THE IMPARTIAL OBSERVER UNDER UNCERTAINTY

STEFAN BERENS AND LASHA CHOCHUA[†]

THE LATEST VERSION

Abstract. This paper extends Harsanyi's Impartial Observer Theorem by introducing Knightian Uncertainty in the form of individual belief systems. It features an axiomatic framework of societal decision-making in the presence of individual uncertainty. The model allows the analysis of scenarios where individuals agree on the ranking but not on the likelihood of social outcomes. The preferences of the impartial observer are expressed by a weighted sum of utilities - each representing individual preferences with different belief systems. To incorporate common criticism of the framework of Harsanyi (1953), our approach is based on the generalized version by Grant et al. (2010). The belief systems are introduced as second-order beliefs following Seo (2009).

Keywords: Impartial Observer, Uncertainty, Utilitarianism

JEL Classification: D71, D81

[†]Kiel Institute for the World Economy.

The following people supported and influenced this paper through various fruitful discussions: Herbert Dawid, Frank Riedel, Gerald Willmann, Jan-Henrik Steg, Thibault Gajdos, Simon Grant, and Luca Rigotti.

1. Introduction

As members of society, individuals are constantly confronted with a myriad of choices concerning moral rules, institutional arrangements, government policies, and wealth distribution patterns. Consequently, they engage in ongoing value judgments aimed at selecting the most appropriate social alternatives. Among different viewpoints, one particular perspective argues for the adoption of decisions guided by a sympathetic interest in the welfare of every member of society, free from any bias favoring specific participants. Among economists, notably, Adam Smith's seminal work in his "Theory of Moral Sentiments" (1759) serves as a foundation for this perspective.

Harsanyi (1953, 1955, 1977) developed a rational theory of societal judgments. According to this theory, such choices should be made based on individuals' 'social' or 'moral' preferences derived from the concept of an 'impartial observer'. As such, you imagine a situation where you do not know your actual place in society when comparing different social arrangements. Instead, you judge the desirability of the alternatives under the personal preferences of all community members.[1] Thus, the original premise remains that this type of theory, unlike the theory of individual rational behavior or game theory, should be independent of selfish considerations.[2]

The main result of this theory, now known as Harsanyi's Impartial Observer Theorem, combines Adam Smith's ideas of a sympathetic and impartial spectator with Kant's universality criterion and the utilitarian tradition of social utility maximization using von Neumann-Morgenstern expected utility theory. In the end, Harsanyi argued that an individual facing risky prospects over social outcomes and a hypothetical lottery over identities in society should rank these according to the weighted aggregate of individuals' expected utilities.

However, Harsanyi's Impartial Observer Theorem, with its implicit utilitarianism, only considers scenarios where each involved individual faces objective risk. It is a theory analyzing societal judgments when each member knows objective probabilities over a set of social outcomes of society. This paper extends Harsanyi's Impartial Observer Theorem to include Knightian Uncertainty in the model. By introducing individual belief systems about the likelihood of the social outcomes (which the impartial observer necessarily considers), our approach allows the application of the original framework

[1]The imaginary construct of impartiality is similar to John Rawls' idea of a 'veil of ignorance' in 'A Theory of Justice' (1971), as these two metaphors are attempts at capturing the same stance.

[2]This understanding of the significance of value judgments extends beyond conventional decision-making domains and holds particular relevance within the realm of artificial intelligence (AI) systems, such as self-driving cars as an example. These sophisticated technological systems require meticulously crafted programmed mechanisms that can navigate complex ethical scenarios.

to a new area of social value judgments. In particular, it will enable the analysis of scenarios where individuals agree on the ranking but not on the likelihood of social outcomes. The impartial observer in our model does not aggregate these belief systems separately, though. When the impartial observer imagines herself being a particular individual, she adopts not only that individual's preferences but the belief system as well.[3]

The main result of our paper is a generalized utilitarian representation of the preferences of the impartial observer under uncertainty. It is a weighted sum of Second-Order Subjective Expected Utility (SOSEU) functions, each representing an individual's preferences. Our framework is based on the generalized version of Harsanyi (1953) by Grant et al. (2010), which accommodates common criticism of Harsanyi's approach, specifically the issue of fairness and attitude towards mixing (risk). The introduction of individual belief systems to our framework follows Seo (2009). As a result, the SOSEU functions supersede the Expected Utility (EU) functions of both Harsanyi's and Grant et al.'s approaches.

According to Fleurbaey (2018), acknowledging the potential emotional responses triggered by ambiguity among individuals might be crucial. However, the social evaluator's objective is to rationalize these emotions without necessarily adopting an ambiguity-averse decision-making approach at a societal level. Our paper meets this conceptual criterion through the presentation of a framework that effectively incorporates individuals' attitudes towards ambiguity. In doing so, we ensure that an impartial observer can address these attitudes.

In addition to the framework, our paper includes two illustrative examples. First, the moral dilemma of the 'Afghan Goatherds' (see Sandel (2010)) showcases a scenario of agreement on the ranking but disagreement on the likelihood of social outcomes. Second, an example of a simple exchange economy with endowments and different alternatives of wealth (re-)distribution demonstrates the framework's suitability for traditional economic problems - it serves as a proof of concept in that regard.

To motivate the introduction of uncertainty to moral value judgments, let us preview the story of the Afghan Goatherds. In 2005 a team of four soldiers, all U.S. Navy SEALs, set out to find a Taliban leader in the Afghan mountains near the Pakistan border. Just as the team set up their base overlooking the area to fulfill their reconnaissance mission, two Afghan goatherds stumbled upon them - a young boy with them. Due to the nature of their mission and other circumstances, the team

[3]In our opinion, this is a natural extension of Harsanyi's concept of impartiality. Therefore, it differs from the group decisions presented in Raiffa (1970), which features and discusses the aggregation of belief systems.

considered killing or releasing the civilians the only two viable options. Eventually, the group cast a vote, where one soldier abstained, two voted two ways, and the unit commander made the decisive call to release them. The civilians later informed the Taliban in a nearby village about the presence of the soldiers. Three of the four soldiers died in the subsequent ambush, leaving the commander as the lone survivor.

In retrospect, it is easy to make the correct call for this specific scenario. However, imagine you wanted to create a guideline for commanders about making such moral value judgments when in the field. In that case, you would naturally assume the role of an impartial observer and evaluate the situation based on individual preferences. To truly accept a person's view, it is also necessary to take on that person's belief system - which is not possible (or included) in the traditional setting. Therefore, our approach extends the framework to include cases where belief systems play an essential role. This paper also contains a formal presentation of this specific moral dilemma and connects the model with reality.

Our paper is related to the literature that uses impartiality's essential philosophical tradition for moral value judgments about collective life. Vickrey (1945) and Harsanyi (1953) both independently introduced the idea to the economic literature, but as Mongin (2001) formulates it: 'All in all, Harsanyi, if perhaps not Vickrey, should count as a major representative of the ethics of impartiality among 20th-century writers.' In addition, our work is related to the literature on decision-making under ambiguity. Both streams of literature are substantial. Our aim is not to review this vast literature, but instead, we concentrate on the building blocks of our work and other closely related research.

As previously said, Grant et al. (2010) and Seo (2009) are the inspiration for the building blocks of our framework. In the first one, the authors revisit Harsanyi's Impartial Observer Theorem; they consider two significant criticisms concerning fairness and different risk attitudes and derive a generalized version of the theorem that accommodates these criticisms. Furthermore, in the particular case of an impartial observer indifferent between identity and outcome lotteries ('accidents of birth' and 'life chances'), the generalized version of the theorem boils down to the standard Harsanyi doctrine. In consequence, the setting of the paper yields a new axiomatization of Harsanyi's utilitarianism. The resulting generalized utilitarianism serves as inspiration for the foundation of our approach. It gives us the possibility to extend the original framework of Harsanyi from risk to uncertainty while also accommodating common criticism of it.

In the second paper mentioned above, Seo formulates a model for decision-making under uncertainty using second-order beliefs, i.e., beliefs over probability measures. Existing models in this stream of literature essentially differ in choosing the domain of preference. Seo takes the domain of Anscombe and Aumann (1963) and a similar axiomatic foundation. Klibanoff et al. (2005), by contrast, require an additional (sub-)domain with preferences. The domain selection of Seo allows us to introduce uncertainty to the (generalized) framework without any other modifications. The Second-Order Subjective Expected Utility (SOSEU) representation of the preferences by Seo, therefore, translates to a corresponding version in the context of an impartial observer.

A variety of papers already deals with Harsanyi's Impartial Observer Theorem under uncertainty. Gajdos and Kandil (2008) extend the framework where the impartial observer considers sets of identity lotteries. In their model, unlike ours, uncertainty is introduced on a societal (not individual) level. The impartial observer's preference (under additional assumptions) is then characterized by a convex combination of Harsanyi's utilitarian and Rawls' egalitarian criteria. Billot and Vergopoulos (2016) develop a framework where individual uncertainties are potentially different and independent of each other. The society represents social uncertainty through the Cartesian product of individual state spaces. Preference aggregation has two components: social utility, the convex combination of individual utilities, and social probability, the independent product of individual subjective probabilities. Contrary to Billot and Vergopoulos (2016), we do not aggregate individual utilities and beliefs separately.[4]

Work closer to ours is Nascimento (2012), which presents a model of aggregating preference orderings under subjective uncertainty. A fundamental difference is the setting of each paper. Namely, the one of Nascimento is that of a group of individuals that necessarily agree on the ranking of particular risky objects.[5] In contrast, our setting is one where a group of individuals does not agree on ranking any objects (risky or ambiguous). The assumption of Nascimento fits a group of experts or specialists in a field where there is a certain consensus. However, in our opinion, this assumption is too restrictive for other cases, like the economic example in this paper (see Section 4 for a formal discussion). Nevertheless, the results are closely related, compare specifically, Theorem 1 by Nascimento and Theorem 1 in this paper. In a sense, our work arrives at an equal representation but with a different axiomatic foundation and applications in mind that are explicitly excluded otherwise. Furthermore,

[4]There are several papers that deal with the decision-making of societies using various mechanisms of aggregating different individual beliefs. See, for example, Crès et al. (2011), Alon and Gayer (2016), Danan et al. (2016), and Qu (2017).

[5]See also Stanca (2021). He develops an axiomatic model of preference aggregation where agents (Bayesian experts) have a common utility function but different beliefs.

as the analysis of the example of the Afghan Goatherds shows, our framework is also able to include scenarios with consent (see Section 4 and Appendix A).

It is worth mentioning (and repeating) that the impartial observer in our model always takes on individual beliefs as part of the preferences, thereby avoiding any aggregation of belief systems. In consequence, our model stays true to Harsanyi's thought experiment and also avoids the impossibility result of Mongin (1995).

The paper is organized as follows. Section 2 presents a minimal version of a framework based on Grant et al. (2010) and then introduces uncertainty following Seo (2009). Section 3 provides an analysis of the model, including a comparison with Grant et al. (2010). Section 4 contains the aforementioned illustrative examples as it revisits the Afghan Goatherds and presents the economic example. Section 5 summarizes and concludes our paper. The appendix consists of some additional details for the example of the Afghan Goatherds.

2. Model

Let (X, τ) be a topological space. Then, denote by $\mathcal{B}_X$ its Borel σ-algebra and by $\Delta(X)$ the set of all probability measures on $(X, \mathcal{B}_X)$. By x for $x \in X$ refer both to the actual element in X and the induced one in $\Delta(X)$ - depending on context. Endow $\Delta(X)$ with the weak convergence topology. Also, endow any product of topological spaces with the product topology.

2.1. **General Setting.** Let $\mathcal{I} = \{i_1, \ldots, i_I\}$, $I \geq 2$, be a finite set of individuals facing a societal decision problem in the presence of individual uncertainty. Each social choice is modeled as a three-layered object (of different types of risk).[6] First, the (final) outcome space is given by $\mathcal{X}$ - a compact metrizable space with $|\mathcal{X}| \geq 2$. The outcome lotteries $p \in \Delta(\mathcal{X})$, also called one-stage lotteries, are the first layer (featuring objective risk). Further, let $\mathcal{S} = \{s_1, \ldots, s_S\}$, be the finite set of states of the world, which introduces uncertainty via individual beliefs about its probability distribution. The functions $h \colon \mathcal{S} \to \Delta(\mathcal{X})$, also called acts, are the second layer (featuring subjective risk or simply ambiguity). Denote by $\mathcal{H}$ the set of all acts. The act lotteries $P \in \Delta(\mathcal{H})$, also called two-stage lotteries, are the third layer (featuring objective risk again).

The individuals in this situation imagine themselves as impartial observer, i.e., treating their (social) identity as an unknown component in the decision problem. As such, they face both identity lotteries $z \in \Delta(\mathcal{I})$ and act lotteries $P \in \Delta(\mathcal{H})$. Thus, the individual preferences $\succeq_i$, $i \in \mathcal{I}$, are

[6] The introduction of uncertainty via these three-layered objects follows Seo (2009) and, by extension, Anscombe and Aumann (1963).

defined on $\Delta(\mathcal{H})$ while that of the impartial observer $\succeq$ is defined on $\Delta(\mathcal{I}) \times \Delta(\mathcal{H})$.[7] For all of these preferences, we assume a couple of 'standard' properties:

Assumption 1 (Individual)**.** *For each i in $\mathcal{I}$ the preference $\succeq_i$ on $\Delta(\mathcal{H})$ is complete, transitive and continuous. Its asymmetric part $\succ_i$ is non-empty.*

Assumption 2 (Impartial Observer)**.** *The preference $\succeq$ on $\Delta(\mathcal{I}) \times \Delta(\mathcal{H})$ is complete, transitive and continuous. Its asymmetric part $\succ$ is non-empty*

Note that by continuous we mean that the weak upper and lower contour sets are closed with respect to the corresponding topologies. In the case of the individual, this means with respect to the weak convergence topology, and in the case of the impartial observer, the product topology of the weak convergence topologies.

Axiom 1 (Acceptance Principle)**.** *For all i in $\mathcal{I}$ and all P, Q in $\Delta(\mathcal{H})$:*

$$P \succeq_i Q \Leftrightarrow (i, P) \succeq (i, Q)$$

The acceptance principle establishes the intuitive link between the preferences of the individuals and that of the impartial observer. The intuition is that when the impartial observer imagines herself to be a particular individual, she takes on the preferences of that individual (including the belief system).

Axiom 2 (Independence over Identity Lotteries)**.** *Suppose elements $(z, P), (z', Q)$ in $\Delta(\mathcal{I}) \times \Delta(\mathcal{H})$ are such that $(z, P) \sim (z', Q)$. Then, for all $\tilde{z}, \tilde{z}'$ in $\Delta(\mathcal{I})$ and all α in $(0, 1]$:*

$$(\tilde{z}, P) \succeq (\tilde{z}', Q) \Leftrightarrow (\alpha\tilde{z} + (1-\alpha)z, P) \succeq (\alpha\tilde{z}' + (1-\alpha)z', Q)$$

The independence over identity lotteries and the acceptance principle are each concerned with the nature of the impartial observer's preferences relating to identities. As our approach considers uncertainty on the level of outcomes and not identities, these two axioms naturally carry over from the traditional setting.

[7] The impartiality that is presented here is based on the framework of Grant et al. (2010), which generalized the concept of Harsanyi (1953). In Grant et al. (2010) the impartial observer's preferences are defined on $\Delta(\mathcal{I}) \times \Delta(\mathcal{X})$, which naturally extends to $\Delta(\mathcal{I}) \times \Delta(\mathcal{H})$ - incorporating the framework of Seo (2009). By contrast, the corresponding set in Harsanyi (1953) is $\Delta(\mathcal{I} \times \mathcal{X})$. See Grant et al. (2010) for a detailed discussion on this difference.

Assumption 3 (Absence of Unanimity)**.** *For all P, Q in $\Delta(\mathcal{H})$:*

$$\exists i \in \mathcal{I} : P \succ_i Q \Rightarrow \exists j \in \mathcal{I} : Q \succ_j P$$

The absence of unanimity can be interpreted as a required heterogeneity in the social alternatives and (preferences of) individuals. It is also not a new addition but controversial enough to require additional comment. First, normative decision-making is trivial when all individuals agree on all rankings. Thus, excluding this extreme case without losing any explanatory power is possible. However, in our opinion, it is too restrictive to completely leave out the opposite where everyone disagrees about everything - as it is done in Nascimento (2012) with the requirement of agreement on risky prospects. In general, we aim to focus on scenarios that exhibit substantial heterogeneity in terms of (dis-)agreement.[8]

Next, let us state a lemma (which will be useful later on) about the representation of the preferences of the impartial observer and the individuals. Now, the structure of the results and the proof itself follow the ideas of Grant et al. (2010):

Lemma 1. *Suppose the absence of unanimity applies. Then, the impartial observer satisfies the acceptance principle and independence over identity lotteries if and only if there exists a continuous function $V\colon \Delta(\mathcal{I}) \times \Delta(\mathcal{H}) \to \mathbb{R}$ that represents $\succeq$ and for each individual i in $\mathcal{I}$ a function $V_i\colon \Delta(\mathcal{H}) \to \mathbb{R}$ that represents $\succeq_i$ such that for all (z, P) in $\Delta(\mathcal{I}) \times \Delta(\mathcal{H})$:*

$$V(z, P) = \sum_{i \in \mathcal{I}} z_i V_i(P)$$

Moreover, the functions V and V_i, $i \in \mathcal{I}$, are unique up to common positive affine transformation.

Proof. To employ Lemma 8 of Grant et al. (2010), it is necessary to show that the set $\mathcal{H}$ is compact and metrizable. For any compact topological space (X, τ), the set $\Delta(X)$ is always compact with the week convergence topology. Therefore, as $\mathcal{X}$ is compact by the initial assumption, $\Delta\mathcal{X}$ is compact as well. Any finite Cartesian product of compact spaces is compact with the product topology. Thus, with $\Delta\mathcal{X}$ compact, $\mathcal{H} = \Delta\mathcal{X}^S$ is compact too.

[8] It might seem that absence of unanimity is too restrictive as well (just in the other direction). Yet, adding a dummy individual that provides the (technically) required heterogeneity allows us to relax the restriction while still staying in our framework. See Section 4 and Appendix A for the formal presentation of the story of the Afghan Goatherds with a demonstration of a dummy.

Using the fact that $\mathcal{X}$ is compact and metrizable (implying separable), it follows that $\Delta\mathcal{X}$ is metrizable - for example, using the Lévi-Prokhorov metric. Furthermore, by combining the metrices of the product, for instance, with a p-norm, $1 < p$, $\mathcal{H} = \Delta\mathcal{X}^S$ is metrizable as desired.

□

Note that the main arguments of the proof work for a general compact set $\mathcal{H}$ as well, i.e., the proof requires no specific structure of $\mathcal{H}$. Thus, a modified Lemma 1 could potentially serve as a foundation for conceptually similar approaches to ours that only differ in terms of additional structure, specifically for individual utilities.

Now, up to this point, all assumptions and axioms follow Grant et al. (2010). Specifically, their axiom of independence over outcome lotteries (for individuals) is the only missing axiom. However, in the next part, the axioms follow Seo (2009) instead and introduce uncertainty to the framework.

2.2. **Introducing Uncertainty.** To formulate the remaining axioms and introduce uncertainty to the model, it is necessary to define additional objects. Namely, let us explain what we mean when talking about mixing two acts or lotteries (of acts). In the end, the two different kinds of mixtures depend on the timing of the resolution of uncertainty (or of the mixing - depending on how you look at it).

First, consider the case where, when combining two (pure) acts, the uncertainty is resolved first, and then the mixing takes place:

Definition 1. *For f, g in $\mathcal{H}$ and α in $[0,1]$ and for $s \in S$, $B \in \mathcal{B}_\mathcal{X}$ we set*

$$(\alpha f \oplus (1-\alpha)g)(s)(B) = \alpha f(s)(B) + (1-\alpha)g(s)(B).$$

This operation is called a second-stage mixture.

Now, with this in mind, we introduce a 'standard' independence axiom with respect to second-stage mixtures:

Axiom 3 (Second-Stage Independence)**.** *For all i in $\mathcal{I}$, all α in $(0,1]$ and lotteries p, q, r in $\Delta(\mathcal{X})$:*

$$\alpha p \oplus (1-\alpha)r \succeq_i \alpha q \oplus (1-\alpha)r \Leftrightarrow p \succeq_i q$$

Second, consider the case where, when combining two lotteries of acts, the mixing takes place first, and then the uncertainty is resolved:

Definition 2. *For P, Q in $\Delta(\mathcal{H})$ and α in $[0,1]$ and for $B \in \mathcal{B}_{\mathcal{H}}$ we set*

$$(\alpha P + (1-\alpha)Q)(B) = \alpha P(B) + (1-\alpha)Q(B).$$

This operation is called a first-stage mixture.

Again, with this in mind, we introduce a 'standard' independence axiom with respect to first-stage mixtures:

Axiom 4 (First-Stage Independence)**.** *For all i in $\mathcal{I}$, all α in $(0,1]$ and lotteries P, Q, R in $\Delta(\mathcal{H})$:*

$$\alpha P + (1-\alpha)R \succeq_i \alpha Q + (1-\alpha)R \Leftrightarrow P \succeq_i Q$$

Finally, it is necessary to introduce an additional technical object that essentially serves as a tool for scenario analysis (for the individuals):

Definition 3. *Each $f \in \mathcal{H}$ and $\mu \in \Delta(\mathcal{S})$ induce a one-stage lottery*

$$\Psi(f, \mu) := \bigoplus_{s \in \mathcal{S}} \mu(s) f(s),$$

each $P \in \Delta(\mathcal{H})$ and $\mu \in \Delta(\mathcal{S})$ induce a two-stage lottery

$$\Psi(P, \mu)(B) := P(\{f \in \mathcal{H} \colon \Psi(f, \mu) \in B\})$$

for $B \in \mathcal{B}_{\mathcal{H}}$.

In other words, the element $\Psi(P, \mu)$ is the (induced) lottery that corresponds to the lottery P in the scenario where μ is the probability distribution over the states.

Axiom 5 (Dominance)**.** *For all i in $\mathcal{I}$, all P, Q in $\Delta(\mathcal{H})$:*

$$\Psi(P, \mu) \succeq_i \Psi(Q, \mu)\ \forall \mu \in \Delta(\mathcal{S}) \Rightarrow P \succeq_i Q$$

Imagine an individual only knows there exists a 'true' probability distribution but does not know which one it is. Then, the axiom of dominance captures the intuition that if the individual prefers one induced lottery over another one - substituting all possible probability distributions as the true one, then this individual should prefer that one lottery over the other (and vice-versa).

Finally, using the additional axioms with Lemma 1, it is possible to formulate the main result of our paper:

Theorem 1. *Suppose absence of unanimity applies. Then, the impartial observer satisfies the acceptance principle as well as independence over identity lotteries, and each individual satisfies first-stage and second-stage independence, and dominance if and only if the impartial observer's preference admits a representation* $\langle \{U_i, \phi_i\}_{i \in \mathcal{I}} \rangle$ *of the form*

$$V(z, P) = \sum_{i \in \mathcal{I}} z_i \phi_i(U_i(P))$$

where each

- $\phi_i \colon \mathbb{R} \to \mathbb{R}$ *is an increasing continuous function*
- $U_i \colon \Delta(\mathcal{H}) \to \mathbb{R}$ *is a SOSEU representation of* $\succeq_i$

i.e. the impartial observer is a generalized (weighted) utilitarian under uncertainty.

In addition, the U_i *are unique up to the uniqueness of the SOSEU representation. Further, the functions* V *and* $\phi_i \circ U_i$ *are unique up to a common positive affine transformation.*

Proof. Let the absence of unanimity apply.

Part 1 ('$\Leftarrow$'):

First, the representation of the impartial observer is affine in identity lotteries and therefore satisfies the acceptance principle and independence over identity lotteries. Note that alternatively, this specific step also follows by application of Lemma 1. Second, the representation of each individual is of SOSEU form and thus satisfies its axioms, that is, first-stage and second-stage independence, and dominance, using Theorem 4.2 of Seo (2009).

Part 2 ('$\Rightarrow$'):

In this part, the result of Lemma 1 is required. Namely, as the impartial observer satisfies the acceptance principle and independence over identity lotteries, it gives us a continuous function $V \colon \Delta(\mathcal{I}) \times \Delta(\mathcal{H}) \to \mathbb{R}$ representing $\succeq$ and for each $i \in \mathcal{I}$ functions $V_i \colon \Delta(\mathcal{H}) \to \mathbb{R}$ representing $\succeq_i$ such that for all $(z, P) \in \Delta(\mathcal{I}) \times \Delta(\mathcal{H})$:

$$V(z, P) = \sum_{i \in \mathcal{I}} z_i V_i(P)$$

The preferences of each individual satisfy first-stage and second-stage independence and dominance, and so by Theorem 4.2 of Seo (2009), each V_i is a SOSEU function. However, this only holds up to transformation via an increasing continuous function. Thus, for each $i \in \mathcal{I}$ it follows that $V_i = \phi_i \circ U_i$ where U_i is a SOSEU function and ϕ_i is a transformation.

The uniqueness of the U_i follows directly from Lemma C.1 in Seo (2009), while the uniqueness of the V and $\phi_i \circ U_i$ follows from Lemma 1. □

In the theorem and its proof, the actual SOSEU representation remained hidden. The following remark (re-)states the formal definition of a SOSEU representation and its uniqueness properties (see Seo (2009)). It will be helpful in the analysis and for the applications later on.

Remark 1. *A SOSEU representation is generally characterized by a triple (u, v, m) and of the form*

$$U(P) = \int_{\mathcal{H}} \int_{\Delta(\mathcal{S})} v\left(\int_{\mathcal{S}} u(f)d\mu\right) dm(\mu)dP(f)$$

for $P \in \Delta(\mathcal{H})$, where

- *u is a bounded continuous mixture linear function, $u\colon \Delta(\mathcal{X}) \to \mathbb{R}$,*
- *v is a bounded continuous strictly increasing function, $v\colon u(\Delta(\mathcal{X})) \to \mathbb{R}$,*
- *m is a probability measure, $m \in \Delta(\Delta(\mathcal{S}))$.*

The uniqueness of the SOSEU representation implies that for any another triple (u', v', m')

i) *u and u' as well as $v \circ u$ and $v' \circ u'$ are each identical up to positive affine transformation,*

ii) *$\int_{\Delta(\mathcal{S})} \varphi dm = \int_{\Delta(\mathcal{S})} \varphi dm'$ for all continuous functions φ on $\Delta(\mathcal{S})$ for which there exists a Borel signed measure λ on $T := [u(\Delta(\mathcal{X}))]^S$ with bounded variation such that for all $\mu \in \Delta(\mathcal{S})$ it holds that $\varphi(\mu) = \int_T v(\mu \cdot t)d\lambda(t)$.*

3. Analysis

Naturally, the starting point of our analysis is the connection between our result and that of Grant et al. (2010). As expected, eliminating uncertainty by reducing it to risk produces their result as a special case of ours (see 3.1).

Furthermore, one (significant) indeterminacy in Theorem 1 is the specific form of the ϕ_i, $i \in \mathcal{I}$. In the current setting, the only restriction is that each of them is an increasing continuous function. Let us, therefore, analyze the specific form of these functions in relation to the preferences' different (additional) properties. In particular, let us compare this to the findings presented in Grant et al.

(2010). Fortunately, their results and proofs do not rely on the underlying structure of the outcome space, and therefore all of their findings translate into our setting without any modifications.

3.1. **The Special Case of Risk.** In the special case of (only) risk, i.e., $\mathcal{S} = \{s\}$, all belief systems are trivial. Consequently, the three-layer objects containing uncertainty in the middle reduce to two-layer objects with only risk present. Further, to identify this with the setting of Grant et al. (2010), assume that each v_i, $i \in \mathcal{I}$, is a linear function (corresponding to indifference to uncertainty) or equivalently assume reversal of order or reduction of compound lotteries.[9] Thus, the two-layer objects with risk collapse to single-layer objects with risk.

In this (particular case of a) setting, first/second-stage independence simply reduces to independence over outcome lotteries, dominance is now an empty statement, and the representation of the impartial observer takes on the form of the generalized (weighted) utilitarian of Grant et al. (2010).

3.2. **Fairness.** One common criticism of Harsanyi's utilitarianism is concerned with fairness (see, for example, Diamond (1967)). It is one of the two issues that Grant et al. (2010) solved by generalizing the theory. The notion of fairness in this context refers to a preference of the impartial observer for mixing act lotteries over mixing identity lotteries.[10]

Consider from now on those tuples of identity and act lotteries between which the impartial observer is indifferent, that is (z, P') and (z', P) with $(z, P') \sim (z', P)$. Following the arguments in favor of fairness, the impartial observer always prefers mixing these pairs on the level of acts over mixing on the level of identities, i.e., $(z, \alpha P + (1-\alpha)P') \succeq (\alpha z + (1-\alpha)z', P)$ for all $\alpha \in (0,1)$, which is also referred to as a preference for life chances (compared to accidents of birth). In the case of risk, Grant et al. (2010) show that this holds if and only if each ϕ_i, $i \in \mathcal{I}$, is concave, which translates one-to-one into our setting.

Conversely, consider a scenario where the impartial observer is indifferent between life chances and accidents of birth, that is $(z, \alpha P + (1-\alpha)P') \sim (\alpha z + (1-\alpha)z', P)$ for all $\alpha \in (0,1)$. In the case

[9] Take $i \in \mathcal{I}$. Then, reversal of order and reduction of compound lotteries each describe a property of the preferences of individual i with respect to first-stage and second-stage mixtures.
Namely, reversal of order is satisfied if for all $f, g \in \mathcal{H}$ and $\alpha \in [0,1]$

$$\alpha f \oplus (1-\alpha)g \sim_i \alpha f + (1-\alpha)g.$$

Similarly, reduction of compound lotteries is satisfied if for all $p, q \in \Delta(\mathcal{X})$ and $\alpha \in [0,1]$

$$\alpha p \oplus (1-\alpha)q \sim_i \alpha p + (1-\alpha)q.$$

The axiom of dominance implies the equivalence of these two properties (see Seo (2009)).

[10] The paper of Grant et al. (2010) also provides an example justifying this notion of fairness.

of risk, Grant et al. (2010) show that this holds if and only if each ϕ_i, $i \in \mathcal{I}$, is affine, which again translates one-to-one into our setting.[11]

3.3. **Mixtures.** In the last part, the focus was on the relation between mixing either identity or act lotteries. Another criticism of Harsanyi's utilitarianism is concerned with different attitudes among individuals towards mixing.[12]

Fix two individuals, say i and j, and consider from now on those act lotteries which the impartial observer ranks equally from each perspective, that is $P, \tilde{P}, Q, \tilde{Q}$ with $(i, P) \sim (j, Q)$ and $(i, \tilde{P}) \sim (j, \tilde{Q})$. Imagine that the impartial observer prefers facing the (first-stage) mixtures of each of those two pairings of act lotteries as i rather than as j, i.e. $(i, \alpha P + (1-\alpha)\tilde{P}) \succeq (j, \alpha Q + (1-\alpha)\tilde{Q})$ for all $\alpha \in (0, 1)$. Now, Grant et al. (2010) show that this holds if and only if the composite function $\phi_i^{-1} \circ \phi_j$ is convex on the domain $\mathcal{U}_{ji} := \{u \in \mathbb{R} \,|\, \exists P, Q \in \Delta(\mathcal{H})\colon (i, P) \sim (j, Q) \wedge U_j(Q) = u\}$ for the case of risk but it actually also applies to our setting under uncertainty.

Alternatively, imagine that the impartial observer is indifferent when comparing to face these mixtures as different individuals: $(i, \alpha\tilde{P} + (1-\alpha)P) \sim (j, \alpha\tilde{Q} + (1-\alpha)Q)$ for all $\alpha \in (0, 1)$ and all $i, j \in \mathcal{I}$. Following Grant et al. (2010) this holds if and only if $\phi_i = \phi$, $i \in \mathcal{I}$, both for risk and under uncertainty. Further, let i_1 and i_2 be a pair of individuals such that there exists a sequence of individuals $j_1, \ldots, j_N$ with $j_1 = i_1$ and $j_N = i_2$ where each $\mathcal{U}_{j_n, j_{n-1}}$ has non-empty interior. Then, the functions U_i, $i \in \mathcal{I}$, are unique up to a common positive affine transformation.

4. Applications

In the following, two applications (or examples) for our approach are presented. The first example picks up the story of the Afghan Goatherds from the introduction; the second one is a simple economic example.

As mentioned in the introduction, the moral dilemma of the Afghan Goatherds features a scenario where individuals agree on the ranking of social outcomes but disagree on the likelihood of these. It showcases the effect of the introduction of uncertainty, as the nature of different belief systems purely drives its results.

On the other hand, the economic example essentially serves as a proof of concept. It illustrates that our framework can accommodate not only pure philosophical but also economic problems. As a

[11] Note that in this case, the resulting representation is actually equivalent to a framework with a single probability distribution over states for every individual and simultaneously a modified probability distribution over individuals that includes belief systems.
[12] In Grant et al. (2010), this issue is referred to as a 'different attitude towards risk.'

bonus, the example allows us to demonstrate the effect of the degree of fairness on the level of the impartial observer.

4.1. **Afghan Goatherds.** First, recall the moral dilemma of the Afghan Goatherds described in the introduction. As mentioned there, our claim is not that the unit commander necessarily made the decision using anything related to our approach (even though it could very well be the case). However, our approach allows a normative analysis of this (and similar) situations. You could, for example, be developing a moral guideline for the military in the vein of the United States Army Field Manuals. Naturally, you would then be in the position of a neutral or impartial observer and evaluate the situation using each involved individual's point of view, including their perception of the situation's uncertainty.

Let us define the formal structure of the decision problem now. It is deliberately kept simplistic compared to the actual events. Hence, it might not be a perfect fit for every aspect of the original story but should still serve as a proper demonstration of an application of our theory. First, let $\mathcal{X} = \{0,1\}^2 \backslash \{(0,0)\}$ be the set of outcomes. Each element $x = (x_1, x_2)$ corresponds to the survival of the soldiers, x_1, and that of the afghan goatherds, x_2. Each entry then indicates either 'alive', 1, or 'dead', 0. Furthermore, let $\mathcal{S} = \{t, u\}$ be the states of the world, where t and u correspond to talking to the Taliban about the soldiers and keeping quiet about them, respectively. Finally, the two available moral choices are killing or sparing civilians, denoted by K and L. Thus, using the previous notation, they are given by:

$$K = \begin{cases} (1,0) & s = t, u \end{cases}$$

$$L = \begin{cases} (0,1) & s = t \\ (1,1) & s = u \end{cases}$$

Note that these elements only contain subjective risk (with respect to the states). Together with the remaining specifications, this is actually going to guarantee that individual belief systems purely drive the results in the example.

In addition, let $I = \{1, \ldots, 3\}$ be the set of individuals - corresponding to the team of soldiers.[13] Following the traditional setting, the identity lottery that is part of the choice problem is going to

[13] In the original story, the team consists of four soldiers including the commander of the unit. Each of the three regular soldiers exhibited different individual preferences, i.e., $K \succ_1 L$, $L \succ_2 K$, and $K \sim_3 L$, which is already enough to construct a simple (yet fascinating) example. Therefore, excluding the commander as an individual has no (significant) consequence for our analysis.

be fixed to $(1/3, 1/3, 1/3)$. Note that in this setting, i.e., with this set of individuals and this (fixed) identity lottery, the specifications of the next part conflict with our initial assumption of the absence of unanimity. Appendix A analyzes and corrects this problem by the introduction of a dummy variable without any changes to the preferences. Therefore, the rest of this analysis considers the absence of unanimity to be fulfilled.

4.1.1. *Specifying the Moral Dilemma.* To conduct a detailed analysis, we need to specify more about the preferences of the individuals and, in particular, the nature of their belief systems. First of all, to ensure that our results are purely driven by the soldiers' beliefs, we assume that their preferences are otherwise completely identical, that is, $u_i = u$ and $v_i = v$ for each $i \in \mathcal{I}$. Similarly, assume that the impartial observer treats the soldiers identical with respect to facing similar mixtures. Consequently, $\phi_i = \phi$, $i \in \mathcal{I}$, by Section 3.3.

Furthermore, let us specify the ranking of all possible survival outcomes. Naturally, you would assume that the soldiers rank the highest survival of all. Additionally, we consider that the soldiers, when confronted with the exclusive survival outcomes, prefer their own survival over that of the civilians. This essentially captures the idea of universal self-preservation instincts. Therefore, after using a positive affine transformation (to specify the lower and upper bound):

$$0 = u(0,1) < u(1,0) < u(1,1) = 1$$

Moreover, the individuals are assumed to be uncertainty-averse, implying a concave function v. In particular, it will be of the form z^q for $q \in (0,1)$. Assume further a preference for life chances (of the impartial observer). Thus, Section 3.2 yields that ϕ is also a concave function. As before, take z^r for $r \in (0,1)$.

In general, the belief systems m_i are probability distributions over $\Delta(\mathcal{S})$, which in this specific scenario is equivalent to probability distributions over $[0,1]$ by simply identifying $\mu \in \Delta(\mathcal{S})$ with $p \in [0,1]$ via $p = \mu(u)$ (alternatively $p = \mu(t)$). Now, the actual belief systems are going to be truncated normal distributions on $[0,1]$.[14] Let (μ_i, σ_i) denote a pair of parameters of the initial normal distributions. Then, assume that $\mu_i = \mu = 0.5$ for $i \in \mathcal{I}$ and furthermore $0 < \sigma_1 < \sigma_2 < \sigma_3 < +\infty$. Hence, the individual belief systems are mean-preserving spreads of each other, which allows us to showcase the effect of introducing uncertainty to the framework. Additionally, the centered mean

[14] Alternatively, take beta distributions $Beta(\alpha, \beta)$ with varying $\alpha = \beta$.

captures the idea of unbiased individuals. Finally, combining everything together yields the following utility of the impartial observer for the two moral choices:

$$\begin{aligned} V(z,K) &= \sum_{i=1}^{3} \frac{1}{3}\phi(v(u(1,0))) \\ &= (u(1,0)^q)^r \\ V(z,L) &= \sum_{i=1}^{3} \frac{1}{3}\phi\left(\int_0^1 v(pu(1,1)+(1-p)u(0,1))dm_i(p)\right) \\ &= \sum_{i=1}^{3} \frac{1}{3}\left(\int_0^1 p^q dm(\mu,\sigma_i)(p)\right)^r \end{aligned}$$

4.1.2. *Numerical Analysis.* In addition to calculating the results for specific values, the aim is to demonstrate the effect of uncertainty in our framework. Therefore, consider a modified version of the framework where each individual only uses a single (subjective) probability distribution over states, but on a societal level, there is an additional probability distribution over these subjective probability distributions.[15] In other words, instead of uncertainty, the model features subjective risk. Denote by $\tilde{V}(z,P)$ the evaluation of (z,P) corresponding to the modified model. Ideally, the comparison between our model and this modified one produces a difference and thus justifies to a certain degree the use of the concept of uncertainty in our model.

Finally, fix $q = 0.75$, $r = 0.25$, $\sigma_1 = 0.01$, $\sigma_2 = 0.1$, $\sigma_3 = 1$, and $u(1,0) = 0.48$. Now, the parameter choices here (and also the previous functional choices) should be understood as part of the example. Our (numerical) analysis uses reasonable but still debatable choices to enable us to showcase interesting phenomena for one of the potentially many formal interpretations of the story within our framework. Anyhow, using these parameters produces the following values when rounded to the fourth decimal:

	V_1	V_2	V_3	V	$\tilde{V}$
K	0.5767	0.5767	0.5767	0.8714	0.8714
L	0.5946	0.5923	0.5723	0.8751	0.8657

[15] Formally, this corresponds to the use of indicator functions as belief systems, i.e., $m_i = \mathbb{1}_{\mu_i}$ for $\mu_i \in \Delta(\mathcal{S})$ and all $i \in \mathcal{I}$, and an extended set of individuals that includes beliefs, i.e., $\mathcal{I} \times \Delta(\mathcal{S})$ where the density function is given by $f_z(i,\mu) = z_i \cdot m_i(\mu)$ for $(i,\mu) \in \mathcal{I} \times \Delta(\mathcal{S})$, instead of $\mathcal{I}$ and the corresponding $z \in \Delta(\mathcal{I})$.

As a consequence, the values produce the following rankings:

$$V_3(L) < V_{1/2/3}(K) < V_2(L) < V_1(L)$$

$$V(z,K) < V(z,L)$$

$$\tilde{V}(z,L) < \tilde{V}(z,K)$$

Therefore, the higher the (perceived) uncertainty on the level of the individuals, the lower the evaluation of L compared to that of K, which stays constant. In fact, the dynamic actually produces different individual rankings of the two alternatives. Additionally, the comparison of our model and the modified one actually produces a different ranking on the level of the impartial observer. Thus, the introduction of uncertainty influences (and to a specific extent drives) the final ranking. Further, the impartial observer agrees with the result of a simple majority vote in this particular case (which in general is not guaranteed).

4.2. **Exchange Economy.** Consider a simple exchange economy with two goods and two individuals, each receiving endowments. In this setting, compare two alternative re-distributions rules, namely the Walrasian auctioneer and the Egalitarian rule.[16] Uncertainty enters the model via a possible bias in the distribution of the endowments. This specific example serves as a proof-of-concept because it shows an interpretation of a traditional economic problem within our framework. It is based on an example by Eichberger and Pethig (1994).

Let $\mathcal{I} = \{1,2\}$ be the set of individuals and let x and y denote the two goods. In order to keep it simple, let us assume that the total endowment for each good is set to 3 and restricted to positive integers. Thus, the possible initial endowments are given by the two by two matrix

$$\begin{pmatrix} e_{1,1} & e_{1,2} \\ e_{2,1} & e_{2,2} \end{pmatrix} = \begin{pmatrix} ((1,1),(2,2)) & ((1,2),(2,1)) \\ ((2,1),(1,2)) & ((2,2),(1,1)) \end{pmatrix}$$

where $e_{i,j}$ in row i and column j corresponds to individual 1 receiving (i,j) and individual 2 receiving $(3-i,3-j)$ of the pair of goods $(e_{x,k}, e_{y,k})$, $k = 1,2$.

In the following, our analysis focuses on two possible re-distributions of these initial endowments, namely the Walrasian auctioneer and the Egalitarian rule. Let the utility function of an individual i for the (re-distributed) goods x_i and y_i be given by the Cobb-Douglas form $(x_i y_i)^{\alpha_i}$, where $\alpha_i \in$

[16] In general, there is also a combination of the two, namely the Walras rule from Equal Division. However, in this scenario, it coincides with the Egalitarian rule. See Nagahisa and Suh (1995) for a characterization of the Walras rules.

$\mathbb{R}_{>0}$. Then, the individuals evaluate the results of the re-distributions as follows (depending on endowments):

In case of the Egalitarian rule, any endowment vector $((e_{x,1}, e_{y,1}), (e_{x,2}, e_{y,2}))$ yields the consumption bundles $((3/2, 3/2), (3/2, 3/2))$. In consequence, the utility for an individual i is always given by $(9/4)^{\alpha_i}$, independent of initial endowments. Now, in case of the Walrasian auctioneer rule, any fixed endowment vector induces a corresponding unique Walrasian equilibrium. The resulting utility of individual i is then given by $((e_{x,i} + e_{y,i})^2/4)^{\alpha_i}$, where $e_{x,i}$ and $e_{y,i}$ are the initial endowments. Using our matrix notation, the following then characterizes the re-distribution rules with respect to the individual utilities for all possible initial endowments:

$$\tilde{E}\colon \begin{pmatrix} e_{1,1} & e_{1,2} \\ e_{2,1} & e_{2,2} \end{pmatrix} \mapsto \begin{pmatrix} ((9/4)^{\alpha_1}, (9/4)^{\alpha_2}) & ((9/4)^{\alpha_1}, (9/4)^{\alpha_2}) \\ ((9/4)^{\alpha_1}, (9/4)^{\alpha_2}) & ((9/4)^{\alpha_1}, (9/4)^{\alpha_2}) \end{pmatrix}$$

$$\tilde{W}\colon \begin{pmatrix} e_{1,1} & e_{1,2} \\ e_{2,1} & e_{2,2} \end{pmatrix} \mapsto \begin{pmatrix} (1, 4^{\alpha_2}) & ((9/4)^{\alpha_1}, (9/4)^{\alpha_2}) \\ ((9/4)^{\alpha_1}, (9/4)^{\alpha_2}) & (4^{\alpha_1}, 1) \end{pmatrix}$$

The notation with the tilde is deliberately distinguishing these descriptions from their counterparts that consider the uncertainty.

As mentioned earlier, the uncertainty is about the probability distribution over the initial endowments. Assume that there are two states, i.e. $\mathcal{S} = \{s_1, s_2\}$ where s_1 corresponds to a bias towards individual 1 and analogously s_2 towards 2. Thus, these are described via the following probability distributions (π_1 for s_1, π_2 for s_2):

$$\pi_1(e_{i,j}) = \begin{cases} \frac{1}{2} & i = j = 2 \\ \frac{1}{4} & i = 1, j = 2 \text{ or } i = 2, j = 1 \\ 0 & i = j = 1 \end{cases}$$

$$\pi_2(e_{i,j}) = \begin{cases} \frac{1}{2} & i = j = 1 \\ \frac{1}{4} & i = 1, j = 2 \text{ or } i = 2, j = 1 \\ 0 & i = j = 2 \end{cases}$$

In a sense, our example exhibits uncertainty about (the state of) the economy instead of a fixed (state of the) economy with inherent uncertainty. Consequently, society chooses the re-distribution rule before knowing the initial distributions.

Finally, to formally state the two rules taking uncertainty into consideration, i.e., to formulate the act lotteries corresponding to them, combine the functions above and probability distributions as follows:

$$E\colon s_k \mapsto \left(\tilde{E}(e_{i,j}) \text{ with probability } \pi_k(e_{i,j})\right)$$
$$W\colon s_k \mapsto \left(\tilde{W}(e_{i,j}) \text{ with probability } \pi_k(e_{i,j})\right)$$

Furthermore, using the inherent symmetries and other similarities simplifies it to:

$$E\colon s \mapsto \left((9/4)^{\alpha_1}, (9/4)^{\alpha_2}\right)$$
$$W\colon s \mapsto \begin{cases} \left(4^{\alpha_1 \mathbb{1}_{\{s_1\}}(s)}, 4^{\alpha_2 \mathbb{1}_{\{s_2\}}(s)}\right) & \text{with probability } 1/2 \\ \left((9/4)^{\alpha_1}, (9/4)^{\alpha_2}\right)) & \text{with probability } 1/2 \end{cases}$$

Following the traditional setting, take the 'fair' uniform identity lottery again, that is $z = (1/2, 1/2)$. As in the previous example, assume that the impartial observer treats individuals identical with respect to similar mixtures; thus, $\phi_i = \phi$, $i \in \mathcal{I}$. In addition, set $\alpha_i = 1/2$ for both i (assuming similar risk-aversion) and fix v_i to be the identity function for both i (assuming uncertainty-neutrality).[17] Also, take $\phi = z^r$ again but with $r \in (0, +\infty)$ this time.

As individual belief systems m_i take truncated normal distributions on $[0, 1]$ again[18], where $p \in [0, 1]$ corresponds to the probability of state s_i realizing. Further, let (μ_i, σ_i) denote a pair of parameters of the initial normal distributions. Then assume that $\mu_1 = 0.75$, $\mu_2 = 0.25$, and $\sigma_i = \sigma = 0.25$, i.e. each of the individuals suspects a bias towards individual 1 and the same level of volatility.

Note that due to the exclusive nature of the game, essentially a zero-sum game, and the additional assumptions, the absence of unanimity requires no modifications. Finally, combine everything for the

[17] It seems completely counter-intuitive to assume uncertainty-neutrality in our framework as the introduction of uncertainty is our main contribution. However, in this example and specifically this (numerical) analysis, our focus is on the effect of different transformations ϕ.

[18] Alternatively, take beta distributions $Beta(\alpha, \beta)$ with varying $\alpha = \beta$.

following evaluations:

$$\begin{aligned} V(z,E) &= \sum_{i=1}^{2} \frac{1}{2}\phi_i \left(v_i \left(\left(\frac{9}{4} \right)^{\alpha_i} \right) \right) \\ &= \left(\frac{3}{2} \right)^{r} \\ V(z,W) &= \sum_{i=1}^{2} \frac{1}{2}\phi_i \left(\int_0^1 v_i \left(\frac{1}{2} \left(\frac{9}{4} \right)^{\alpha_i} + \frac{1}{2} \left(4^{\alpha_i} p + 1\,(1-p) \right) \right) dm_i(p) \right) \\ &= \sum_{i=1}^{2} \frac{1}{2} \left(\frac{5}{4} + \frac{1}{2} \int_0^1 p dm(\mu_i, \sigma)(p) \right)^{r} \end{aligned}$$

Consider different values for $r \in (0, +\infty)$ now, which captures different degrees of fairness on the level of the impartial observer. It results in the following values when rounded to the fourth decimal:

	V_1	V_2	$V_{\|r=1.5}$	$V_{\|r=1}$	$V_{\|r=0.5}$
W	1.5897	1.4103	1.8396	1.5000	1.2242
E	1.5000	1.5000	1.8371	1.5000	1.2247

Evidently, the ranking of the impartial observer depends on the exact value of r, and the rankings of the individuals are diametrically opposed:

$$V_2(W) < V_2(E) = V_1(E) < V_1(W)$$

$$V(z,E)_{|r=1.5} < V(z,W)_{|r=1.5}$$

$$V(z,E)_{|r=1} = V(z,W)_{|r=1}$$

$$V(z,E)_{|r=0.5} > V(z,W)_{|r=0.5}$$

In other words, a preference for life chances of the impartial observer actually leads to a selection of the Egalitarian rule over that of the Walrasian auctioneer in a setting with a clear bias towards one specific individual. It essentially provides another example for the discussion on the issue of fairness.

5. Conclusion

The focus of this paper is the normative decision-making of an individual when considering all other individuals in society and under the influence of uncertainty. Based on the works of Grant et al. (2010) and Seo (2009), we provide an axiomatic foundation for an extension of Harsanyi's Impartial Observer Theorem that includes Knightian uncertainty while also accomodating specific common criticism of the traditional approach. The main result in the paper shows that the impartial observer's preferences admit a representation in the form of a weighted average of the individual (transformed) second-order subjective expected utilities. This representation allows for a tractable analysis of the normative choice problems under consideration. Furthermore, the framework re-establishes links between additional properties of the preferences of the impartial observer on the one hand (e.g., the issues of fairness and mixtures) and the specific form of the individual transformations on the other hand.

The main appeal of our model is the extension to normative decision-making in situations where a group of individuals faces subjective instead of objective risk. The story of the Afghan Goatherds is an example of such moral value judgments. It is purely driven by individual belief systems and thus justifies the introduction of uncertainty to the framework. Moreover, the example shows that our model extends beyond the limitations of the absence of unanimity via the use of a dummy individual. Finally, the economic example applies our theory to a scenario that demonstrates the effect of a preference for life chances - compared to accidents of birth - of the impartial observer. It also serves as a proof-of-concept for applying our theory to economic problems in general.

As a final note, Fleurbaey (2018) writes: "If ambiguity triggers emotions in the population, the social evaluator may rationally want to take account of them, but this is very different from adopting an ambiguity averse decision criteria." In our setup, ambiguity is introduced on an individual level, and we believe that individuals' attitudes toward ambiguity can be dealt with by the impartial observer (see 3.3). Moreover, as our exchange economy example demonstrates, veil-of-ignorance arguments are not necessarily prone to unfairness, and such criticism can be handled with the proper choice of the ϕ function.

References

Alon, S. and G. Gayer (2016): "Utilitarian Preferences With Multiple Priors," *Econometrica*, 84, 1181–1201.

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

Billot, A. and V. Vergopoulos (2016): "Aggregation of Paretian preferences for independent individual uncertainties," *Social Choice and Welafre*, 47, 973–984.

Crès, H., I. Gilboa, and N. Vieille (2011): "Aggregation of multiple prior opinions," *Journal of Economic Theory*, 146, 2563–2582.

Danan, E., T. Gajdos, B. Hill, and J.-M. Tallon (2016): "Robust Social Decisions," *American Economic Review*, 106, 2407–25.

Diamond, P. (1967): "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparison of Utility: Comment," *Journal of Political Economy*, 75, 765–766.

Eichberger, J. and R. Pethig (1994): "Constitutional choice of rules," *European Journal of Political Economy*, 10, 311 – 337.

Fleurbaey, M. (2018): "Welfare economics, risk and uncertainty," *Canadian Journal of Economics*, 51, 5–40.

Gajdos, T. and F. Kandil (2008): "The ignorant observer," *Social Choice and Welafre*, 31, 193–232.

Grant, S., A. Kajii, B. Polak, and Z. Safra (2010): "Generalized Utilitarianism and Harsanyi's Impartial Observer Theorem," *Econometrica*, 78, 1939–1971.

Harsanyi, J. (1953): "Cardinal Utility in Welfare Economics and in the Theory of Risk-taking," *Journal of Political Economy*, 61, 434–435.

——— (1955): "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility," *Journal of Political Economy*, 63, 309–321.

——— (1977): *Rational Behaviour and Bargaining Equilibrium in Games and Social Situations*, Campbirdge University Press.

Klibanoff, P., M. Marinacci, and S. Mukerji (2005): "A Smooth Model of Decision Making under Ambiguity," *Econometrica*, 73, 1849–1892.

Mongin, P. (1995): "Consistent Bayesian Aggregation," *Journal of Economic Theory*, 66, 313–351.

——— (2001): "The impartial observer theorem of social ethics," *Economics and Philosophy*, 17, 147–179.

Nagahisa, R. I. and S. C. Suh (1995): "A characterization of the Walras rule," *Social Choice and Welfare*, 12, 335–352.

Nascimento, L. (2012): "The ex ante aggregation of opinions under uncertainty," *Theoretical Economics*, 7, 535–570.

Qu, X. (2017): "Separate aggregation of beliefs and values under ambiguity," *Economic Theory*, 63, 503–519.

Raiffa, H. (1970): *Decision Analysis - Introductory lectures on choices under uncertainty*, Reading, MA: Addison Wesley.

Rawls, J. (1971): *A Theory of Justice*, Belknap.

Sandel, M. (2010): *Justice: What's the right thing to do?*, Farrar, Straus and Giroux.

Seo, K. (2009): "Ambiguity and Second-Order Belief," *Econometrica*, 77, 1575–1605.

Smith, A. (1759): *The Theory of Moral Sentiments*, London: A. Millar.

Stanca, L. (2021): "Smooth aggregation of Bayesian experts," *Journal of Economic Theory*, 196.

Vickrey, W. (1945): "Measuring marginal utility by reaction to risk," *Econometrica*, 13, 319–333.

Appendix A. Construction of a Dummy

As mentioned in chapter 4, in order to satisfy the assumption of the absence of unanimity, without distorting any preferences, the introduction of a dummy individual d is necessary. The following explains this necessity:

Assuming sufficiently heterogeneous beliefs, it is certainly possible to imagine a ranking of the form $K \succ_1 L$, $L \succ_2 K$ and $K \sim_3 L$ - essentially mirroring reality. However, the assumption of the absence of unanimity also applies to all degenerate outcome lotteries, like the one always yielding the outcome $(1, 1)$ irrespective of the state of the world and the one yielding $(0, 1)$. Indeed, it is counter-intuitive to assume that, in reality, one of the soldiers would prefer the second over the first one, i.e., preferring being dead over being alive with everything else fixed. Now, the dummy individual takes care of this problem by choosing $(0, 1)$ over $(1, 1)$ and maintaining the absence of unanimity. At the same time, the probability of imagining yourself as the dummy individual is set to zero (for both options) to prevent any distortion on the level of preferences of the impartial observer.

A dummy individual seems artificial, especially necessary because of an assumption that we impose on the model. Yet, it is not an actual restriction or invalidates the assumption. It is merely a technical solution to a technical problem. The two presented (degenerate) lotteries that conflict with the absence of unanimity otherwise are (or were) not part of the set of feasible options in reality anyway. A dummy individual in this example is necessary due to the homogeneity of the individuals (and their preferences), resulting from a simplistic structure. Thus, a dummy individual allows us to apply our theory to examples where the size of the choice set collides with the absence of unanimity otherwise. Essentially, this weakens the absence of unanimity while remaining in the framework of our theory.

Formally: Consider $\mathcal{I}' = \mathcal{I} \cup \{d\}$ with the (fixed) identity lottery $(1/3, 1/3, 1/3, 0)$ and set $u_d(x_1, x_2) = u(x_1, x_2)$ and $v_d(y) = 1 - v(y)$ with $m_d = m_1$ (or $m_d = m_{2/3}$). It results in the following (ultimately irrelevant) utilities for the two moral choices: $V_d(K) = 0.4233$ and $V_d(L) = 0.4054$; Consequently: $V_d(L) < V_d(K)$.