eta_{\mathbf{s},t} \right) > \epsilon, \quad \text{for all } i \in \mathcal{I}, \tag{A12}$$

$$\text{while} \quad \limsup_{t\to\infty} \left(\int\limits_{\mathcal{S}^{(t..\infty)}} W \circ {}^{t}\vec{\alpha} \, \mathrm{d}\eta_{\mathbf{s},t} - \int\limits_{\mathcal{S}^{(t..\infty)}} W \circ {}^{t}\vec{\beta} \, \mathrm{d}\eta_{\mathbf{s},t} \right) < -\epsilon. \tag{A13}$$

But $W = \sum\limits_{i\in\mathcal{I}} c^i \, u^i$, where $c^i \geqslant 0$ for all $i \in \mathcal{I}$. So (A12) and (A13) yield a contradiction.

"$\Longrightarrow$" (by contradiction) Recall that $\mathcal{J} := \mathcal{I} \cup \{0\}$. Let $u^0 := W$. Let $z \in \mathcal{X}$. For all $j \in \mathcal{J}$, by replacing u^j with $u^j - u^j(z)$, we can assume without loss of generality that $u^j(z) = 0$. (This does not affect the SEU representations.) Let $\mathcal{C}$ be the closed, convex cone in $\mathcal{U}$ spanned by $\{u^i\}_{i\in\mathcal{I}}$ and 0. Then u^0 is weakly utilitarian if and only if $u^0 \in \mathcal{C}$.

To get a contradiction, suppose that u^0 is *not* in $\mathcal{C}$. Then Lemma A.2 yields finitely additive probability measures ν_1 and ν_2 on $\mathcal{X}$ satisfying the inequalities (A3). For all $j \in \mathcal{J}$, let $\epsilon^j := \left|\int_{\mathcal{X}} u^j \, \mathrm{d}\nu_1 - \int_{\mathcal{X}} u^j \, \mathrm{d}\nu_2\right|$. Let

$$\epsilon \quad := \quad \frac{1}{5} \min_{j\in\mathcal{J}} \epsilon^j. \tag{A14}$$

Then $\epsilon > 0$. Inequalities (A3) and definition (A14) yield

$$\int_{\mathcal{X}} u^0 \, \mathrm{d}\nu_2 - \int_{\mathcal{X}} u^0 \, \mathrm{d}\nu_1 \;>\; 5\,\epsilon, \quad \text{and} \quad \int_{\mathcal{X}} u^i \, \mathrm{d}\nu_1 - \int_{\mathcal{X}} u^i \, \mathrm{d}\nu_2 \;>\; 5\,\epsilon \;\; \text{for all } i \in \mathcal{I}. \tag{A15}$$

Let $R := \max\left\{\|u^j\|_\infty\right\}_{j\in\mathcal{J}}$. Then R is finite because $\{u^j\}_{j\in\mathcal{J}}$ are bounded functions. Let $N := \lceil R/\epsilon \rceil + 1$; then $N\epsilon > R$, so the interval $[-N\,\epsilon,\ N\,\epsilon)$ contains the ranges of $\{u^j\}_{j\in\mathcal{J}}$. For all $j \in \mathcal{J}$ and all $n \in [-N..N]$, let $\mathcal{Y}^j_n := (u^j)^{-1}[n\,\epsilon, (n{+}1)\,\epsilon)$. Then $\mathfrak{Y}^j := \{\mathcal{Y}^j_n\}_{n=-N}^N$ is a measurable partition of $\mathcal{X}$. Let $\mathfrak{Y}$ be the common refining partition of $\{\mathfrak{Y}^j\}_{j\in\mathcal{J}}$. This is a measurable partition of $\mathcal{X}$. Suppose it has K cells, and write $\mathfrak{Y} = \{\mathcal{Y}_k\}_{k=1}^K$. For all $k \in [1..K]$, let $p^1_k := \nu_1(\mathcal{Y}_k)$ and $p^2_k := \nu_2(\mathcal{Y}_k)$. Then $\mathbf{p}^1 := (p^1_k)_{k=1}^K$ and $\mathbf{p}^2 := (p^2_k)_{k=1}^K$ are K-dimensional probability vectors.

Claim 1: *Fix $\ell \in \{1, 2\}$. Let $\nu \in \Delta(\mathcal{X})$ be any measure such that*

$$\left|\nu(\mathcal{Y}_k) - p^\ell_k\right| \quad < \quad \frac{1}{K\,N}, \qquad \textit{for all } k \in [1..K]. \tag{A16}$$

Then $\left|\int_{\mathcal{X}} u^j \, \mathrm{d}\nu - \int_{\mathcal{X}} u^j \, \mathrm{d}\nu_\ell\right| < 2\,\epsilon$ *for all $j \in \mathcal{J}$.*

Proof. Fix $j \in \mathcal{J}$. For any $k \in [1..K]$, there is some $n \in [-N \dots N]$ such that $\mathcal{Y}_k \subseteq \mathcal{Y}^j_n$. Suppose that $n \in [-N \dots -1]$. (The argument when $n \in [0 \dots N]$ is similar.) Then

$$n\,\epsilon\,\nu(\mathcal{Y}_k) \leqslant \int_{\mathcal{Y}_k} u^j \, \mathrm{d}\nu \quad < \quad (n{+}1)\,\epsilon\,\nu(\mathcal{Y}_k)$$

$$\text{and } n\,\epsilon\,\nu_\ell(\mathcal{Y}_k) \leqslant \int_{\mathcal{Y}_k} u^j \, \mathrm{d}\nu_\ell \quad < \quad (n{+}1)\,\epsilon\,\nu_\ell(\mathcal{Y}_k),$$

because $n\,\epsilon \leqslant u^j(y) < (n{+}1)\,\epsilon$ for all $y \in \mathcal{Y}^j_n$, by definition. Thus,

$$\left|\int_{\mathcal{Y}_k} u^j \, \mathrm{d}\nu - \int_{\mathcal{Y}_k} u^j \, \mathrm{d}\nu_\ell\right| \tag{A17}$$

$$< \max\left\{\left|(n{+}1)\,\epsilon\,\nu_\ell(\mathcal{Y}_k) - n\,\epsilon\,\nu(\mathcal{Y}_k)\right|,\ \left|(n+1)\,\epsilon\,\nu(\mathcal{Y}_k) - n\,\epsilon\,\nu_\ell(\mathcal{Y}_k)\right|\right\}.$$

Now,

$$\begin{aligned}
\Big|(n{+}1)\,\epsilon\,\nu_\ell(\mathcal{Y}_k) - n\,\epsilon\,\nu(\mathcal{Y}_k)\Big| &= \Big|(n{+}1)\,\epsilon\,\Big(\nu_\ell(\mathcal{Y}_k) - \nu(\mathcal{Y}_k)\Big) + \epsilon\,\nu(\mathcal{Y}_k)\Big| \\
&\leqslant |n{+}1|\,\epsilon\,\Big|\nu_\ell(\mathcal{Y}_k) - \nu(\mathcal{Y}_k)\Big| + \epsilon\,\nu(\mathcal{Y}_k) \;\underset{(\diamond)}{\leqslant}\; |n|\,\epsilon\,\Big|\nu_\ell(\mathcal{Y}_k) - \nu(\mathcal{Y}_k)\Big| + \epsilon\,\nu(\mathcal{Y}_k) \\
&\underset{(*)}{\leqslant} |n|\,\epsilon\,\Big|p_k^\ell - \nu(\mathcal{Y}_k)\Big| + \epsilon\,\nu(\mathcal{Y}_k) \;\underset{(\dagger)}{<}\; N\,\epsilon \cdot \frac{1}{K\,N} + \epsilon\,\nu(\mathcal{Y}_k) \;=\; \frac{\epsilon}{K} + \epsilon\,\nu(\mathcal{Y}_k),
\end{aligned} \tag{A18}$$

while

$$\begin{aligned}
\Big|(n{+}1)\,\epsilon\,\nu(\mathcal{Y}_k) - n\,\epsilon\,\nu_\ell(\mathcal{Y}_k)\Big| &= \Big|n\,\epsilon\,\Big(\nu(\mathcal{Y}_k) - \nu_\ell(\mathcal{Y}_k)\Big) + \epsilon\,\nu(\mathcal{Y}_k)\Big| \\
&\underset{(*)}{\leqslant} |n|\,\epsilon\,\Big|\nu(\mathcal{Y}_k) - p_k^\ell\Big| + \epsilon\,\nu(\mathcal{Y}_k) \;\underset{(\dagger)}{<}\; N\,\epsilon \cdot \frac{1}{K\,N} + \epsilon\,\nu(\mathcal{Y}_k) \;=\; \frac{\epsilon}{K} + \epsilon\,\nu(\mathcal{Y}_k).
\end{aligned} \tag{A19}$$

Here, $(\diamond)$ is because $n < 0$, so that $n < n+1 \leqslant 0$, hence $|n+1| < |n|$. Meanwhile, both $(*)$ are by the definition of $\mathbf{p}^\ell$, while both $(\dagger)$ use hypotheses (A16) and the fact that $|n|\epsilon \leqslant N\,\epsilon$.

Inequalities (A17), (A18) and (A19) imply that

$$\left|\int\limits_{\mathcal{Y}_k} u^j \,\mathrm{d}\nu - \int\limits_{\mathcal{Y}_k} u^j \,\mathrm{d}\nu_\ell\right| < \frac{\epsilon}{K} + \epsilon\,\nu(\mathcal{Y}_k). \tag{A20}$$

This holds for all $k \in [1..K]$. Thus,

$$\begin{aligned}
\left|\int\limits_{\mathcal{X}} u^j \,\mathrm{d}\nu - \int\limits_{\mathcal{X}} u^j \,\mathrm{d}\nu_\ell\right| &\underset{(*)}{\leqslant} \sum_{k=1}^{K} \left|\int\limits_{\mathcal{Y}_k} u^j \,\mathrm{d}\nu - \int\limits_{\mathcal{Y}_k} u^j \,\mathrm{d}\nu_\ell\right| \;\underset{(\dagger)}{<}\; \sum_{k=1}^{K} \left(\frac{\epsilon}{K} + \epsilon\,\nu(\mathcal{Y}_k)\right) \\
&= K\,\frac{\epsilon}{K} + \epsilon \sum_{k=1}^{K} \nu(\mathcal{Y}_k) \;=\; \epsilon + \epsilon \;=\; 2\epsilon,
\end{aligned}$$

as claimed. Here, $(*)$ is because $\mathcal{X} = \bigsqcup_{k=0}^{K} \mathcal{Y}_k$, while $(\dagger)$ is by inequality (A20). ◇ Claim 1

Let η be the concordance measure. For each $j \in \mathcal{J}$, $\eta \ll \rho^j$ by hypothesis, so Lemma A.1 yields a measurable subset $\mathcal{G}_j \subseteq \mathcal{S}^{\mathbb{N}}$ with $\eta(\mathcal{G}_j) = 1$, such that

$$\lim_{t\to\infty} \left\|\rho^j_{\mathbf{s},t} - \eta_{\mathbf{s},t}\right\| \;=\; 0, \qquad \text{for all } \mathbf{s} \in \mathcal{G}_j. \tag{A21}$$

Let $\mathcal{G} := \bigcap_{j\in\mathcal{J}} \mathcal{G}_j$. Then $\eta(\mathcal{G}) = 1$, and for all $\mathbf{s} \in \mathcal{G}$, the limit (A21) holds for all $j \in \mathcal{J}$. Thus, for all $\mathbf{s} \in \mathcal{G}$, there exists $T'_{\mathbf{s}} \in \mathbb{N}$ such that

$$\left\|\rho^j_{\mathbf{s},t} - \eta_{\mathbf{s},t}\right\| \;<\; \frac{1}{2K\,N}, \qquad \text{for all } j \in \mathcal{J} \text{ and all } t \geqslant T'_{\mathbf{s}}. \tag{A22}$$

Claim 2: *There is an event $\mathcal{F} \subseteq \mathcal{S}^{\mathbb{N}}$ with $\eta(\mathcal{F}) > 0$, and a finite collection of nonatomic measures $\{\mu_1, \ldots, \mu_M\}$ such that, for any $\mathbf{s} \in \mathcal{F}$ and event $\mathcal{B} \subseteq \mathcal{S}^{\mathbb{N}}$, there exists $T''_{\mathbf{s},\mathcal{B}} \in \mathbb{N}$ such that for all $t \geqslant T''_{\mathbf{s},\mathcal{B}}$, there is some $m \in [1..M]$ with $\left|\eta_{\mathbf{s},t}(\vec{{}^t}\mathcal{B}) - \mu_m(\mathcal{B})\right| < \dfrac{1}{2K\,N}$.*

Proof. (by contradiction) Let $\delta := \frac{1}{2KN}$. The process η is coalescent, so there is an event $\mathcal{F} \subseteq \mathcal{S}^{\mathbb{N}}$ with $\eta(\mathcal{F}) > 0$, and a finite collection of nonatomic measures $\{\mu_1, \ldots, \mu_M\}$ such that for any measurable $\mathcal{B} \subseteq \mathcal{S}^{\mathbb{N}}$, the set of values $\{\mu_1(\mathcal{B}), \ldots, \mu_M(\mathcal{B})\}$ is δ-dense in the set of cluster points $\mathcal{C}_{\mathbf{s},\mathcal{B}}$ for all $\mathbf{s} \in \mathcal{F}$.

Suppose there was some $\mathbf{s} \in \mathcal{F}$ and event $\mathcal{B} \subseteq \mathcal{S}^{\mathbb{N}}$ falsifying the claim. Thus, for any $T \in \mathbb{N}$, there exists $t > T$ such that $\left|\eta_{\mathbf{s},t}({}^{\vec{t}}\mathcal{B}) - \mu_m(\mathcal{B})\right| \geqslant \delta$ for all $m \in [1..M]$. Thus, there is an infinite sequence $t_1 < t_2 < t_3 < \cdots$ such that $\left|\eta_{\mathbf{s},t_n}[{}^{\vec{t}_n}\mathcal{B}] - \mu_m(\mathcal{B})\right| \geqslant \delta$ for all $m \in [1..M]$ and $n \in \mathbb{N}$. But $\{\eta_{\mathbf{s},t_n}[{}^{\vec{t}_n}\mathcal{B}]\}_{n=1}^{\infty}$ is a subset of $[0, 1]$, which is compact. So by dropping to a subsequence if necessary, we can suppose that this sequence converges to some $c \in [0, 1]$. It follows that $|c - \mu_m(\mathcal{B})| \geqslant \delta$ for all $m \in [1..M]$. But $c \in \mathcal{C}_{\mathbf{s},\mathcal{B}}$, so this contradicts coalescence. ◇ Claim 2

Let $\mathcal{H} := \mathcal{F} \cap \mathcal{G}$. Then $\eta(\mathcal{H}) > 0$. For all $\mathbf{s} \in \mathcal{H}$ and all measurable $\mathcal{B} \subseteq \mathcal{S}^{\mathbb{N}}$, define $T_{\mathbf{s},\mathcal{B}} := \max\{T'_{\mathbf{s}}, T''_{\mathbf{s},\mathcal{B}}\}$. For all $j \in \mathcal{J}$ and $t \geqslant T_{\mathbf{s},\mathcal{B}}$, Claim 2 and inequality (A22) imply

$$\left|\rho^j_{\mathbf{s},t}({}^{\vec{t}}\mathcal{B}) - \mu_m(\mathcal{B})\right| \;<\; \frac{1}{KN} \qquad \text{for some } m \in [1..M]. \tag{A23}$$

Recall $(p^1_k)_{k=1}^K$ and $(p^2_k)_{k=1}^K$, defined just before Claim 1. The measures $\{\mu_1, \ldots, \mu_M\}$ are nonatomic. Thus, for both $\ell \in \{1, 2\}$, the Dubins-Spanier Theorem yields a measurable partition $\mathfrak{P}^\ell := \{\mathcal{P}^\ell_k\}_{k=1}^K$ of $\mathcal{S}^{\mathbb{N}}$ such that

$$\mu_m(\mathcal{P}^\ell_k) \;=\; p^\ell_k, \quad \text{for all } k \in [1..K] \text{ and all } m \in [1..M]. \tag{A24}$$

(Aliprantis and Border, 2006, Theorem 13.34, p. 478). For all $k \in [1..K]$, let $y_k \in \mathcal{Y}_k$. For both $\ell \in \{1, 2\}$, define $\alpha^\ell : \mathcal{S}^{\mathbb{N}} \longrightarrow \mathcal{X}$ by setting $(\alpha^\ell)^{-1}\{y_k\} := \mathcal{P}^\ell_k$ for all $k \in [1..K]$. Then α^ℓ is measurable, and $\alpha^\ell(\mathbf{s}) \in \{y_k\}_{k=1}^K$ (a finite set) for all $\mathbf{s} \in \mathcal{S}^{\mathbb{N}}$; thus $\alpha^\ell \in \mathcal{A}$. I will now show that α^1 and α^2 violate the Eventual Pareto axiom (9).

For any $\mathbf{s} \in \mathcal{H}$, all $k \in [1..K]$ and all $\ell \in \{1, 2\}$, define $T_{\mathbf{s},\mathcal{P}^\ell_k}$ as prior to inequality (A23). Then let $T_{\mathbf{s}} := \max\{T_{\mathbf{s},\mathcal{P}^\ell_k};\ \ell \in \{1, 2\} \text{ and } k \in [1..K]\}$. Thus, for all $\mathbf{s} \in \mathcal{H}$, if $t \geqslant T_{\mathbf{s}}$, then for all $j \in \mathcal{J}$, $k \in [1..K]$, and $\ell \in \{1, 2\}$ statement (A23) implies that

$$\left|\rho^j_{\mathbf{s},t}({}^{\vec{t}}\mathcal{P}^\ell_k) - \mu_m(\mathcal{P}^\ell_k)\right| \;<\; \frac{1}{KN}, \qquad \text{for some } m \in [1..M]. \tag{A25}$$

Combining statements (A24) and (A25), we deduce

$$\left|\rho^j_{\mathbf{s},t}({}^{\vec{t}}\mathcal{P}^\ell_k) - p^\ell_k\right| \;<\; \frac{1}{KN}, \quad \text{for all } j \in \mathcal{J},\ k \in [1..K],\ \ell \in \{1, 2\}, \text{ and } t \geqslant T_{\mathbf{s}}. \tag{A26}$$

Now, for all $k \in [1..K]$ and $\ell \in \{1, 2\}$, the construction of α^ℓ implies that $(\alpha^\ell)^{-1}(\mathcal{Y}_k) = \mathcal{P}^\ell_k$; thus $({}^{\vec{t}}\alpha^\ell)^{-1}(\mathcal{Y}_k) = {}^{\vec{t}}\mathcal{P}^\ell_k$. For all $j \in \mathcal{J}$, $\ell \in \{1, 2\}$, $\mathbf{s} \in \mathcal{H}$, and $t \geqslant T_{\mathbf{s}}$, let $\nu^{j,\ell}_{\mathbf{s},t} := {}^{\vec{t}}\alpha^\ell(\rho^j_{\mathbf{s},t})$ (i.e. $\nu^{j,\ell}_{\mathbf{s},t}(\mathcal{W}) := \rho^j_{\mathbf{s},t}\left[({}^{\vec{t}}\alpha^\ell)^{-1}(\mathcal{W})\right]$, for any measurable subset $\mathcal{W} \subseteq \mathcal{X}$). Then

$$\nu^{j,\ell}_{\mathbf{s},t}[\mathcal{Y}_k] \;=\; \rho^j_{\mathbf{s},t}({}^{\vec{t}}\mathcal{P}^\ell_k), \quad \text{for all } j \in \mathcal{J},\ k \in [1..K], \text{ and } \ell \in \{1, 2\}. \tag{A27}$$

Substituting equation (A27) into inequality (A26) yields

$$\left\|\nu^{j,\ell}_{\mathbf{s},t}[\mathcal{Y}_k] - p^\ell_k\right\| \;<\; \frac{1}{KN}, \quad \text{for all } j \in \mathcal{J},\ k \in [1..K],\ \ell \in \{1, 2\}, \text{ and } t \geqslant T_{\mathbf{s}}. \tag{A28}$$

Thus, Claim 1 and the inequalities (A28) yield

$$\left| \int_{\mathcal{X}} u^j \, \mathrm{d}\nu^{j,\ell}_{\mathbf{s},t} - \int_{\mathcal{X}} u^j \, \mathrm{d}\nu_\ell \right| < 2\epsilon, \quad \text{for all } j \in \mathcal{J}, \text{ both } \ell \in \{1,2\}, \text{ and all } t \geqslant T_{\mathbf{s}}. \tag{A29}$$

A change of variables theorem (see e.g. Petersen 1989, Proposition 1.4.1, p. 13) yields

$$\int_{\mathcal{X}} u^j \, \mathrm{d}\nu^{j,\ell}_{\mathbf{s},t} \quad = \quad \int_{\mathcal{S}^{\mathbb{N}}} \left(u^j \circ \vec{^t\alpha}^\ell \right) \mathrm{d}\rho^j_{\mathbf{s},t}, \qquad \text{for all } j \in \mathcal{J} \text{ and } \ell \in \{1,2\}. \tag{A30}$$

For all $t \geqslant T_{\mathbf{s}}$, combining the equations (A30) with inequalities (A29) and (A15) yields

$$\begin{aligned} &\int_{\mathcal{S}^{\mathbb{N}}} u^0 \circ \vec{^t\alpha}^2 \, \mathrm{d}\rho^0_{\mathbf{s},t} - \int_{\mathcal{S}^{\mathbb{N}}} u^0 \circ \vec{^t\alpha}^1 \, \mathrm{d}\rho^0_{\mathbf{s},t} > \epsilon, \quad \text{while} \\ &\int_{\mathcal{S}^{\mathbb{N}}} u^i \circ \vec{^t\alpha}^2 \, \mathrm{d}\rho^i_{\mathbf{s},t} - \int_{\mathcal{S}^{\mathbb{N}}} u^i \circ \vec{^t\alpha}^1 \, \mathrm{d}\rho^i_{\mathbf{s},t} < -\epsilon \quad \text{for all } i \in \mathcal{I}. \end{aligned} \tag{A31}$$

By defining formula (5), the inequalities (A31) imply that

$$\alpha^1 \; {}_{\epsilon}{\prec}_{\mathbf{s},t} \; \alpha^2, \quad \text{while} \quad \alpha^1 \; {}_{\epsilon}{\succ}^i_{\mathbf{s},t} \; \alpha^2 \quad \text{for all } i \in \mathcal{I} \text{ and all } t \geqslant T_{\mathbf{s}}. \tag{A32}$$

This holds for all $\mathbf{s} \in \mathcal{H}$. But $\eta(\mathcal{H}) > 0$, and $\eta \ll \rho^j$ for all $j \in \mathcal{J}$, so $\mathcal{H}$ is unanimously non-null. Comparing statement (A32) and definition (6) yields $\alpha^1 \prec_{\mathcal{H}} \alpha^2$ while $\alpha^1 \succ^i_{\mathcal{H}} \alpha^2$ for all $i \in \mathcal{I}$, contradicting statement (9), and thereby contradicting Eventual Pareto.

To avoid the contradiction, $W = u^0$ must be an element of the cone $\mathcal{C}$, which means that W is weakly utilitarian. □

Remark A.3. (a) Note that concordance is only used in the "$\Longrightarrow$" direction of the proof. The proof of "$\Longleftarrow$" works for any collection of measures $\{\rho^j\}_{j \in \mathcal{J}}$.

(b) The argument leading up to inequality (A23) can be used to show that the measures $\{\rho_j\}_{j \in \mathcal{J}}$ themselves are coalescent, if they are concordant.

(c) Diaconis and Freedman (1986, Theorem 3) proved a version of the Blackwell-Dubins theorem for *exchangeable* stochastic processes (i.e. mixtures of independent coin-tossing processes). Their result does not require any absolute continuity assumption, but it only yields weak* convergence of conditional beliefs rather than convergence in total variation norm. However, weak* convergence is all that is needed for the proof above, and any coin-tossing process is coalescent. This yields a version of result which replaces concordance with the assumption that the beliefs of all agents take form of exchangeable processes.

Appendix B. Proofs of other statements

Proofs of coalescence (from Section 2). Parts (i), (ii), and (iii) are immediate.

(iv) Let $\eta \in \Delta(\mathcal{S}^{\mathbb{N}})$ be uniformly and fully coalescent, and let $\eta' := \Phi(\eta)$; we must show that η' is also uniformly and fully coalescent. Let $\epsilon > 0$. Let $\epsilon' := \epsilon/3$. Since η is fully coalescent, there is a finite collection $\{\mu_1, \ldots, \mu_M\} \subset \Delta(\mathcal{S}^{\mathbb{N}})$ such that, for any event $\mathcal{B} \subseteq \mathcal{S}^{\mathbb{N}}$, the set $\{\mu_1(\mathcal{B}), \ldots, \mu_M(\mathcal{B})\}$ is ϵ'-dense in the set $\mathcal{C}_{\mathcal{B}} := \bigcup_{\mathbf{s} \in \mathcal{S}^{\mathbb{N}}} \mathcal{C}_{\mathbf{s},\mathcal{B}}$. Let $\mu'_m := \Phi(\mu_m)$ for all

$m \in [1..M]$. Let $\mathcal{K}$ be the closed convex hull of $\{\mu'_1, \ldots, \mu'_M\}$ in $\Delta(\mathcal{X})$. This is a compact subset of $\Delta(\mathcal{X})$ with respect to the total variation norm (because it is the continuous image of an M-dimensional simplex, which is compact). Thus, there is a finite subset $\Lambda_\epsilon \subseteq \mathcal{K}$ that is ϵ'-dense in $\mathcal{K}$ in the total variation norm. I will show that Λ_ϵ satisfies the coalescence property for η'. To be precise for any event $\mathcal{A} \subseteq \mathcal{R}^{\mathbb{N}}$, I will show that the set $\{\lambda(\mathcal{A})\}_{\lambda \in \Lambda_\epsilon}$ is ϵ-dense in the set $\mathcal{C}_\mathcal{A} := \bigcup_{\mathbf{r} \in \mathcal{R}^{\mathbb{N}}} \mathcal{C}_{\mathbf{r},\mathcal{A}}$.

For any $t \in \mathbb{N}$, define ${}^{\vec{t}}\Phi : \mathcal{S}^{(t..\infty)} \longrightarrow \mathcal{R}^{(t..\infty)}$ in the obvious way. Let $\mathcal{B} := \Phi^{-1}(\mathcal{A})$. For all $t \in \mathbb{N}$, it is easily verified that ${}^{\vec{t}}\Phi^{-1}({}^{\vec{t}}\mathcal{A}) = {}^{\vec{t}}\mathcal{B}$. For any $\mathbf{r} \in \mathcal{R}^{\mathbb{N}}$ and $t \in \mathbb{N}$, recall that $[\mathbf{r}_{[0..t]}] := \{\mathbf{r}' \in \mathcal{R}^{\mathbb{N}};\ r'_n = r_n \text{ for all } n \in [0..t]\}$. A simple computation shows that

$$\eta'_{\mathbf{r},t}({}^{\vec{t}}\mathcal{A}) \quad = \quad \frac{1}{\eta'[\mathbf{r}_{[0..t]}]} \int_{\mathcal{Q}_t} \eta_{\mathbf{s},t}({}^{\vec{t}}\mathcal{B})\, \mathrm{d}\eta[\mathbf{s}], \quad \text{where} \quad \mathcal{Q}_t \ := \ \Phi^{-1}[\mathbf{r}_{[0..t]}]. \tag{B1}$$

Since η is *uniformly* coalescent, there exists $T_\epsilon \in \mathbb{N}$ such that, for all $\mathbf{s} \in \mathcal{S}^{\mathbb{N}}$ and all $t \geqslant T_\epsilon$, there is some $c_t \in \mathcal{C}_\mathcal{B}$ such that $\left|\eta_{\mathbf{s},t}({}^{\vec{t}}\mathcal{B}) - c_t\right| < \epsilon'$. Meanwhile, there is some $m_t(\mathbf{s}) \in [1..M]$ such that $\left|c_t - \mu_{m_t(\mathbf{s})}(\mathcal{B})\right| < \epsilon'$, because $\{\mu_1(\mathcal{B}), \ldots, \mu_M(\mathcal{B})\}$ is ϵ'-dense in $\mathcal{C}_\mathcal{B}$. Combining these inequalities, we conclude that $\left|\eta_{\mathbf{s},t}({}^{\vec{t}}\mathcal{B}) - \mu_{m_t(\mathbf{s})}(\mathcal{B})\right| < 2\epsilon'$. For all $m \in [1..M]$ and $t \geqslant T_\epsilon$, let $\mathcal{Q}_t^m := \{\mathbf{s} \in \mathcal{Q}_t;\ m_t(\mathbf{s}) = m\}$. Then

$$\left|\eta_{\mathbf{s},t}({}^{\vec{t}}\mathcal{B}) - \mu_m(\mathcal{B})\right| \ < \ 2\epsilon', \quad \text{for all } \mathbf{s} \in \mathcal{Q}_t^m. \tag{B2}$$

Let $q_t^m := \dfrac{\eta(\mathcal{Q}_t^m)}{\eta(\mathcal{Q}_t)}$. Note that $\mathcal{Q}_t = \bigsqcup_{m=1}^{M} \mathcal{Q}_t^m$. Thus, $\sum_{m=1}^{M} q_t^m = 1$, so $\sum_{m=1}^{M} q_t^m\, \mu'_m \in \mathcal{K}$. Thus,

$$\begin{aligned}
\inf_{\kappa \in \mathcal{K}} \left|\eta'_{\mathbf{r},t}({}^{\vec{t}}\mathcal{A}) - \kappa(\mathcal{A})\right| &\leqslant \left|\eta'_{\mathbf{r},t}({}^{\vec{t}}\mathcal{A}) \ - \ \sum_{m=1}^{M} q_t^m\, \mu'_m(\mathcal{A})\right| \\
&\underset{(*)}{=} \left|\frac{1}{\eta'[\mathbf{r}_{[0..t]}]} \int_{\mathcal{Q}_t} \eta_{\mathbf{s},t}({}^{\vec{t}}\mathcal{B})\, \mathrm{d}\eta[\mathbf{s}] \ - \ \sum_{m=1}^{M} q_t^m\, \mu_m(\mathcal{B})\right| \\
&\underset{(\dagger)}{=} \frac{1}{\eta(\mathcal{Q}_t)} \left|\int_{\mathcal{Q}_t} \eta_{\mathbf{s},t}({}^{\vec{t}}\mathcal{B})\, \mathrm{d}\eta[\mathbf{s}] \ - \ \sum_{m=1}^{M} \eta(\mathcal{Q}_t^m)\, \mu_m(\mathcal{B})\right|.
\end{aligned} \tag{B3}$$

Here, $(*)$ is by equation (B1), and the fact that $\eta' = \Phi(\eta)$ while $\mathcal{B} = \Phi^{-1}(\mathcal{A})$. Meanwhile, $(\dagger)$ is because $\eta' = \Phi(\eta)$, $\mathcal{Q}_t = \Phi^{-1}[\mathbf{r}_{[0..t]}]$, and $q_t^m = \eta(\mathcal{Q}_t^m)/\eta(\mathcal{Q}_t)$. But

$$\left|\int_{\mathcal{Q}_t} \eta_{\mathbf{s},t}({}^{\vec{t}}\mathcal{B})\, \mathrm{d}\eta[\mathbf{s}] - \sum_{m=1}^{M} \eta(\mathcal{Q}_t^m)\, \mu_m(\mathcal{B})\right| \quad = \quad \left|\int_{\mathcal{Q}_t} \eta_{\mathbf{s},t}({}^{\vec{t}}\mathcal{B})\, \mathrm{d}\eta[\mathbf{s}] - \sum_{m=1}^{M} \int_{\mathcal{Q}_t^m} \mu_m(\mathcal{B})\, \mathrm{d}\eta\right|$$

$$\underset{(\dagger)}{=} \left|\sum_{m=1}^{M} \int_{\mathcal{Q}_t^m} \left(\eta_{\mathbf{s},t}({}^{\vec{t}}\mathcal{B}) - \mu_m(\mathcal{B})\right)\, \mathrm{d}\eta[\mathbf{s}]\right| \quad \leqslant \quad \sum_{m=1}^{M} \int_{\mathcal{Q}_t^m} \left|\eta_{\mathbf{s},t}({}^{\vec{t}}\mathcal{B}) - \mu_m(\mathcal{B})\right|\, \mathrm{d}\eta[\mathbf{s}]$$

$$\underset{(*)}{\leqslant} \sum_{m=1}^{M} \int_{\mathcal{Q}_t^m} 2\,\epsilon' \, \mathrm{d}\eta[\mathbf{s}] \quad = \quad 2\,\epsilon' \sum_{m=1}^{M} \eta(\mathcal{Q}_t^m) \quad \underset{(\dagger)}{=} \quad 2\,\epsilon'\, \eta(\mathcal{Q}_t). \tag{B4}$$

Here, $(*)$ is by inequality (B2), and both $(\dagger)$ are because $\mathcal{Q}_t = \bigsqcup_{m=1}^{M} \mathcal{Q}_t^m$. Combining (B3) and (B4), we get $\inf_{\kappa \in \mathcal{K}} \left| \eta'_{\mathbf{r},t}({}^{\vec{t}}\!\mathcal{A}) - \kappa(\mathcal{A}) \right| \leqslant 2\,\epsilon'$. By construction, Λ_ϵ is ϵ'-dense in $\mathcal{K}$. Thus, $\inf_{\lambda \in \Lambda_\epsilon} \left| \eta'_{\mathbf{r},t}({}^{\vec{t}}\!\mathcal{A}) - \lambda(\mathcal{A}) \right| < 3\,\epsilon' = \epsilon$.

This argument works for any event $\mathcal{A} \subseteq \mathcal{R}^{\mathbb{N}}$, any $\mathbf{r} \in \mathcal{R}^{\mathbb{N}}$, and any $t \geqslant T_\epsilon$. We can construct such a finite subset $\Lambda_\epsilon \subset \Delta(\mathcal{R}^{\mathbb{N}})$ and $T_\epsilon \in \mathbb{N}$ for any $\epsilon > 0$. We conclude that η' is fully and uniformly coalescent.

(v) Suppose $\mathcal{S}$ is finite and η is quasimarkovian, with Markov function $\mu : \mathcal{S}^* \longrightarrow \Delta(\mathcal{S}^{\mathbb{N}})$, such that $\mu(\mathbf{s})$ is nonatomic for all $\mathbf{s} \in \mathcal{S}^*$. I claim that η is coalescent. To see this, let $0 < \epsilon < \epsilon'$. There is some $M > 0$ and event $\mathcal{F} \subseteq \mathcal{S}^{\mathbb{N}}$ with $\eta(\mathcal{F}) > 0$ such that for all $\mathbf{s} \in \mathcal{F}$, the limsup inequality (2) is satisfied. For all $\mathbf{r} \in \mathcal{S}^M$, let $\mu_{\mathbf{r}} := \mu(\mathbf{r}) \in \Delta(\mathcal{S}^{\mathbb{N}})$. Let $\mathcal{M} := \{\mu_{\mathbf{r}};\ \mathbf{r} \in \mathcal{S}^M\}$. This is a finite collection of measures, because $\mathcal{S}$ is finite.

Let $\mathbf{s} \in \mathcal{F}$ and let $\mathcal{B} \subseteq \mathcal{S}^{\mathbb{N}}$ be an event. I claim that $\{\mu(\mathcal{B});\ \mu \in \mathcal{M}\}$ is ϵ'-dense in $\mathcal{C}_{\mathbf{s},\mathcal{B}}$. To see this, let $c \in \mathcal{C}_{\mathbf{s},\mathcal{B}}$. Then $c = \lim_{n\to\infty} \eta_{\mathbf{s},t_n}({}^{\vec{t_n}}\!\mathcal{B})$ for some sequence $t_1 < t_2 < t_3 < \cdots$. Inequality (2) says that there is some $T_{\mathbf{s},\mathcal{B}} \geqslant M$ such that $\left| \eta_{\mathbf{s},t}({}^{\vec{t}}\!\mathcal{B}) - \mu(\mathbf{s}_{(t-M\,..t]})(\mathcal{B}) \right| \leqslant \epsilon$ for all $t \geqslant T_{\mathbf{s},\mathcal{B}}$. Let $N := \min\{n \in \mathbb{N};\ t_n \geqslant T_{\mathbf{s},\mathcal{B}}\}$. Then $\left| \eta_{\mathbf{s},t_n}({}^{\vec{t_n}}\!\mathcal{B}) - \mu(\mathbf{s}_{(t_n-M\,..t_n]})(\mathcal{B}) \right| \leqslant \epsilon$ for all $n \geqslant N$. By dropping to a subsequence if necessary, we can fix $\mathbf{r} \in \mathcal{S}^M$ such that $\mathbf{s}_{(t_n-M\,..t_n]} = \mathbf{r}$ for all $n \geqslant N$ (because $\mathcal{S}^M$ is finite). Thus, we have $\left| \eta_{\mathbf{s},t_n}({}^{\vec{t_n}}\!\mathcal{B}) - \mu_{\mathbf{r}}(\mathcal{B}) \right| \leqslant \epsilon$ for all $n \geqslant N$. Thus, we must have $|c - \mu_{\mathbf{r}}(\mathcal{B})| \leqslant \epsilon < \epsilon'$, as desired.

This argument works for any $\epsilon' > 0$. We conclude that η is coalescent. □

Proof of Example 1. Suppose ρ is quasimarkovian, with Markov function $\mu : \mathcal{S}^* \longrightarrow \Delta(\mathcal{S}^{\mathbb{N}})$. Let $\alpha, \beta \in \mathcal{A}$, and suppose there is some $\epsilon' > 0$ and $N \in \mathbb{N}$ satisfying inequality (7). Let $K := \|u\|_\infty$ and let $\epsilon := \epsilon'/(2\,K + 1)$. The quasimarkovian property yields some $M \geqslant N$ and measurable $\mathcal{F} \subseteq \mathcal{S}^{\mathbb{N}}$ with $\rho(\mathcal{F}) > 0$ such that for any $\mathbf{s} \in \mathcal{F}$ and measurable subsets $\mathcal{B} \subseteq \mathcal{S}^{\mathbb{N}}$, the limsup inequality (2) holds; hence there is some $T_{\mathbf{s},\mathcal{B}} \geqslant M$ such that $\left| \rho_{\mathbf{s},t}({}^{\vec{t}}\!\mathcal{B}) - \mu_{\mathbf{s}_{(t-M\,..t]}}(\mathcal{B}) \right| \leqslant \epsilon$ for all $t \geqslant T_{\mathbf{s},\mathcal{B}}$. Since α and β are finitely valued, there is a measurable partition $\mathfrak{P} = \{\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_J\}$ of $\mathcal{S}^{\mathbb{N}}$ such that both α and β are measurable with respect to $\mathfrak{P}$. Thus, the integrals of $u \circ \alpha$ and $u \circ \beta$ over $\mathcal{S}^{\mathbb{N}}$ are weighted sums involving the measures of $\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_J$, whereas the integrals of $u \circ {}^{\vec{t}}\alpha$ and $u \circ {}^{\vec{t}}\beta$ over $\mathcal{S}^{(t\,..\infty)}$ are the corresponding weighted sums involving the measures of ${}^{\vec{t}}\mathcal{P}_1, {}^{\vec{t}}\mathcal{P}_2, \ldots, {}^{\vec{t}}\mathcal{P}_J$. For any $\mathbf{s} \in \mathcal{F}$, define $T_{\mathbf{s}} := \max\{T_{\mathbf{s},\mathcal{P}_1},\ T_{\mathbf{s},\mathcal{P}_2}, \ldots, T_{\mathbf{s},\mathcal{P}_J}\}$; then $T_{\mathbf{s}}$ is finite, and for all $t \geqslant T_{\mathbf{s}}$, we have $\left| \rho_{\mathbf{s},t}[{}^{\vec{t}}\mathcal{P}_j] - \mu_{\mathbf{s}_{(t-M\,..t]}}[\mathcal{P}_j] \right| \leqslant \epsilon$ for all $j \in [1..J]$. Thus,

$$\left| \int_{\mathcal{S}^{(t..\infty)}} u \circ {}^{\vec{t}}\alpha \; \mathrm{d}\rho_{\mathbf{s},t} - \int_{\mathcal{S}^{\mathbb{N}}} u \circ \alpha \; \mathrm{d}\mu_{\mathbf{s}_{(t-M\,..t]}} \right| \leqslant K\,\epsilon \tag{B5}$$

$$\text{and} \qquad \left| \int_{\mathcal{S}^{(t..\infty)}} u \circ {}^{\vec{t}}\beta \; \mathrm{d}\rho_{\mathbf{s},t} - \int_{\mathcal{S}^{\mathbb{N}}} u \circ \beta \; \mathrm{d}\mu_{\mathbf{s}_{(t-M\,..t]}} \right| \leqslant K\,\epsilon, \tag{B6}$$

for all $t \geqslant T_{\mathbf{s}}$. (Recall $K := \|u\|_\infty$.) Combining inequalities (7), (B5) and (B6) yields

$$\int_{\mathcal{S}^{(t..\infty)}} u \circ \overset{t}{\vec{\alpha}}\, \mathrm{d}\rho_{\mathbf{s},t} \quad > \quad \epsilon + \int_{\mathcal{S}^{(t..\infty)}} u \circ \overset{t}{\vec{\beta}}\, \mathrm{d}\rho_{\mathbf{s},t}, \qquad \text{for all } t \geqslant T_{\mathbf{s}},$$

hence, $\alpha \;{}_{\epsilon}{\succ}_{\mathbf{s},t}\; \beta$, for all $t \geqslant T_{\mathbf{s}}$. This holds for all $\mathbf{s} \in \mathcal{F}$; thus, $\alpha \succ_{\mathcal{F}} \beta$, as claimed. □

Finally, here is a technical result that was mentioned in footnote 16 and in §4.3.

Proposition B.1. *Let $\{\rho^j\}_{j\in\mathcal{J}}$ be a collection of probability measures on $\mathcal{S}^{\mathbb{N}}$. There exists $\eta \in \Delta(\mathcal{S}^{\mathbb{N}})$ such that $\eta \ll \rho^j$ for all $j \in \mathcal{J}$ if and only if $\{\rho^j\}_{j\in\mathcal{J}}$ is not singular.*

Proof. "$\Longrightarrow$" (by contradiction) Suppose there exists $\eta \in \Delta(\mathcal{S}^{\mathbb{N}})$ such that $\eta \ll \rho^j$ for all $j \in \mathcal{J}$, but $\{\rho^j\}_{j\in\mathcal{J}}$ is singular. Let $\{\mathcal{B}_j\}_{j\in\mathcal{J}}$ be a measurable partition of $\mathcal{S}^{\mathbb{N}}$ such that $\rho^j(\mathcal{B}_j) = 0$ for all $j \in \mathcal{J}$. Then for all $j \in \mathcal{J}$, we have $\eta(\mathcal{B}_j) = 0$, because $\eta \ll \rho^j$. Thus, $\eta(\mathcal{S}^{\mathbb{N}}) = \sum_{j\in\mathcal{J}} \eta(\mathcal{B}_j) = 0$, contradicting the fact that $\eta \in \Delta(\mathcal{S}^{\mathbb{N}})$.

"$\Longleftarrow$" (by induction on $|\mathcal{J}|$) The case $|\mathcal{J}| = 2$ follows from the Lebesgue Decomposition Theorem: for any two measures if $\rho^1, \rho^2 \in \Delta(\mathcal{S}^{\mathbb{N}})$, we can write $\rho^1 = \widetilde{\rho} + \rho_\perp$, where $\widetilde{\rho} \ll \rho^2$, while $\rho_\perp$ and ρ^2 are singular. It is easily verified from this equation that $\widetilde{\rho} \ll \rho^1$ as well. If ρ^1 and ρ^2 are *not* singular, then $\widetilde{\rho} \neq 0$. Thus, let $\eta := \widetilde{\rho}/\widetilde{\rho}(\mathcal{S}^{\mathbb{N}})$; then $\eta \in \Delta(\mathcal{S}^{\mathbb{N}})$, and by construction $\eta \ll \rho^1$ and $\eta \ll \rho^2$.

Now let $J \geqslant 3$, and suppose inductively that the claim is true for $|\mathcal{J}| = J - 1$. For simplicity, suppose $\mathcal{J} = \{1, 2, 3, \ldots, J\}$. Since $\{\rho^j\}_{j\in\mathcal{J}}$ is not singular, in particular the measure ρ^J is not singular versus any of $\rho^1, \ldots, \rho^{J-1}$. Thus, for all $i \in [1..J)$, the Lebesgue Decomposition Theorem yields two measures $\widetilde{\rho}^i$ and $\rho^i_\perp$ on $\mathcal{S}^{\mathbb{N}}$ with $\rho^i = \widetilde{\rho}^i + \rho^i_\perp$, such that $0 \neq \widetilde{\rho}^i \ll \rho^J$ and $\rho^i_\perp$ is singular to ρ^J. This means there are disjoint measurable sets $\mathcal{B}^i, \mathcal{C}^i \subseteq \mathcal{S}^{\mathbb{N}}$ such that $\mathcal{S}^{\mathbb{N}} = \mathcal{B}^i \sqcup \mathcal{C}^i$, with $\rho^i_\perp(\mathcal{B}^i) = 0$ and $\rho^J(\mathcal{C}^i) = 0$. Let

$$\mathcal{B} \;:=\; \bigcap_{i=1}^{J-1} \mathcal{B}^i \quad \text{and} \quad \mathcal{C} \;:=\; \bigcup_{i=1}^{J-1} \mathcal{C}^i,$$

Then $\mathcal{S}^{\mathbb{N}} = \mathcal{B} \sqcup \mathcal{C}$ (by de Morgan's Law), and $\rho^i_\perp(\mathcal{B}) = 0$ for all $i \in [1..J)$, while $\rho^J(\mathcal{C}) = 0$, and hence $\widetilde{\rho}^i(\mathcal{C}) = 0$ for all $i \in [1..J)$.

Claim 1: $\{\widetilde{\rho}^1, \ldots, \widetilde{\rho}^{J-1}\}$ *is not singular.*

Proof. (by contradiction) Suppose this collection was singular. Then there would be a measurable partition $\mathcal{S}^{\mathbb{N}} = \mathcal{D}^1 \sqcup \cdots \sqcup \mathcal{D}^{J-1}$ such that $\widetilde{\rho}^i(\mathcal{D}^i) = 0$ for all $i \in [1..J)$. For all $i \in [1..J)$, let $\mathcal{E}^i := \mathcal{B} \cap \mathcal{D}_i$. Then $\mathcal{B} = \mathcal{E}^1 \sqcup \cdots \sqcup \mathcal{E}^{J-1}$, and $\widetilde{\rho}^i(\mathcal{E}^i) = 0$ for all $i \in [1..J)$. Now let $\mathcal{E}^J := \mathcal{C}$. Then $\mathcal{E}^1 \sqcup \cdots \sqcup \mathcal{E}^{J-1} \sqcup \mathcal{E}^J = \mathcal{B} \sqcup \mathcal{C} = \mathcal{S}^{\mathbb{N}}$, so the collection $\{\mathcal{E}^j\}_{j=1}^J$ is a measurable partition of $\mathcal{S}^{\mathbb{N}}$. For any $i \in [1..J)$, we have

$$\rho^i(\mathcal{E}^i) \;\underset{(*)}{=}\; \widetilde{\rho}^i(\mathcal{E}^i) + \rho^i_\perp(\mathcal{E}^i) \;\underset{(\dagger)}{=}\; 0 + 0,$$

where $(*)$ is because $\rho^i = \widetilde{\rho}^i + \rho^i_\perp$, while $(\dagger)$ is because $\widetilde{\rho}^i(\mathcal{E}^i) = 0$ and $\rho^i_\perp(\mathcal{E}^i) \leqslant \rho^i_\perp(\mathcal{B}) = 0$. Finally, $\rho^J(\mathcal{E}^J) = \rho^J(\mathcal{C}) = 0$, as already noted. Thus, the partition $\{\mathcal{E}^j\}_{j=1}^J$ makes the collection $\{\rho^j\}_{j\in\mathcal{J}}$ singular, contradicting the hypothesis of the theorem. ◇ Claim 1

Given this claim, we can apply the induction hypothesis to construct some probability measure $\eta \in \Delta(\mathcal{S}^{\mathbb{N}})$ such that $\eta \ll \widetilde{\rho}^i$ for all $i \in [1..J)$. For any $i \in [1..J)$, we have $\widetilde{\rho}^i \ll \rho^i$ and thus, $\eta \ll \rho^i$. Meanwhile, $\widetilde{\rho}^i \ll \rho^J$, and thus, $\eta \ll \rho^J$. Thus, $\eta \ll \rho^j$ for all $j \in \mathcal{J}$. □

References

Aliprantis, C.D., Border, K.C., 2006. Infinite Dimensional Analysis: A Hitchhiker's Guide, 3rd edition. Springer, Berlin.

Alon, S., Gayer, G., 2016. Utilitarian preferences with multiple priors. Econometrica 84 (3), 1181–1201.

Baccelli, J., 2017. Do bets reveal beliefs? Synthese 194 (9), 3393–3419.

Bhaskara Rao, K., Bhaskara Rao, M., 1983. Theory of Charges: A Study of Finitely Additive Measures. Academic Press.

Billot, A., Qu, X., 2021. Utilitarian aggregation with heterogeneous beliefs. Am. Econ. J. Microecon. 13 (3), 112–123.

Billot, A., Vergopoulos, V., 2016. Aggregation of Paretian preferences for independent individual uncertainties. Soc. Choice Welf. 47 (4), 973–984.

Blackwell, D., Dubins, L., 1962. Merging of opinions with increasing information. Ann. Math. Stat. 33 (3), 882–886.

Blume, L.E., Cogley, T., Easley, D.A., Sargent, T.J., Tsyrennikov, V., 2018. A case for incomplete markets. J. Econ. Theory 178, 191–221.

Brandl, F., 2021. Belief-averaging and relative utilitarianism. J. Econ. Theory 198, 105368.

Brunnermeier, M., Simsek, A., Xiong, W., 2014. A welfare criterion for models with distorted beliefs. Q. J. Econ. 129 (4), 1753–1797.

Ceron, F., Vergopoulos, V., 2019. Aggregation of Bayesian preferences: unanimity vs monotonicity. Soc. Choice Welf. 52 (3), 419–451.

Chambers, C., Hayashi, T., 2006. Preference aggregation under uncertainty: Savage vs. Pareto. Games Econ. Behav. 54, 430–440.

Chambers, C., Hayashi, T., 2014. Preference aggregation with incomplete information. Econometrica 82 (2), 589–599.

Clemen, R.T., Winkler, R.L., 2007. Aggregating probability distributions. In: Edwards, W., Miles, R., von Winterfeldt, D. (Eds.), Advances in Decision Analysis. Cambridge University Press, Cambridge, UK, pp. 154–176.

Conway, J.B., 1990. A Course in Functional Analysis, 2nd edition. Graduate Texts in Mathematics, vol. 96. Springer-Verlag, New York.

Danan, E., Gajdos, T., Hill, B., Tallon, J.-M., 2016. Robust social decisions. Am. Econ. Rev. 106 (9), 2407–2425.

Desai, N., Critch, A., Russell, S.J., 2018. Negotiable reinforcement learning for Pareto optimal sequential decision-making. In: Advances in Neural Information Processing Systems, pp. 4712–4720.

Diaconis, P., Freedman, D., 1986. On the consistency of Bayes estimates. Ann. Stat., 1–26.

Diaconis, P., Freedman, D., 1990. On the uniform consistency of Bayes estimates for multinomial probabilities. Ann. Stat., 1317–1327.

Dietrich, F., 2021. Fully Bayesian aggregation. J. Econ. Theory 194, 105255.

Dietrich, F., List, C., 2016. Probabilistic opinion pooling. In: Hájek, A., Hitchcock, C. (Eds.), The Oxford Handbook of Probability and Philosophy. Oxford UP, pp. 519–544. Ch. 25.

Duffie, D., 2014. Challenges to a policy treatment of speculative trading motivated by differences in beliefs. J. Leg. Stud. 43 (S2), S173–S182.

Dunford, N., Schwartz, J.T., 1958. Linear Operators Part i: General Theory, vol. 243. Interscience Publishers, New York.

Fleurbaey, M., 2018. Welfare economics, risk and uncertainty. Can. J. Econ. 51 (1), 5–40.

Gayer, G., Gilboa, I., Samuelson, L., Schmeidler, D., 2014. Pareto efficiency with different beliefs. J. Leg. Stud. 43 (S2), S151–S171.

Genest, C., Zidek, J.V., 1986. Combining probability distributions: a critique and an annotated bibliography. Stat. Sci. 1 (1), 114–148.

Gilboa, I., Samet, D., Schmeidler, D., 2004. Utilitarian aggregation of beliefs and tastes. J. Polit. Econ. 112, 932–938.

Gilboa, I., Samuelson, L., Schmeidler, D., 2014. No-betting Pareto dominance. Econometrica 82, 1405–1442.

Hanson, R., 2013. Shall we vote on values, but bet on beliefs? J. Polit. Philos. 21, 151–178.

Harsanyi, J.C., 1955. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. J. Polit. Econ. 63, 309–321.

Hayashi, T., Lombardi, M., 2019. Fair social decision under uncertainty and responsibility for beliefs. Econ. Theory 67 (4), 775–816.

Hild, M., Jeffrey, R., Risse, M., 2003. Flipping and ex post aggregation. Soc. Choice Welf. 20, 267–275.

Hild, M., Jeffrey, R., Risse, M., 2008. Preference aggregation after Harsanyi. In: Fleurbaey, M., Salles, M., Weymark, J.A. (Eds.), Justice, Political Liberalism, and Utilitarianism: Themes from Harsanyi and Rawls. Cambridge UP, pp. 198–217.

Kalai, E., Lehrer, E., 1994. Weak and strong merging of opinions. J. Math. Econ. 23 (1), 73–86.

Kreps, D.M., Porteus, E.L., 1978. Temporal resolution of uncertainty and dynamic choice theory. Econometrica, 185–200.

Lehrer, E., Smorodinsky, R., 1996a. Compatible measures and merging. Math. Oper. Res. 21 (3), 697–706.

Lehrer, E., Smorodinsky, R., 1996b. Merging and learning. In: Statistics, Probability and Game Theory. In: Institute of Mathematical Statistics Lecture Notes—Monograph Series, vol. 30. Institute of Mathematical Statistics, Hayward, CA, pp. 147–168.

List, C., Pivato, M., 2015. Emergent chance. Philos. Rev. 124 (1), 119–152.

Miller, R.I., Sanchirico, C.W., 1999. The role of absolute continuity in "merging of opinions" and "rational learning". Games Econ. Behav. 29 (1–2), 170–190.

Mongin, P., 1995. Consistent Bayesian aggregation. J. Econ. Theory 66, 313–351.

Mongin, P., 1997. Spurious unanimity and the Pareto principle. Tech. Rep. THEMA, Université de Cergy-Pontoise. published as Mongin (2016).

Mongin, P., 1998. The paradox of the Bayesian experts and state-dependent utility theory. J. Math. Econ. 29, 331–361.

Mongin, P., 2016. Spurious unanimity and the Pareto principle. Econ. Philos. 32, 511–532 (earlier circulated as Mongin (1997)).

Mongin, P., Pivato, M., 2016. Social evaluation under risk and uncertainty. In: Adler, M.D., Fleurbaey, M. (Eds.), Handbook of Well-Being and Public Policy. Oxford University Press, pp. 711–745. Ch. 24.

Mongin, P., Pivato, M., 2020. Social preference under twofold uncertainty. Econ. Theory 70 (3), 633–663.

Morris, P.A., 1974. Decision analysis expert use. Manag. Sci. 20 (9), 1233–1241.

Petersen, K., 1989. Ergodic Theory. Cambridge University Press, New York.

Pivato, M., 2013. Voting rules as statistical estimators. Soc. Choice Welf. 40 (2), 581–630.

Pivato, M., 2017. Epistemic democracy with correlated voters. J. Math. Econ. 72, 51–69.

Posner, E., Weyl, G., 2013. FDA for financial innovation: applying the insurable interest doctrine to twenty-first-century financial markets. Northwest. Univ. Law Rev. 107 (3), 1307.

Qu, X., 2017. Separate aggregation of beliefs and values under ambiguity. Econ. Theory 63 (2), 503–519.

Raiffa, H., 1968. Decision Analysis: Introductory Lectures on Choices Under Uncertainty. Addison-Wesley Series in Behavioral Science. Addison-Wesley.

Risse, M., 2001. Instability of ex post aggregation in the Bolker-Jeffrey framework and related instability phenomena. Erkenntnis 55 (2), 239–270.

Risse, M., 2003. Bayesian group aggregation and two modes of aggregation. Synthese 135, 347–377.

Savage, L.J., 1954. The Foundations of Statistics. John Wiley & Sons, New York.

Schervish, M., Seidenfeld, T., 1990. An approach to consensus and certainty with increasing evidence. J. Stat. Plan. Inference 25 (3), 401–414.

Sprumont, Y., 2018. Belief-weighted Nash aggregation of Savage preferences. J. Econ. Theory 178, 222–245.

Sprumont, Y., 2019. Relative utilitarianism under uncertainty. Soc. Choice Welf. 53 (4), 621–639.

Weymark, J., 1991. A reconsideration of the Harsanyi-Sen debate on utilitarianism. In: Elster, J., Roemer, J. (Eds.), Interpersonal Comparisons of Well-Being. Cambridge University Press, Cambridge, pp. 255–320.

Zuber, S., 2016. Harsanyi's theorem without the sure-thing principle: on the consistent aggregation of Monotonic Bernoullian and Archimedean preferences. J. Math. Econ. 63, 78–83.