We examine the subjective value of gaining knowledge in a version of Savage’s model for decisions under uncertainty in which the received outcome provides information about which event has obtained. Decision makers commonly value such knowledge either because they want to use it in future decisions or because they are intrinsically interested in it. We find that in our model, the sure-thing principle and several other axioms of Savage are inconsistent with this value for knowledge about events. We provide a representation theorem for a subjective value of knowledge consisting of the sum of expected utility and a function of the information partition generated by the outcomes of an act. Bayesian updating of likelihood judgments and stationarity of preferences imply that decision makers rank acts by the sum of expected utility and the Shannon entropy. Our results also provide a novel critique of the necessity of Savage’s axioms for rational decisions under uncertainty.

Keywords: Decision Theory, Uncertainty, Learning, Expected Utility, Preferences, Entropy, Knowledge, Information

JEL Classification: D81, C44
1 INTRODUCTION

Decisions often involve choices between alternatives from which knowledge can be gained. Consider the number of test subjects recruited by an experimenter, or the too few hours of sleep left after finishing an addictive novel — both of these reveal a preference for acquiring knowledge. Sometimes, this knowledge is acquired as a means to achieve other goals; other times, this knowledge is completely inconsequential. The experimenter hopes to gain useful knowledge but Agatha Christie’s *Murder on the Orient Express* is read for pleasure only. In this paper, we construct a theory of decision under uncertainty that accounts for preferences for knowledge in general.

We embed our analysis into Savage (1954)’s framework and axiomatization of rational behavior under uncertainty. In this model, decision makers have preferences over acts. An act yields an outcome for every possible state of the world. We assume that after an act is resolved, the decision maker infers from the obtained outcome and the chosen act what event obtained. The decision maker can therefore only distinguish two events via an act that yields distinct outcomes on the two events.

We argue that if a decision maker is interested in gaining knowledge about which event obtains, then several of Savage’s axioms lose their normative appeal. This is easiest to see for the monotonicity axiom that requires that the preferences over achieving outcomes on an event are the same irrespective of what other outcomes are achieved on the other events. If the decision maker cares about knowledge, replacing an outcome by a worse outcome on some event may allow the decision maker to identify if this event obtains and may thus be desirable. For example, in order to find out whether an event obtains, you may reject a sure payoff of 100 Euros in favor of a gamble that pays you 100 Euros only if an event obtains and nothing otherwise.

We provide an axiomatization of a decision model that accounts for a preference for acquiring knowledge about events. For this, we weaken the axioms of Savage (1954). Savage’s sure-thing principle imposes separability of preferences on conditional acts given an event from the preferences on conditional acts on the complement. We only assume this separability of preferences whenever the event can be distinguished from its complement. Similar adjustments are made to the monotonicity axiom and the likelihood outcome independence axiom. We add one axiom that guarantees that the preference for knowledge is separable from the preference for achieving good
outcomes. Our representation theorem characterizes what we call a subjective knowledge utility in which the utility of an act is additively separable into the expected utility of the outcomes and the value of the information partition revealed by the act. The value of the information partition is in turn additively separable in the revealed events. We then analyze the value of knowledge in more detail both conceptually and axiomatically.

Conceptually, our axiomatic analysis gives rise to a utility representation of the value of knowledge from which we derive further concepts. First, we define a classification of decision makers by their knowledge preferences into knowledge-seeking, knowledge-neutral, and knowledge-averse and show that these can be linked to subadditivity and superadditivity properties of our utility representation of knowledge preferences. Second, we define a measure of the value of knowledge about events and a knowledge equivalent which roughly play the role of risk attitude measures and certainty equivalents in decisions under uncertainty. Third, we derive a curiosity relation with which decision makers can be ranked by their value of knowledge.

Axiomatically, we sharpen our decision model further, by adding axioms that guarantee that the decision maker does not care about obtaining specific pieces of knowledge but rather cares about the (quantified) amount of information gained about the state of the world.

Probabilistic knowledge utility consists of the sum of expected utility and a very general information measure. It is characterized by assuming that the decision maker’s value of knowledge only depends on the likelihoods of the involved events. This representation turns out not to be stationary in the sense that the value of information may change as additional knowledge is gained.

If the information measure is the Shannon (1948) entropy, we obtain entropic knowledge utility. It is characterized by a stationarity condition which ensures that after Bayesian updating of the probabilities on an event, the tradeoff remains the same between achieving good outcomes and being able to distinguishing two events with a fixed relative probability. Essentially, this allows the decision maker to use the same utility functions for evaluating conditional acts after Bayesian updating of the probabilities.

In the context of probabilistic knowledge utility we also define what we call the elasticity of curiosity that describes the change in intensity of knowledge preferences as the decision maker becomes more and more informed. From this definition, we derive constant elasticity of curiosity utility which is a
convenient two parameter functional form that incorporates two important aspects of knowledge preferences: the intensity of knowledge preferences and the elasticity of curiosity.

Stationary information cost utility accounts for the fact that decision makers may sometimes only be interested in a specific statistical variable but consider any information received as costly to process. For example, a decision maker who is interested in the mean of a random variable would prefer to know the sample mean of a data set instead of being informed about every single data point. Stationary information cost utility attaches a prior-independent value to each posterior probability measure over a random variable. The cost of processing the information is given by the Shannon entropy. We characterize this functional form by weakening the stationarity axioms used to characterize entropic knowledge utility. An example of a stationary information cost utility is the mutual information between the known events and the variable the decision maker is interested in.

In the light of our results, we briefly revisit Savage’s endeavour to provide a foundation of statistics. Savage (1954) considered statistics as a practice of solving partition problems in which a decision maker tries to distinguish if an event obtains or not. In our framework, acts not only provide more or less good outcomes but also solve partition problems. This allows us to consider the choice of a hypothesis test or the choice of the number of subjects invited to an experiment as choices over acts, enabling us to study them within our theory. As an application we therefore briefly sketch how estimation problems, hypothesis tests, or optimal data collection fall into the domain of the decision problems to which our model applies.

The paper proceeds as follows. In section 2, we briefly review the related literature. The notation is introduced in section 3. In section 4, we introduce our model of knowledge in a Savage framework, propose our adjustments to Savage’s axioms, and characterize subjective knowledge utility based on these axioms. Several concepts to analyze the value of knowledge of decision makers are introduced in section 5. We then axiomatically characterize various forms of knowledge preferences, specifically, probabilistic knowledge utility in subsection 6.1, entropic knowledge utility in subsection 6.2, constant elasticity of curiosity utility in subsection 6.3, and stationary information cost utility in subsection 6.4. Section 7 translates aspects of statistical estimation into our framework. Finally, section 8 discusses alternative ways of modeling knowledge
and the necessity of Savage’s axioms for rationality before section 9 concludes.

2 Literature

Our paper is closely related to the literature on the intrinsic value of information, which has been studied theoretically (Grant, Kajii, & Polak, 1998; Golman & Loewenstein, 2018) and empirically (Bennett, Bode, Brydevall, Warren, & Murawski, 2016; Falk & Zimmermann, 2016; Kops & Pasichnichenko, 2020; Masatlioglu, Orhun, & Raymond, 2017). Kadane, Schervish, and Seidenfeld (2008) discuss whether it can be rational for a decision maker to pay for not receiving information and how a theory of decision can incorporate this. Our adjustment to Savage’s model allows for a negative value of information and therefore also for such preferences. Wakker (1988a) shows that violations of the independence axiom imply that a decision maker displays information-averse behavior for some choices. Safra and Sulganik (1995) shows that for every information structure there exists a nonexpected utility maximizer that prefers less information to more. In a working paper, Alaoui (2012) studies under which circumstances behavioral decision makers attach value to useless information. Luce, Ng, Marley, and Aczél (2008a, 2008b) study the functional form of our entropic knowledge utility representation in the context of utility for gambling. In a recent working paper, Liang (2019) characterizes intrinsic information preferences in an Anscombe and Aumann (1963) model. Information in Liang (2019) is modeled as objective probabilities over the state space as revealed by an outcome. Similar to our model, the decision maker infers information about the state space from the outcomes. Unlike our model, inference is about the objective probabilities of the second stage of the Anscombe-Aumann model. Thus, the remaining uncertainty after an act is resolved is an objective probability distribution over states of the world, which enters the decision maker’s utility function. We do not require the availability of objective probabilities for our axiomatization. Our decision maker values knowledge about which event obtains and assigns these events a subjective probability. Knowledge about which event obtains of course provides information about what the state of the world is. However, in our model this information may be entirely subjective and different decision makers may disagree on how informative an event is about a state. Another important aspect that distinguishes our model is that since our decision maker values knowledge of events, the decision maker may
have a stronger preference for knowing the combination of some states than the combination of other states.

Cooke (2017) models a tradeoff between current consumption and (taste) information using a multi-stage model in which choices over intial consumption and choices over menus of later consumption are observed. Caplin and Leahy (2001) model anticipatory feelings over the resolution of lotteries, such as suspense. In contrast to their model, in which probabilities are fully objective and are separated into two stages, the uncertainty in our paper is fully subjective on arbitrary event spaces.


The conception of knowledge about the state of the world that we employ is the usual one employed in decision and game theory, using partitions of the state space. Aumann (1999) links this to a syntactic representation of knowledge. Gilboa and Lehrer (1991) characterize the value of information partitions that is expected utility rationalizable using methods from cooperative games. Their paper provides an answer to the question under which conditions the value of knowledge we characterize may arise from expected utility maximization of a subsequent decision problem. Azrieli and Lehrer (2008) extend their analysis to stochastic information structures. Ghirardato (2002) discusses how a dynamic consistency axiom that resembles the sure-thing principle expresses a nonnegative instrumental value of information. In a recent working paper, Galanis (2019) further extends the analysis of this relation between dynamic consistency and information value to ambiguity averse decision makers.

If the value of knowledge in our paper is seen as purely instrumental for
solving future decision problems, then there is a natural relation to the value of flexibility (Kreps, 1979). Flexibility allows decision makers to react to future information while knowledge allows decision makers to use flexibility. Nehring (1999) axiomatizes preference for flexibility in a Savage framework. Studying the duality between these axiomatizations is an interesting avenue for further research.

3 Notation

Let $S$ be a set of states of nature. $\mathcal{E}$ denotes a $\sigma$-algebra of events on $S$. For any event $E$, let $\overline{E} = S - E$ be the complement of $E$ in $S$. $\mathcal{X}$ is a set of outcomes. An act is a function $a : S \rightarrow \mathcal{X}$ that is measurable, i.e., for all subsets of outcomes $X \subseteq \mathcal{X}$, the preimage is an event, $a^{-1}(X) \in \mathcal{E}$. In this paper we only consider simple acts, i.e., acts that are finite-valued $|a(S)| < \infty$.

The set of (simple) acts is denoted by $A$. If $E \in \mathcal{E}$ is an event, and $f$ an act then $f_E : E \rightarrow \mathcal{X}$ denotes the restriction of $f$ to $E$ which we will call a conditional act. The set of conditional acts on an event $E$ is denoted $A_E$. The set of outcomes resulting from a conditional act $f_E$ and an event $F \subseteq E$ is defined as the image $f_E(F) = \{ \alpha \in \mathcal{X} : f_E(s) = \alpha \text{ for some } s \in F \}$. For convenience, we introduce several ways of denoting acts. If $f$ and $g$ are acts and $E$ is an event, then $f_Eg$ denotes the act that agrees with $f$ on the event $E$ and with $g$ on the event $\overline{E}$. Constant acts are simply denoted by their outcome, $\alpha \in \mathcal{X}$.

A decision maker ranks acts via a preference relation $\succeq$. $f \succeq g$ means the act $f$ is at least as good as the act $g$, $f \sim g$ is the symmetric part of the relation and means that the decision maker is indifferent between the two acts. $\succ$ denotes the asymmetric part of the relation and indicates strict preference. A function $U : A \rightarrow \mathbb{R}$ represents the relation $\succeq$ if $f \succeq g \iff U(f) \geq U(g)$.

An event $E$ is nonnull if for some outcomes $\gamma \succ \beta$. We denote the set of null events by $\mathcal{N} \subseteq \mathcal{E}$. An atom is a nonnull event $E$ that cannot be partitioned into two nonnull events. We assume throughout the paper that there are no atoms. This guarantees that the event space is sufficiently rich.

We make the following additional richness assumption that greatly simplifies our analysis. For every outcome $\alpha$, there exists a countable number of indifferent outcomes $\alpha', \alpha'', \ldots$. We can interpret the outcome $\alpha'$ as the same as the outcome $\alpha$ but with a label attached to it that allows the decision maker to distinguish between $\alpha$ and $\alpha'$.
For some results, we also assume that the set of outcomes is rich enough such that conditional acts are outcome-solvable, i.e., for all events $E$ and all conditional acts $f_E$ and some $\gamma \not\in f_E(E)$ there exists an outcome $\epsilon \neq \gamma$, such that $\epsilon_E \gamma \sim f_E \gamma$.

4 Axiomatization of a Preference for Knowledge

Savage (1954) did not specify what information the decision maker acquires about the state space as an act is resolved. To start, we clarify that Savage’s axioms maintain their normative appeal in case the exact state of the world is always revealed when the act is resolved. However, there are many decisions in which different information is acquired depending on which act is chosen. To account for such decisions, we must clarify what information is acquired as an act is resolved.

We assume that the decision maker finds out which event obtains by combining the memory of which act she played with the information which outcome she experienced. Since this is a justified true belief, we say that the decision knows that this event obtains. Thus, if the decision maker receives outcome $a$ from act $f$, then the decision maker knows that the event $f^{-1}(a)$ obtains. The knowledge that will be acquired from this act can therefore be described by the information partition $\iota(f) = \{f^{-1}(a) : a \in X\}$. An example is provided in Table 1. $a$ is a constant act that does not reveal any knowledge about events. $b$ yields distinct outcomes on events $E$ and $\overline{E}$ and the decision maker can conclude from the outcome which of these two events obtains. We justify this choice of modeling knowledge and discuss alternative possibilities in section 8.

<table>
<thead>
<tr>
<th></th>
<th>$E$</th>
<th>$\overline{E}$</th>
<th>$\iota$</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>$a$</td>
<td>$a$</td>
<td>${S}$</td>
</tr>
<tr>
<td>b</td>
<td>$a$</td>
<td>$\beta$</td>
<td>${E, \overline{E}}$</td>
</tr>
<tr>
<td>c</td>
<td>$\beta$</td>
<td>$\alpha$</td>
<td>${E, \overline{E}}$</td>
</tr>
<tr>
<td>d</td>
<td>$\beta$</td>
<td>$\beta$</td>
<td>${S}$</td>
</tr>
</tbody>
</table>

Table 1: Example Acts and their Information Partition

We next examine whether Savage’s axioms maintain their normative appeal under this interpretation of the knowledge gained from an act and propose

---

3The state space of a roulette wheel fulfills this condition. The joint state space of a roulette wheel and a blackjack table violates this condition in case one can only stand at either table and not observe the other.
alternative axioms where necessary. The normative position from which we
discuss the axioms can be summarized as follows:

“\(\text{A decision maker may gain joy from acquiring knowledge, irrespective of}
whether this knowledge will be useful in future decisions. The decision maker
may be willing to forego better outcomes in favor of gaining this knowledge. A}
theory of rational choice under uncertainty must allow for this behavior.\)”

We think this statement is uncontroversial to most decision makers. If
you ever bought and read a book purely for entertainment and were excited
about the events revealed in the book, most likely you will need to accept this
statement. We say that a decision maker who exhibits the described behavior
has a preference for knowledge. Our theory of a preference for knowledge
about events naturally encompasses a theory of preference for information
about events or states of the world. This is because knowledge of an event may
provide information about other, correlated events. This will be made more
precise in subsection 6.4.

We now show that Savage’s axioms of expected utility are incompatible with
a preference for knowledge, and provide alternative axioms that characterize a
novel representation of preferences.

P 1 (Weak Order). \(\succeq\) is complete and transitive.

We will maintain the Weak Order assumption. The Blackwell order is an
example of an informativeness ranking that does not fulfill completeness as it
ranks experiments by whether they reveal unambiguously more information.
Relaxing the assumption of completeness may therefore be an interesting
avenue for future research.

P 2 (Sure-Thing Principle). For all \(f, g \in \mathcal{A}\), and all \(\alpha, \beta \in \mathcal{X}\),

\[
\alpha_{Ef} \succeq \alpha_{Eg} \iff \beta_{Ef} \succeq \beta_{Eg}
\]  

(1)

The sure-thing principle guarantees that we can ignore identical conditional
subacts when comparing two acts. A simple example shows that the sure-thing
principle is in direct conflict with a preference for knowledge. Consider the
acts in Table 1. The sure-thing principle requires that \(a \succeq b\) if and only if \(c \succeq d\).
However, \(b\) and \(c\) provide knowledge about \(E\) while \(a\) and \(d\) do not. A decision
maker who would like to know whether event \(E\) obtains may therefore prefer \(b\)
to \(a\) and \(c\) to \(d\) as long as the outcomes \(\alpha\) and \(\beta\) are sufficiently close in value.
The issue is that \(a\) and \(b\) as well as \(c\) and \(d\) not only differ by a change in the
outcomes but also by a change in the information partition. The sure-thing principle demands that the change in the information partition is irrelevant for preference. Notice however that in case the outcomes yielded by $f_E$ and $g_E$ on $E$ all differ from $\alpha$ and $\beta$, then $P_2$ is not inconsistent with a preference for knowledge because $E$ is always part of the information partition. In this case, the sure-thing principle maintains its normative appeal.

We now formalize this idea using disjoint conditional acts.

**Definition 1 (Disjoint Conditional Acts).** A conditional act $f_E$ is disjoint from a conditional act $g_F$ if the images $f_E(E)$ and $g_F(F)$ are disjoint: $f_E(E) \cap g_F(F) = \emptyset$.

Two conditional acts are therefore disjoint if they have no outcomes in common. This definition helps us to specify which events the decision maker can identify after an act is resolved. If $f_E$ and $f_E$ are disjoint, then after the act $f$, the decision maker will know with certainty whether event $E$ has happened or not. We adjust the sure-thing principle in the following manner:

**Axiom 2 (Information-Neutral Sure-Thing Principle).** Let $E \in \mathcal{E}$ be an event. Suppose $f_E$ and $g_E$ are disjoint from $h_E$ and $k_E$. Then,

$$f_E h \succ g_E h \iff f_E k \succ g_E k \quad (2)$$

The information-neutral sure-thing principle differs from the sure-thing principle in two respects. First, full separability of the preferences across $E$ and $\overline{E}$ is limited to acts in which the decision maker learns with certainty whether event $E$ obtains. This is achieved by requiring disjointness of the conditional acts on $E$ and the conditional acts on its complements. Second, instead of only changing a single outcome, we allow for changes of conditional acts, i.e., the change in Equation (2) from $h_E$ to $k_E$ may not only change a particular outcome but may change several outcomes. This means that also the knowledge about events contained in $h_E$ and $k_E$ may differ. The independence condition therefore not only imposes separability of the preferences on outcomes but also on knowledge about events. For example, $h_E$ may achieve a single good outcome and $k_E$ achieves several poor outcomes but informs the decision maker about several subevents of $\overline{E}$.

Savage’s monotonicity axiom conflicts in a similar manner as the sure-thing principle with a preference for knowledge.

---

For easier comparison, we number our axioms by their counterparts $P_1$-$P_6$ in Savage (1954).
P 3 (Monotonicity). For all $f \in \mathcal{A}$, $E \in \mathcal{E} - \mathcal{N}$, and all $\beta, \gamma \in \mathcal{X}$,

$$\gamma \succsim \beta \iff \gamma Ef \succsim \beta Ef$$

(3)

Monotonicity states that outcomes are ranked the same way on every event. It implies in our previous example in Table 1 that $a$ is preferred to $d$ if and only if $b$ is preferred to $d$. However, note that $b$ reveals whether event $E$ obtains but $a$ and $d$ do not. The decision maker may therefore judge $d$ to be better than $a$ because the outcome $\beta$ is preferable to the outcome $\alpha$ and at the same time judge $b$ to be better than $d$ because $b$ reveals event $E$. Note again that there is no issue with imposing the monotonicity axiom with respect to event $E$ in case the outcomes on $E$ are disjoint from the outcomes of $\overline{E}$.

We therefore make the following adjustment to the monotonicity axiom:

Axiom 3 (Information-Neutral Monotonicity). Suppose $h_E$ is disjoint from $\gamma$ and $\beta$. Then,

$$\gamma Eh_E \succsim \beta Eh_E \iff \gamma \succsim \beta$$

(4)

Information-neutral monotonicity implies that outcomes are identically ranked on all events as long as a change in outcomes does not change the information partition. By requiring disjointness, we guarantee that when combining $\beta_E$ and $\gamma_E$ with $h_E$, the information about the event $E$ is identical across the two acts. For example, $\beta_E \gamma_E \succ \gamma$ and $\gamma \succ \beta$ is a legitimate preference in our model. It means that the outcome $\gamma$ is preferable to $\beta$ but obtaining information about the event $E$ may be more valuable than receiving the better outcome $\gamma$ instead of $\beta$ on the event $E$.

P 4 (Likelihood Outcome Independence). For all $E, F \in \mathcal{E}$ and all $\beta, \beta', \gamma, \gamma' \in \mathcal{X}$, such that $\gamma \succ \beta$ and $\gamma' \succ \beta'$,

$$\gamma_E \beta \succsim \gamma_F \beta \iff \gamma'_E \beta' \succsim \gamma'_F \beta'$$

(5)

The likelihood outcome independence axiom ensures that the decision maker has preferences consistent with the existence of a likelihood relation over events. For an expected utility maximizing decision maker who prefers $\gamma_E \beta$ to $\gamma_F \beta$, we can conclude that the decision maker judges $E$ to be more likely than $F$ if $\gamma \succ \beta$. No matter whether $\gamma'$ is a lot or just slightly better than $\beta'$, the decision maker should prefer to realize $\gamma'$ on the more likely event. Thus, the
decision maker should also prefer $\gamma_E' \beta' \to \gamma_F' \beta'$.

For a decision maker who cares about knowledge, $\gamma_E \beta > \gamma_F \beta$ need not mean that $E$ is strictly more likely than $F$. Instead, the decision maker may simply prefer to know whether $E$ obtains to knowing whether $F$ obtains. Indeed, if $F = S$, then $\gamma_F \beta = \gamma$ is completely uninformative and $\gamma_E \beta > \gamma_F \beta$ simply means that the decision maker is willing to accept the loss of realizing $\beta$ instead of $\gamma$ on event $\overline{E}$ in order to find out whether event $E$ or event $\overline{E}$ obtains. Likelihood outcome independence then requires the decision maker to not only accept this loss but also any arbitrarily large loss of realizing $\beta'$ instead of $\gamma'$ on $\overline{E}$. In this manner, likelihood outcome independence disallows tradeoffs between the value of knowledge and outcomes.

Yet again there are also cases in which likelihood outcome independence is consistent with a preference for knowledge. From $\gamma_E \beta \succeq \beta_E \gamma = \gamma_F \beta$, we can indeed conclude that $E$ is at least as likely as $\overline{E}$; in both acts the decision maker learns whether the event $E$ obtains. The former act must then be at least as good as the latter act in virtue of event $E$ being at least as likely as its complement. We can generalize this idea to arbitrary disjoint events $E$ and $F$ by concluding that $E$ is at least as likely as $F$ if the decision maker weakly prefers to receive the good outcome $\gamma$ on $E$ and the bad outcome $\beta$ on $F$ as long as both acts yield the same information. We therefore change the likelihood outcome independence axiom to be consistent with this way of identifying likelihoods.

**Axiom 4** (Information-Neutral Likelihood Outcome Independence). For all disjoint events $E, F \in \mathcal{E}$ and all distinct outcomes $\alpha, \beta, \beta', \gamma, \gamma' \in \mathcal{X}$, if $\gamma > \beta$, $\gamma' > \beta'$, then

$$
\gamma_E \beta F \alpha \succeq \beta_E \gamma F \alpha \iff \gamma'_E \beta' F \alpha \succeq \beta'_E \gamma' F \alpha.
$$

In fact, it is straightforward to show that the above axiom implies likelihood outcome independence given the sure-thing principle. Under the weaker information-neutral sure-thing principle however, the two axioms are distinct.
For a preference \( \succeq \) fulfilling information-neutral likelihood outcome independence, there exists a well-defined likelihood relation \( \succeq^* \) over events.\(^3\)

**Definition 2 (Likelihood Relation).** Let \( E, F \in \mathcal{E} \) be events and \( \alpha, \beta, \gamma \in \mathcal{X} \) be distinct outcomes such that \( \gamma \succ \beta \). \( E \) is revealed by \( \succeq \) to be more likely than \( F \), \( E \succeq^* F \) if

\[
\gamma E - F \beta F - E \succ \beta E - F \gamma F - E \alpha.
\] (7)

**P 5 (Nontriviality).** There are \( \beta, \gamma \in \mathcal{X} \) such that \( \gamma \succ \beta \).

Nontriviality is consistent with a preference for knowledge. However, it imposes that the decision maker also has some nontrivial consequentialist preference and thus cannot only have a preference for knowledge.

Finally, Savage introduces a continuity assumption that guarantees that one can find sufficiently small events to measure the utility of every act in terms of an arbitrarily chosen outcome.

**P 6 (Event Continuity).** If \( f, g \in \mathcal{A} \) are such that \( f \succ g \) and \( \alpha \in \mathcal{X} \), then there is a finite partition \( \mathcal{H} \) of \( \mathcal{S} \) such that for every \( H \in \mathcal{H} \),

\[
\alpha_H f \succ g
\] and
\[
f \succ \alpha_H g.
\] (8)

While we technically would be able to work with this axiom, we find that it is not quite in the spirit of our paper. Specifically, the change from \( f \) to \( \alpha_H f \) may involve different changes to the information partition than the change from \( g \) to \( \alpha_H g \). We therefore employ a continuity axiom that ensures that preferences are continuous in monotone changes in events on which outcomes are achieved.

We state our continuity properties in terms of convergent sequences of acts in which both the information and the outcomes converge. Stating continuity properties using convergent sequences is quite standard in the context of decisions within a topological space such as consumer choice. To avoid introducing a topological structure on outcomes or acts, we use set-theoretic limits of events. For a sequence \( f^k, k = 1, 2, \ldots \), the notation \( f^k \to f \) means that the acts in the sequence \( f^k \) become arbitrarily similar to \( f \) with the following condition.

---

\(^3\)We note at this point, that empirically, obtaining likelihoods judgments in this manner may be preferable if the exact state cannot be revealed to the subject. In case subjects have a preference for knowledge, then the above way of obtaining likelihood judgments is robust to this preference.
Every outcome \( \alpha \) is either acquired on more and more states or on less and less states as the sequence progresses. Moreover, the set of states in which the outcome obtains is in the limit the same as in \( f \). We make this notion precise in Appendix A.

Continuity of the preference relation is then defined as follows:

**Axiom 6 (Continuity).** If \( f^k \to f \) and \( g^k \to g \) and for all \( k, f^k \succeq g^k \), then \( f \succeq g \).

As sequences of acts converge to an act, the preference is not permitted to “jump” – similar acts must be similarly ranked.

This concludes our adjustments to Savage’s original axioms. In addition, we assume that the intrinsic value of information is separable from the outcomes obtained. In principle, a decision maker may find information especially valuable in case a particular outcome is obtained. For example, knowing about how to drive a car will generally be more useful if one acquires a car. There are many cases in statistics, research, and media consumption in which it is plausible to assume that the benefits from knowledge are separable from the outcomes obtained. Nevertheless, we fully acknowledge that the following axiom is not a pure axiom of rationality.

**Axiom 7 (Learning Independence).** Suppose \( \alpha \sim \alpha', \delta \sim \delta' \) are disjoint outcomes from \( f, g \), and \( E \cap F = \emptyset \). Then,

\[
\alpha_E \alpha_F f \succeq \delta_E \delta_F g \iff \alpha_E \alpha'_F f \succeq \delta_E \delta'_F g
\]

This axiom guarantees that the value of learning whether event \( E \) or \( F \) obtains is unrelated to whether the outcomes \( \alpha \) or \( \delta \) arise on these events.

Our main representation theorem based on the previously introduced axioms characterizes the sum of expected utility and the value of learning whether particular events obtain or not. We call this value a subjective knowledge utility because the value of knowing if an event obtains is subjectively determined. A decision maker may prefer learning about event \( E \) rather than event \( F \) simply because she finds \( E \) more interesting than \( F \).

**Definition 3 (Subjective Knowledge Utility).** \( \succeq \) on the set \( \mathcal{A} \) defined on events \( \mathcal{E} \) has a subjective knowledge utility representation if there exist functions \( U : \mathcal{A} \to \mathbb{R}, u : \mathcal{X} \to \mathbb{R}, \) a monotonely continuous function \( h : \mathcal{E} \to \mathbb{R} \) and a
probability measure \( \mu : E \to \mathbb{R} \) such that \( U(f) \geq U(g) \) if and only if \( f \succsim g \) and

\[
U(a) = \sum_{\alpha \in \mathcal{X}} u(\alpha)\mu(a^{-1}(\alpha)) + \sum_{\alpha \in \mathcal{X}} h(a^{-1}(\alpha))
\]  

(10)

The first component of the representation is the expected utility of the outcomes and will be called the instrumental component or the instrumental preference of the decision maker. The second component contains the utility from gaining knowledge about what event obtains. After choosing act \( a \), the event \( a^{-1}(\alpha) \) is known if outcome \( \alpha \) is achieved. \( h(E) + h(F) \) expresses how much the decision maker enjoys being able to distinguish between the event \( E \) and \( F \). If \( h(E) + h(F) \) is greater than \( h(E \cup F) \), then the decision maker would like to know if \( E \) or \( F \) obtains.

**Theorem 1.** Suppose all conditional acts are outcome-solvable. Then the following statements are equivalent:

1. \( \succsim \) has a subjective knowledge utility representation with a unique probability measure.

2. \( \succsim \) fulfills weak order, the information-neutral sure-thing principle, information-neutral monotonicity, information-neutral likelihood outcome independence, non-triviality, continuity, and learning independence.

By adjusting Savage’s axioms we have therefore obtained a representation that accounts for the value of knowing which event obtains.\(^4\) These preferences encompass a wealth of possible attitudes to information as we will see in section 6. They can express a varying intensity of preference for knowledge about some events but a preference against knowledge for other events. In the following section we therefore analyze the subjective value of knowledge a decision maker attaches to being able to distinguish between events in more detail.

## 5 Value of Knowledge

Regarding risk attitudes, we often distinguish preferences that are risk-seeking, risk-neutral, and risk-averse. A similar classification can also be done with knowledge preferences which we will call the knowledge attitude of a decision

\(^4\)Gilboa and Lehrer (1991) provide conditions such that the preference for knowledge about events expressed by \( h \) can be derived from expected utility maximization with respect to a subsequent decision problem.
maker. A decision maker may exhibit knowledge aversion or knowledge seeking behavior. An individual is knowledge seeking about distinguishing event \( E \) from event \( F \) if for outcomes \( \alpha \sim \alpha', \beta \) we have that \( \alpha E \alpha' \beta \succsim \alpha E \cup F \beta \). Similarly, the decision maker is knowledge averse about these events if the preference is reversed.

If for all nonnull, disjoint events \( E, F \) the decision maker is knowledge seeking (averse), we say simply that the decision maker is knowledge seeking (averse). An expected utility maximizer is both knowledge seeking and knowledge averse and thus knowledge-neutral. Since a set function \( h \) defined on a \( \sigma \)-algebra is subadditive if \( h(E \cup F) \leq h(E) + h(F) \) for all \( E, F \), the following result is straightforward.

**Proposition 1.** A subjective knowledge utility is knowledge-seeking if and only if \( h \) is a subadditive set function. The utility is knowledge-averse if and only if \( h \) is a superadditive set function.

Subadditivity and superadditivity of \( h \) are therefore the defining property of knowledge attitudes. We will see below that this property extends also to comparisons of decision makers.

For a decision maker with a subjective knowledge utility representation \( U \), we define the value of knowledge as follows.

**Definition 4 (Value of Knowledge).** For a decision maker with a subjective knowledge utility representation, the value of distinguishing a finite set of disjoint events \( \{E_1, \ldots, E_n\} \) is

\[
H(E_1, \ldots, E_n) = \frac{U(\gamma^{(1)}_{E_1} \ldots \gamma^{(n)}_{E_n} \delta) - U(\gamma_{E_1 \cup \ldots \cup E_n} \delta)}{\mu(\bigcup_{i=1}^{n} E_i)(U(\gamma) - U(\beta))} = \frac{\left(\sum_{i=1}^{n} h(E_i)\right) - h(\bigcup_{i=1}^{n} E_i)}{\mu(\bigcup_{i=1}^{n} E_i)(U(\gamma) - U(\beta))} \tag{11}
\]

where \( \gamma^{(1)}, \ldots, \gamma^{(n)} \) are indifferent, mutually distinct outcomes.

This definition has the desirable property that it fulfills the identity \( H(E, F, G) = \mu(E \cup F \cup G)H(E \cup F, G) + \mu(E \cup F)H(E, F) \) and thus the value of gaining knowledge about several events can be decomposed into the expected value of knowledge of being able to distinguish between binary events.

In decisions under risk, certainty equivalents are a useful indicator of risk preferences. For knowledge preferences, we define the knowledge-equivalent (KE) as follows:
**Definition 5** (Knowledge Equivalent). At outcome $\beta$, the knowledge equivalent (KE) of whether $E$ or $F$ obtains is the outcome $\gamma$ that fulfills:

$$\beta E \beta F \alpha \sim \gamma_{E \cup F} \alpha$$

for some $\alpha$ and $\beta' \sim \beta$. We denote the knowledge equivalent by $KE(E, F, \beta) = \gamma$.

The following result is straightforward.

**Proposition 2.** Suppose conditional acts are outcome-solvable and the decision maker’s preferences have a subjective knowledge representation. Then the decision maker has a higher value of knowledge for distinguishing $E$ from $F$ than distinguishing $G$ from $H$ if and only if its knowledge equivalent is preferable, i.e.,

$$H(E, F) \preceq H(G, H)$$

$$\Leftrightarrow KE(E, F, \beta) \succeq KE(G, H, \beta) \forall \beta \in X$$

Comparing the knowledge preferences of different decision makers is non-trivial in a subjective model. Decision makers may differ in their utilities of outcomes, $u$, probabilities, $\mu$, and knowledge preferences, $h$. Consider as an example the decision between $\beta E \beta F \delta$ and $\gamma_{E \cup F} \delta$ which a decision maker $1$ may perceive as the decision between a better outcome $\gamma \succ^1 \beta \sim^1 \beta'$ and knowledge about $E$ in case $E \cup F$ comes about. A decision maker $2$ who disagrees on $u$ and $h$ may see this as a decision between a lottery over outcomes $\beta \succ^2 \beta'$ or a safe payoff $\gamma$. Yet another decision maker $3$ may think that $E$ is null and learning about $E \cup F$ is sufficient to know that $F$ obtains. Comparisons between decision makers therefore need to carefully account for the various aspects decision makers can differ on.

To compare knowledge attitudes, we therefore assume that all decision makers agree on the likelihood of events, $\succ^*$. We say that two decision makers $1$ and $2$ have identical instrumental preferences if for all acts $f$ and all acts $g$ such that $\iota(f) = \iota(g)$, we have that $f \succ^1 g$ iff $f \succ^2 g$.

**Definition 6** (Curiosity Relation). For two decision makers with preferences $\succ^1$ and $\succ^2$, yielding identical likelihood relations $\succ^1 = \succ^2*$, and identical instrumental preferences, we say that $\succ^1$ is more curious about (disjoint events)
for all $\alpha, \beta \sim \beta', \gamma \in \mathcal{X}$. We say that $\succeq^1$ is more curious than $\succeq^2$ if the former is more curious than the latter about all disjoint events $E$ and $F$.

The following proposition establishes that a DM is more curious than another if she has a higher knowledge equivalent.

**Proposition 3.** Suppose two decision makers with a subjective knowledge representation agree on $\succeq^*$ and have the same instrumental preferences. Then the following statements are equivalent:

1. $KE^1(E, F, \beta) \succeq KE^2(E, F, \beta)$ for all $\beta \in \mathcal{X}, E, F \in \mathcal{E}$.

2. $\succeq^1$ is more curious than $\succeq^2$.

3. $D(E) = \frac{h^1(E)}{U^1(\gamma) - U^1(\beta)} - \frac{h^2(E)}{U^2(\gamma) - U^2(\beta)}$ is a subadditive function.

The function $D(E)$ can be seen as the difference between the two decision makers in the (normalized) value of knowing whether event $E$ obtains. The equivalence of the second and third statement is very natural given that subadditivity of $h$ implies knowledge-seeking behavior. The equivalence between the two statements means that “higher subadditivity” means a higher curiosity. This is similar to analogous results for risk preferences; the subadditivity of $h$ is with respect to knowledge preferences what the concavity of $u$ is with respect to risk preferences.

Equipped with an understanding of how $h$ represents knowledge preferences, we characterize specific functional forms of $h$ in the following section.

### 6 Characterizations of the Value of Knowledge

While the appeal of the subjective knowledge utility representation lies in its generality, for practical applications it is often desirable to have more restrictive decision models. In this section we axiomatically characterize several functional forms that are more tractable in practical applications.
6.1 Probabilistic Knowledge Utility

A decision maker might only care about the probability of the known events and ignore all other aspects of these events. This is an appealing condition if the event space is very homogeneous and there are no events that are intrinsically more interesting than others. In practical applications it may also be a useful approximation in case the decision maker has already previously selected what information to gain and the primary tradeoff is how much information is gained.

Definition 7 (Probabilistic Knowledge Utility). $\succeq$ on the set $A$ of simple acts on events $E$ has a probabilistic knowledge utility representation if there exist functions $U : A \to \mathbb{R}$, $u : X \to \mathbb{R}$, a continuous function $\hat{h} : [0,1] \to \mathbb{R}$, and a probability measure $\mu : E \to \mathbb{R}$ such that $U(f) \geq U(g)$ if and only if $f \succeq g$ and

$$U(a) = \sum_{\alpha \in X} u(\alpha) \mu(|\alpha|) + \sum_{\alpha \in X} \hat{h}(\mu(|\alpha|))$$  

We call this utility representation a probabilistic knowledge utility because the knowledge about events is only valued in terms of the subjective probability attached to the events. Thus, acts are treated as signals about the state space and no special value is attached to knowing particular events. We formalize the notion of a decision maker only caring about the likelihood of events for their value of knowledge as follows.

Axiom 8 (Principle of Indifference of Information). If $E \sim^* F$, then,

$$\alpha_E \beta_{F \alpha G} \beta_{H \gamma} \sim \beta_{E \alpha F \alpha G} \beta_{H \gamma}.$$  

This principle of indifference of information states that if $E$ and $F$ are equally likely, then the decision maker is indifferent between being informed about whether $E$ or $G$ obtains or being informed about whether $F$ or $G$ obtains. Together with our previous axioms, it characterizes a probabilistic knowledge utility.\footnote{Unlike our first result, this result does not use outcome solvability and therefore does not implicitly assume a continuum of outcomes. We derive this result for a countably infinite number of outcomes for technical convenience. In principle, the result can also be derived for a sufficiently large finite number of outcomes.}

Theorem 2. The following statements are equivalent:

1. $\succeq$ has a probabilistic knowledge utility representation with a unique probability measure.
2. $\succsim$ fulfills weak order, the information-neutral sure-thing principle, information-neutral monotonicity, information-neutral likelihood outcome independence, continuity, learning independence, and the principle of indifference of information.

6.2 Entropic Knowledge Utility

A probabilistic knowledge utility representation is dynamically consistent in the sense that for every event we can find a probabilistic knowledge utility representation that represents the preferences over conditional acts on this event. However, this utility representation does not need to be the one obtained from applying Bayesian updating to the probabilities and then using the same representation. It is in this sense dynamically consistent but not stationary: suppose a decision maker chooses between acts that yield distinct outcomes on events $E, F, G$, which partition the state space. If the decision maker learns that event $E$ does not obtain, updates beliefs according to Bayes’ rule, and then uses equation (16) to evaluate the conditional acts given $F \cup G$, preference reversals may occur. In contrast, the following representation based on the Shannon entropy does not exhibit such preference reversals.

**Definition 8** (Entropic Knowledge Utility). $\succsim$ on the set $\mathcal{A}$ of simple acts on events $\mathcal{E}$ has an entropic knowledge utility representation if there exist functions $U : \mathcal{A} \to \mathbb{R}$, $u : \mathcal{X} \to \mathbb{R}$, a real number $v$, and a probability measure $\mu : \mathcal{E} \to \mathbb{R}$ such that $U(f) \geq U(g)$ if and only if $f \succsim g$ and

$$U(a) = \sum_{a \in \mathcal{X}} u(a)\mu(a^{-1}(a)) - v \cdot \sum_{a \in \mathcal{X}} \mu(a^{-1}(a)) \ln \mu(a^{-1}(a)) \quad (18)$$

The parameter $v$ determines how strong the knowledge preference is and whether the decision maker is knowledge-seeking or knowledge-averse. The entropic knowledge utility is indeed unique in the respect that if $U$ is the entropic knowledge utility representation, then $U$ can also be used to compare conditional acts in $\mathcal{A}_E$ after updating probabilities. It therefore deserves its own characterization which we prepare with the following definitions.

**Definition 9** (Bayesian Updating). A relation $\succsim^*_E$ on the subevents intersecting $E$ is obtained from Bayesian updating of $\succsim^*$ if it is a quantitative probability with representation $\mu_E$ and $\mu_E(F) = \mu(F \cap E) / \mu(E)$.

From the probabilities obtained by Bayesian updating we can derive comparisons between conditional likelihoods. Let $\mathcal{C} = \mathcal{E} \times \mathcal{E}$ be the set of conditional
events, with elements denoted by $E|F$. We can define a likelihood relation on $\mathcal{C}$ as follows. We impose that $E|F \succ^+ G|H$ if and only if $\mu_F(F \cap E) \geq \mu_H(G \cap H)$.\footnote{This definition has a natural analogue using equiprobable partitions of events that does not directly refer to the probability measure $\mu$. Simply define $E|F \succ^+ G|H$ if for partitions with $F^1 \sim^* F^2 \sim^* \ldots \sim^* F^k$ of $F$ and $H^1 \sim^* H^2 \sim^* \ldots \sim^* H^k$ of $H$, we have that for some $l \leq k$, $E \supseteq F^1 \cup \ldots \cup F^l$ and $G \subseteq H^1 \cup \ldots \cup H^l$. Under monotone continuity this is identical to the definition in terms of probabilities.}

The following stationarity condition sharpens the principle of indifference of information and guarantees that using the same utility function on conditional acts after Bayesian updating of beliefs does not yield preference reversals.

**Definition 10 (Stationarity I).** Suppose for all $i \leq k$, $E_i \cup_{i=1}^k E_i \sim^* G_i \cup_{i=1}^k G_i$. Then, for all (not necessarily distinct) $\alpha^1, \ldots, \alpha^k, \beta^1, \ldots, \beta^k \in \mathcal{X} - \{\gamma\}$,

$$\alpha^1_{E_1} \ldots \alpha^k_{E_k} \gamma \succ^*_E \beta^1_{E_1} \ldots \beta^k_{E_k} \gamma \iff \alpha^1_{G_1} \ldots \alpha^k_{G_k} \gamma \succ^*_G \beta^1_{G_1} \ldots \beta^k_{G_k} \gamma$$

(19)

Let $f_F$ be an arbitrary conditional act on the event $F$ and $\psi(f_F)$ an act fulfilling (19) with $H = S$. Thus, $\psi$ maps conditional acts into acts such that the conditional probabilities of the outcomes are unchanged. If $\succ^*$ fulfills stationarity I, then the function $U \circ \psi$ can be used to evaluate conditional acts. This means that the relative value of knowledge remains the same after updating on an event.

The previous definition of stationarity employed the somewhat nonstandard conditional likelihood relation $\succ^+$. We therefore also provide an alternative definition that does not rely on $\succ^+$ but is equivalent given our axioms of a subjective knowledge representation.

**Definition 11 (Stationarity II).** For all events $E, F, G, H$ such that $E \sim F$ and $G \sim H$ and for all (not necessarily distinct) outcomes $\alpha, \beta, \gamma, \delta \in \mathcal{X} - \{\epsilon\}$,

$$\alpha_E \beta_F \epsilon \succ G \beta_E \gamma \delta \epsilon \iff \alpha_G \beta_H \epsilon \succ G \gamma \delta_H \epsilon$$

(20)

The next theorem shows that the two stationarity definitions are equivalent and that the entropic knowledge utility is the only probabilistic knowledge utility that fulfills these conditions.
Theorem 3. Suppose $\succsim$ has a subjective knowledge utility representation and fulfills outcome solvability. Then the following statements are equivalent.

1. $\succsim$ fulfills stationarity I.
2. $\succsim$ fulfills stationarity II.
3. $\succsim$ has an entropic knowledge utility representation.

A good comparison of this result is the result that the only stationary time discounting rule is exponential discounting (Koopmans, 1960). Under exponential discounting of a time-independent utility over commodities, we can use the same utility function on substreams of commodities without preference reversals. Similarly, in an entropic knowledge utility representation, the value of distinguishing an event of (conditional) probability $p$ from an event of probability $1 - p$ is the same irrespective of how much knowledge the decision maker has gained already. This raises the more general question of how the value of knowledge in probabilistic knowledge utility representations changes as additional knowledge is gained, which we address next.

### 6.3 Constant Elasticity of Curiosity Utility

For decision makers with probabilistic knowledge preferences, the value of knowledge can be expressed as a function of probabilities of events.

$$H(E, F) = \hat{H}(\mu(E), \mu(F)) = \frac{\hat{h}(\mu(E)) + \hat{h}(\mu(F)) - \hat{h}(\mu(E \cup F))}{\mu(E \cup F)(U(\gamma) - U(\beta))}$$  \hspace{1cm} (21)

Using the function $\hat{H}$, we can express how the value of knowledge changes as events become more or less likely. Since the value of distinguishing $E$ and $F$ is only meaningful in case $E \cup F$ is already known, this at the same time expresses how the value of knowledge changes as more knowledge is gained. We thus define the the elasticity of curiosity as the percentage change of the value of knowledge for a percentage change in likelihood.

**Definition 12 (Elasticity of Curiosity).** The coefficient of the elasticity of curiosity at probability $p$ for distinguishing events of probabilities $q$ and $1 - q$ is

$$el(p, q) = \frac{\partial \hat{H}(pq, p(1 - q))}{\partial p} \frac{p}{\hat{H}(pq, p(1 - q))}.$$  \hspace{1cm} (22)
The coefficient expresses how many percent the value of knowledge changes for a one percent increase in probability $p$ of acquiring the information. The relative probabilities of the events to be distinguished, $q$ and $1 - q$, are held constant in this comparison. It is straightforward to verify that entropic knowledge utility preferences have a zero elasticity of curiosity.

For many decision makers, the elasticity of curiosity may be nonzero. A decision maker who is knowledge seeking and has a high elasticity of curiosity will generally only choose to give up instrumental value for a coarsely grained partition of the state space. A decision maker with an entropic knowledge utility representation may also be willing to forego instrumental value to achieve a more detailed partition of the state space. A decision maker may even find a book “addictive” in the sense that reading the second half of the book (allowing to distinguish between equally likely events $E$ and $F$) has a higher knowledge equivalent than reading the first half (allowing to distinguish between the equally likely events $E \cup F$ and $\overline{E \cup F}$). This is the case if the elasticity of curiosity is negative.

A functional form that is mathematically tractable and can account for both the intensity of knowledge preferences and different elasticities of curiosity is therefore desirable. The following two-parameter family of preferences achieves this:

**Definition 13** (Constant Elasticity of Curiosity (CEC) Utility). $\succsim$ on the set $A$ of simple acts on events $E$ has a constant elasticity of curiosity representation if there exist functions $U : A \to \mathbb{R}$, $u : X \to \mathbb{R}$, real numbers $r$ and $v$, and a probability measure $\mu : E \to \mathbb{R}$ such that $U(f) \geq U(g)$ if and only if $f \succsim g$ and

$$U(a) = \sum_{\alpha \in X} u(\alpha) \mu(a^{-1}(\alpha)) - v \cdot \left( \sum_{\alpha \in X} \mu(a^{-1}(\alpha))^r - 1 \right)$$  \hspace{1cm} (23)

The above representation consists of the expected utility and the Rényi (1961) entropy. $v$ is a parameter that determines whether the decision maker is knowledge seeking or knowledge averse and how strong this preference is compared with the instrumental preferences. $r$ is a parameter that determines the elasticity of curiosity.

Table 2 provides an example of CEC preferences for a risk neutral decision maker. The willingness to pay is only approximate for greater clarity of exposition. The decision maker has the choice to acquire knowledge of distinguishing event $E$ from event $F$ in case either of the two obtains. In situation A, the
The decision maker reports a willingness to pay for the information of 200 EUR. The decision maker is sure that she will learn that either $E$ or $F$ obtains since the probabilities $q$ and $1-q$ sum to one. In situation B, the events $E$ and $F$ only make up 98% probability but their relative likelihood is the same as in A. Despite learning to distinguish between $E$ and $F$ being 2% less likely, the willingness to pay only decreases by 1%. Comparing situations C and D, we observe the same pattern; a 2% decrease in the likelihood of distinguishing $E$ from $F$ leads to a 1% decrease in the willingness to pay. Note that the relative likelihood of $E$ and $F$ may differ in situations C and D from situations A and B.

We now justify the name CEC preferences for the functional form given above.

**Proposition 4** (Characterization of CEC Preferences). Suppose a decision maker’s preferences have a probabilistic knowledge utility representation. Then the following statements are equivalent:

1. The elasticity of curiosity $e_l(p, q)$ is constant in $p$.

2. The preferences have an entropic knowledge utility representation or a constant elasticity of curiosity representation.

In some sense, one can see CEC preferences as corresponding to CRRA risk preferences and entropic knowledge utility preferences as corresponding to CARA risk preferences. CARA preferences are invariant under additions and subtractions of income – the willingness to pay for a gamble does not depend on the starting income. Entropic knowledge utility preferences are invariant under additions or removal of initial information - the willingness to forego good outcomes in favor of more information in case of an event does not depend on the likelihood of this event. CRRA preferences change in a consistent manner as income changes while CEC preferences change in a consistent manner as initial information changes.
6.4 Stationary Information Cost Utility

So far, the characterized functional forms of knowledge preferences treat all information about the event space symmetrically. Often, a decision maker may however only be interested in a particular random variable and uninterested in other knowledge that can be gained. For example, an econometrician may be interested in a parameter of a distribution but not interested in every data point generated from this distribution. Nonetheless, knowledge about the data points may be useful to acquire knowledge about the parameter.

We make this idea precise using random variables. A variable $V : S \rightarrow S_V$ is a measurable map from the state space $S$ into the state space of the variable, $S_V$ with $\sigma$-algebra $\mathcal{E}_V$. That is, for all $E_V \in \mathcal{E}_V$, there exists an event $E \in \mathcal{E}$ such that $E = V^{-1}(E_V)$. The state space of the variable, $S_V$ may for example specify a parameter value of an econometric model the decision maker is uncertain about. We define $\mu^V_E = \mu_E \circ V^{-1}$ as the probability measure over the variable given event $E \in S$.

A natural way to evaluate the knowledge gained from an act is to simply attach a value to each possible posterior probability measure over the random variable generated by the knowledge about events and calculate its expected value.

**Definition 14** (Stationary Information Utility). $\succeq$ has a stationary information representation over variable $V$ if for all events $E \in \mathcal{E}$, the conditional relation $\succeq_E$ can be represented by

$$U_E(a_E) = \sum_{\alpha} \mu_E(a_E^{-1}(\alpha)) \left( u(\alpha) + v(\mu^V_{a_E^{-1}(\alpha)}) - v(\mu^V_E) \right)$$

In words, the decision maker perceives a tradeoff between expected utility and the expected valuation of the information gained about the variable $V$.

Analogous to the Bernoulli utility function $u$ for the evaluation of outcomes, the function $v$ evaluates the value of the information $\mu^V_{a_E^{-1}(\alpha)}$ gained by learning that outcome $\alpha$ obtains and compares it with the value of the information $\mu^V_E$ the decision maker had initially.\textsuperscript{8} What is special about stationary information utility

\textsuperscript{7}To clarify that the evaluation of the posterior probability measures over the variable $V$ do not depend on the prior, we state this and the following representation as representations for all conditional acts. Since for a subjective knowledge representation the preferences over acts determine the preferences over conditional acts, this is without loss of generality.

\textsuperscript{8}The representation is unique up to joint linear transformations of $u$ and $v$ and separate additive transformations of $u$ and $v$. The functional form is chosen such that $U_E(\gamma_E) = u(\gamma)$. 

24
is that the evaluation of the posterior does not depend on the prior. An example of a stationary information representation is the mutual information in which each conditional measure is valued according to \( v(\mu^V_E) = \int \mu^V_E(\theta) \ln \mu^V_E(\theta) d\theta \).

Similarly to the entropy, in mutual information the value of knowing about an event of the variable only depends on the likelihood of the event of the variable. However, in some cases this is not a plausible condition. For example, we may be more interested in the first than the 100th digit of a uniform random variable. If the state space of the variable is more structured, other interesting functional forms of \( v \) are possible such as the posterior variance of a real-valued parameter. If \( S_V = \mathbb{R} \), i.e., if the parameter is real-valued, then using (the negative of) the variance, \( v(\mu^V_E) = \int \mu^V_E(\theta)(\theta - \int \theta \mu^V_E(\theta') d\theta')^2 d\theta \), is a plausible way of measuring how informed the decision maker is about \( V \).

A stationary information representation is only sensible under the assumption that it is costless for the decision maker to process the information. If a decision maker is for example interested in the mean of a random variable, the decision maker may rather choose to take an action that estimates the mean from a data set instead of choosing an act that informs the decision maker about all data in the data set. We can account for this by introducing a cost of “too much information”.

**Definition 15 (Stationary Information Cost Utility).** \( \succcurlyeq \) has a stationary information cost representation over variable \( V \) if for all events \( E \in \mathcal{E} \), the conditional relation \( \succcurlyeq_E \) can be represented by

\[
U_E(a_E) = \sum_a \mu_E(a^{-1}_E(a)) \left( u(a) + v(\mu^{-V}_{a^{-1}_E(a)}) - v(\mu^V_E) - r \cdot \ln(\mu_E(a^{-1}_E(a))) \right)
\]

The cost of too much information is expressed via the Shannon entropy of the information partition. If \( r \) is negative, then the decision maker perceives a tradeoff between gaining as much information as possible about the variable \( V \) and having an information partition that is as simple as possible.

In the remainder of this section, we characterize the stationary information cost utility. The stationary information utility then follows trivially from the additional condition that the decision maker only cares about knowledge of events in \( V \) as stated by the following condition.

Omitting the term \(-v(\mu^V_E)\) from the representation yields identical conditional preferences but does not fulfill this condition.
**Definition 16** (Indifference to Irrelevant Information). Suppose $E$ and $F$ fulfill $\mu^V_E = \mu^V_F$ and are nonnull. Then,

$$a_{E \cup F} \sim a_E a'_F b.$$  

(26)

Thus, if two events yield the same conditional probability measure (and thus the same information) about the variable, then being able to distinguish between the two events is irrelevant to the decision maker.

We adjust the stationarity conditions to account for the fact that not only the likelihoods of events are important but also how well these correlate with events in $\mathcal{E}_V$.

**Definition 17** (Stationarity I*). Suppose for all $i \leq k$, $E_i|F \sim^* G_i|H$ and $\mu'_{E_i} = \mu'_{G_i}$. Then, for all (not necessarily distinct) $\alpha^1, \ldots, \alpha^k, \beta^1, \ldots, \beta^k \in \mathcal{X} - \{\gamma\}$,

$$\alpha^1_{E_1} \ldots \alpha^k_{E_k} \gamma \succ \beta^1_{E_1} \ldots \beta^k_{E_k} \gamma \Leftrightarrow \alpha^1_{G_1} \ldots \alpha^k_{G_k} \gamma \succ \beta^1_{G_1} \ldots \beta^k_{G_k} \gamma$$

(27)

The corresponding change to the second stationarity axiom is the following condition.

**Definition 18** (Stationarity II*). Suppose for all events $E, F, G, H$ if $E \sim F$ and $G \sim H$ as well as $\mu'_{E} = \mu'_{G}$ and $\mu'_{F} = \mu'_{H}$. Then for all (not necessarily distinct) $\alpha, \beta, \gamma, \delta \in \mathcal{X} - \{\epsilon\}$,

$$a_E b_F c \succ \gamma d F E \Leftrightarrow a_G b_H c \succ \gamma d H E$$

(28)

We now state a similar result to the previous theorem that yielded an entropic knowledge representation but allow for a decision maker’s special interest in $V$. We can understand this stationarity condition using the following simple example. Let’s say a decision maker is willing to pay a certain amount of money for information about a parameter. Suppose now somebody offers to flip a coin and sell the information only in case the coin comes up “heads” but neither payment nor information is exchanged on “tails”. In case the decision maker’s willingness to pay changes, this would violate stationarity.

The additive separability of the evaluation of prior and posterior distribution turns out to be intrically linked with the stationarity of the value of knowledge.

**Theorem 4.** Suppose $\succ$ has a subjective knowledge utility representation and fulfills outcome solvability. Then the following statements are equivalent.
1. $\succsim$ fulfills stationarity $I^*$.  
2. $\succsim$ fulfills stationarity $II^*$.  
3. $\succsim$ has a stationary information cost utility representation.

In stationary information representations and stationary information cost representations the actions become more or less valuable via the knowledge gained about the variable $V$. This gives the actions taken by the decision maker the character of estimators. If $V$ is an unknown parameter of a distribution or an economic model, then the data $\mathcal{E}$ may be used to inform the decision maker about this parameter. This leads us very naturally to the problem of statistical estimation in general. In the following section, we therefore briefly sketch how certain practices of statistics reveal certain aspects of knowledge preferences.

7 Partition Problems

Savage (1954) considered statistics as the practice of solving partition problems. A partition problem is the problem of determining whether an event $E$ or an event $F$ obtains. Our representation theorem for subjective knowledge utility provides a subjective value of solving partition problems. This provides a novel perspective on partition problems in general because statistical estimators themselves can be considered as acts in our model. This in turn allows us to impose our axioms on the preferences over estimators and analyze these preferences using the tools developed in the previous sections. To see this, we define in the language of our model estimation problems, data collection problems, and hypothesis testing problems in the following paragraphs.

Let $\mathcal{E}_V$ be a $\sigma$-algebra on a parameter space $S_V$. From the perspective of the decision maker, the parameter is a variable $V$ of the more general state space $S$ of states of the world. For simplicity, we assume that $S = \prod_{i=1}^n S_i$ consists of $n$ realizations of a distribution based on the parameter space.\footnote{In fact, we could assume an arbitrary, rich event space and let the variable and thus the parameter space the decision maker has in mind be fully subjective and only be revealed via the knowledge preferences.} We endow $S$ with a product $\sigma$-algebra $\mathcal{E}$. The set of outcomes $\mathcal{X} = \mathbb{R}_+ \times \mathcal{R}$ consists of elements $(c, \hat{r})$ containing a monetary cost $c \in \mathbb{R}_+$ and estimation results $\hat{r} \in \mathcal{R}$, where $\mathcal{R}$ is the set of possible estimation results.\footnote{Given our way of modeling knowledge with state-independent evaluation of outcomes, the estimation results are simply messages from a signal and are only meaningful in conjunction with the parameters.} The set of estimators is simply the
set of acts in this model. An estimation problem is a choice from the set of estimators given a subjective knowledge utility.

The concept of an estimation problem is rather abstract. While solutions to specific estimation problems are beyond the scope of our paper, we show how certain assumptions employed in statistics can be readily translated into our framework.

Curiosity about parameter: The decision maker attaches a positive value to information about the variable \( V \) but a nonpositive value to information about data that does not reveal additional information about \( V \). A stationary information cost utility is an example that exhibits such preferences.

Distribution assumptions: If the decision maker believes that the observations are independent and identically distributed, then this is straightforward to impose on the beliefs \( \mu \) about events and the variable \( V \). Moreover, for such estimation problems it is sensible to impose that the \( h \) function treats each observation symmetrically. Similarly, all other assumptions about the relation between parameters and observations are either expressed in the beliefs \( \mu \) or the variable \( V \). Since \( V \) is only revealed by the knowledge preferences, the distribution assumptions can be fully subjective as they are fully determined by the subjectively determined \( h \) and \( \mu \). \( \mu^V \) is the prior belief of the decision maker about the parameter. The likelihood that event \( E \in \mathcal{E} \) is observed from an observation given an event \( E_V \) in the parameter space is the conditional probability

\[
\mathbb{P}(E|E_V) = \frac{\mu(E)\mu^V(E_V)}{\mu^V(E_V)}.
\]

Hypothesis testing: A hypothesis testing problem is an estimation problem in which the decision maker is interested in distinguishing whether a hypothesis about the parameters is true or not. Thus, we have a null hypothesis \( E_V \in \mathcal{E}_V \) and its alternative \( \overline{E}_V \) for which the decision maker attaches a positive value of knowledge, \( H(V^{-1}(E_V), V^{-1}(\overline{E}_V)) > 0 \), and the value of any further partitioning of \( E_V \) is nonpositive. In classical hypothesis testing, the value of \( h(V^{-1}(E_V)) \) for some \( E_V \in \mathcal{E}_V \) only depends on the likelihoods of observations given the parameters, \( l(E|E_V) \) but not on prior probabilities.

Properties of maximum likelihood estimation: Contrary to Bayesian estimators, maximum likelihood estimators do not depend on the prior probabilities \( \mu^V \). In our framework, this means that in most cases, a maximum likelihood estimator can only be a solution to an estimation problem if \( h \) has a negative elasticity of curiosity. The less likely an event in \( \mathcal{E}_V \) is, the greater must be the value of
knowledge attached to that event. Moreover, the elasticity must be -1 because the value of distinguishing two equally, highly likely events must be the same as distinguishing two equally, but less likely events. This in turn means that the value of knowing the exact parameter is infinite if $E_V$ is atomless. Maximum likelihood estimation therefore may therefore violate our continuity assumption in case the decision maker attaches zero prior probability to some parameter values. However, it may be approximated by an elasticity of curiosity that approaches -1.

Data collection costs: A data collection problem is an estimation problem involving a choice from a subset of the set of estimators in which the cost of the estimator depends on the number of observations used. Note that an estimator $f$ need not depend on all the data. An estimator uses at most the first $\hat{n}$ observations if for all outcomes $(c, \hat{r})$, the preimages $f^{-1}(c, \hat{r})$ are cylinder sets of the form $S_V \times \prod_{i=\hat{n}+1}^{\hat{n}} S_i$ with $S_V \subseteq \prod_{i=1}^{\hat{n}} S_i$. A data collection cost function $c : \mathbb{N} \to \mathbb{R}$ specifies the monetary cost associated with the number of observations used. A data collection problem is a set of estimators $A'$ such that for all $a \in A'$, the outcome $(c, \hat{r})$ is only in the support of $a$ if $c = c(\hat{n}(a))$. The decision maker therefore faces a tradeoff between using more observations to improve the information about $V$ and the cost associated with using these observations.

The characterization of statistical estimators from assumptions on the value of information may be an exciting avenue for further research. From a normative perspective, this may be more interesting for a classical statistician than for a Bayesian statistician as subjective knowledge utility appears to justify a greater variety than expected-utility motivated Bayesian statistics. The revealed preference perspective may however be even more interesting: what does the number of test subjects in an experiment and their pay tell us about the researcher’s value of knowledge of the hypothesis being tested? Do the efforts and resources spent indicate that society places a higher value of knowledge on world temperatures in 2050 than on knowledge about the lives of certain celebrities? Subjective knowledge utility can address these questions within a unified decision theoretic framework.
8 Discussion

In this section, we discuss alternative ways of modeling the gain of knowledge in decisions under uncertainty and implications of our work for the necessity of Savage’s axioms for rational decision making under uncertainty.

It is important to realize that the standard way of modeling decisions under uncertainty that we employed comes with some underlying maintained assumptions. The most crucial implicit assumption is that any outcome $\alpha$ can be realized on any event. The following property therefore directly follows from the way decisions under uncertainty are modeled:

$\textbf{P o (Outcome Permutability).}$ If $\sigma : f(S) \rightarrow f(S)$ is a permutation of the outcomes resulting from an act $f \in A$, then permuting the outcomes of $f$ yields also an act $\sigma \circ f \in A$.

Thus, starting from any act we can permute all outcomes and obtain another act. This assumption is implicit in almost all of axiomatic decision theory in one form or another. However, this imposes severe restrictions on the way the knowledge gained from an act (and thus its value) can be modeled. For example, defining an outcome as “knowing event $E$ obtains” is inconsistent with outcome permutability — under outcome permutability, this outcome could also result from event $\overline{E}$ implying that the decision maker knows (and therefore has the correct belief) that $E$ obtains when in fact it does not obtain! Including beliefs in the description of outcomes faces similar issues. In this case, the value of “believing event $E$ obtains” should be state-dependent, depending on whether $E$ actually obtains or not.\footnote{Savage responded to the critique that the evaluation of outcomes should be state dependent with the famous bonmot “I should not mind being hung so long as it be done without damage to my health and reputation”. If the reason for state dependence is the inclusion of beliefs into outcomes, the bonmot would go “I should not mind being hung as long as I believe that I am not being hung”.}

It seems therefore that if one accepts outcome permutability and wants to avoid state-dependent utility, then knowledge and beliefs about events cannot be (part of) outcomes.\footnote{For difficulties with including beliefs into outcomes to explain anticipatory feelings, see the extensive discussion in Eliaz and Spiegler (2006).} It is for this reason that we assumed that knowledge about which event obtains is derived from the chosen act and the observed outcomes.

Another possible way of modeling knowledge would be to explicitly introduce an information partition in the description of an act. However, consider an act that yields outcome $\alpha$ on event $E$ and $\beta$ on event $\overline{E}$ but has a trivial
information partition. In this act, the decision maker could refine the information partition to \( \{E, \overline{E}\} \) by realizing that according to her chosen act outcome \( \alpha \) occurs if and only if event \( E \) obtains. Thus, either the outcome must be unobservable after it is realized or the decision maker must “forget” what act she has chosen. Our way of modeling knowledge avoids such issues – indeed, the knowledge gained by the decision maker is exactly what can be concluded from observing an outcome and remembering which act has been chosen.

If the outcomes in Savage’s axiomatization cannot contain descriptions of knowledge or beliefs without running into conceptual difficulties, then expected utility as axiomatized by Savage does not fulfill our normative position stated in section 4. This is because Savage’s axiomatization of expected utility cannot properly account for an intrinsic preference for knowledge. Moreover, in case of an instrumental preference for knowledge (to improve future decisions), Savage’s axiomatization makes excessive requirements on the decision maker to integrate the future decision problems into one grand decision problem. Our model allows decision makers to integrate a preference for knowledge into their preferences without specifying the exact way in which this knowledge will be used.

Our results extend previous critiques of the necessity of Savage’s axioms for rational decisions (Gilboa, Postlewaite, & Schmeidler, 2009, 2012). However, the motivation for Gilboa et al. (2012) to reject the sure-thing principle refers to the inability of the decision maker to fix a precise prior. A tough-minded expected utility maximizer may state that this inability is simply a mistake. Our critique instead refers to the tastes of the decision maker for knowledge. De gustibus non est disputandum - clearly a theory of rational choice should allow a decision maker to give up a good night’s sleep in return for finding out who the murderer on the orient express is. It is up to further research how expected utility can be salvaged as a normatively compelling criterion when decision makers have a preference for knowledge.

9 Concluding Remarks

While the instrumental value of information has been covered to great extent in the literature, it is perhaps surprising that the standard model of rational choice does not account for the value of knowledge simpliciter. This is especially striking as an intrinsic preference for knowledge is present in many contexts
of everyday life, be it media, news, and entertainment consumption, scientific exploration, education, and social relationships. With this paper, we hope to provide a starting point for a systematic analysis of rational decisions in the presence of a subjective value of knowledge.
A Monotone Sequences of Acts and Events

In this section of the appendix, we make our definitions regarding monotone continuity precise.

Definition 19 (Monotonely Continuous Function). A function \( h : \mathcal{E} \to \mathbb{R} \) is monotonely continuous if for any convergent monotone sequence of events \( E^1 \subset E^2 \subset \ldots \to E \), we have that \( \lim_{k \to \infty} h(E^k) = h(E) \).

Definition 20 (Monotone Sequence). A sequence of acts \( \{f^k\}_k \) is monotone in outcome \( \alpha \) if \( k \geq l \) implies \( (f^k)^{-1}(\alpha) \subseteq (f^l)^{-1}(\alpha) \) or if \( k \geq l \) implies \( (f^k)^{-1}(\alpha) \subseteq (f^l)^{-1}(\alpha) \).

A monotone sequence of acts therefore makes a particular outcome either increasingly likely or increasingly unlikely by adding or removing states in which the event obtains.

Definition 21 (Monotone Convergence). \( \{f^k\}_k \) converges monotonically to \( f \), denoted \( f_k \to f \), if for all \( \alpha \in \text{supp}(f) \), \( \{f^k\}_k \) is monotone in \( \alpha \) and \( \bigcup_k (f^k)^{-1}(\alpha) = f^{-1}(\alpha) \) or \( \bigcap_k (f^k)^{-1}(\alpha) = f^{-1}(\alpha) \).

A sequence of acts converges monotonically to an act \( f \) if it is a monotone sequence and the set-theoretic limit of the events on which each outcome is obtained is the event on which \( f \) obtains that outcome.

B Proof of Theorem 1

The proof proceeds as follows. First, we show that any set of conditional acts forms a linear continuum. This is highly nontrivial because the value of knowledge prevents us from using standard results. Since in a linear continuum the order topology is connected, this allows us to use additive representation theorems to obtain additive separability of utility across conditional acts. Next, for arbitrary information partitions, we derive an additive utility representation for conditional acts that in turn depends on the additive utility representations of its conditional acts. This allows us to use the uniqueness of additive representations to further refine the representation into the desired subjective knowledge utility.

Lemma 1. If \( E \in \mathcal{E} \) is null, then for all \( f_E, g_E \in \mathcal{A}_E \), \( f_E \sim_E g_E \).
Proof. We prove this by outcome solvability. If $f_E \succ_E g_E$, then for some outcomes $\gamma, \beta, \gamma_E \succ_E \beta_E$ and thus $E$ is not null.

**Definition 22 (Supremum).** $f$ is a supremum of $S'$ under relation $\succ$, denoted $f \in \sup_\succ (S')$ if for all $g \in S'$, $f \succ g$ and for all $h$ such that $h \succ g \forall g \in S'$, $h \succ f$.

**Definition 23 (Linear Continuum).** A set $S$ ordered by $\succ$ is a linear continuum if for all $f, h \in S$ and all $S' \subseteq S$:

$$f \succ h \Rightarrow \exists g : f \succ g \succ h$$ (29)

$$\sup_{S'} \subseteq S.$$ (30)

**Definition 24 (Interval of Acts).** An interval $I[f, h]$ of acts between $f$ and $h$ is defined as the set of all acts that are part of some monotone sequence from $f$ to $g$, i.e.,

$$I[f, h] = \left\{ g : \forall \alpha \in \mathcal{X} : \left( f^{-1}(\alpha) \subseteq g^{-1}(\alpha) \subseteq h^{-1}(\alpha) \right) \lor \left( h^{-1}(\alpha) \subseteq g^{-1}(\alpha) \subseteq f^{-1}(\alpha) \right) \right\}.$$ (31)

**Lemma 2.** If $f \succ h$, and $I[f, h] \neq \emptyset$, then there exists $g \in I[f, h]$ such that $f \neq g \neq h$.

**Proof.** By Lemma 1 and $f \succ h$, it cannot be the case that $f$ and $h$ are identical on all nonnull events and therefore there must be an outcome $\alpha \in \mathcal{X}$ such that $f^{-1}(\alpha) - h^{-1}(\alpha)$ is nonnull. Since $\mathcal{E}$ contains no atoms, we can find a non-null proper subset, denoted $A$, of $f^{-1}(\alpha) - h^{-1}(\alpha)$. Then $A_A h$ is the desired act on the interval.

**Lemma 3.** For any two acts $f, h$, there exists a finite sequence of nonempty intervals $I[f, g'], I[g', g''], \ldots, I[g'', h]$.

**Proof.** This follows from the assumption that all acts are simple. Since the acts are simple, there are finitely many events on which the two acts differ. If two acts differ only on a single event, it is straightforward to show that a nonempty interval exists. We can therefore construct the finite sequence of intervals by changing a single event on each interval.

**Lemma 4 (Villegas (1964), Theorem 5).** In a qualitative probability $\sigma$-algebra the following propositions are equivalent:
1. There are no atoms;

2. Every event can be partitioned into two equally probable events.

**Lemma 5.** Let $\mathcal{I}[f', g']$ be a nonempty interval. For any bounded subset $I'$ of the indifference curves $\mathcal{I}[f', g']/\sim$, there exists an interval $I^0$ between acts containing a bound and elements worse or indifferent to the indifference classes of $I'$.

**Proof.** Assume the bound is an indifference class containing the act $h$. We consider the interval $\mathcal{I}[g', h]$. If we have that $\mathcal{I}[g', h]$ contains indifference classes worse or equal to indifference classes in $I'$, then $I^0 \equiv \mathcal{I}[g', h]$. Otherwise, $\mathcal{I}[g', h]$ contains only elements that are strictly preferred to all elements of the indifference classes in $i \in I'$. Then $g'$ is a bound for $I'$ and for any element $i \in I'$ with representative act $h'$, we can choose $I^0 \equiv \mathcal{I}[h', g']$. \qed

**Lemma 6.** For any nonempty $\mathcal{I}[f, h]$, there exists an element $g \in \mathcal{I}[f, h]$ s.t. the following holds

1. For $\alpha \in \mathcal{X}$ with $f^{-1}(\alpha) \subseteq g^{-1}(\alpha) \subseteq h^{-1}(\alpha)$, we have $g^{-1}(\alpha) - f^{-1}(\alpha)$ is as likely as $h^{-1}(\alpha) - g^{-1}(\alpha)$.

2. For $\alpha \in \mathcal{X}$ with $h^{-1}(\alpha) \subseteq g^{-1}(\alpha) \subseteq f^{-1}(\alpha)$, we have $h^{-1}(\alpha) - g^{-1}(\alpha)$ is as likely as $f^{-1}(\alpha) - g^{-1}(\alpha)$.

We call this element $g$ the half distance element for $\mathcal{I}[f, h]$.

**Proof.** This directly follows from Villegas (1964), Theorem 5. \qed

**Lemma 7.** If $f \succ h$, and $\mathcal{I}[f, h] \neq \emptyset$, then there exists $g \in \mathcal{I}[f, h]$ such that $f \succ g \succ h$.

**Proof.** Let $g^0 \equiv f$ and $g^0 \equiv h$. Define $\mathcal{I}_{1/2}[f, g]$ as the set of half-distance elements between $f$ and $g$. Define $g^i$ as an arbitrary element of $\mathcal{I}_{1/2}[\hat{g}^{i-1}, \hat{g}^{i-1}]$. If $f \succ g^i \succ h$, then we have found the desired element $g$. If $g^i \sim f$, then define $\hat{g}^i \equiv g^i$ and $\hat{g}^i \equiv \hat{g}^{i-1}$. If $h \sim g^i$, then define $\hat{g}^i \equiv \hat{g}^{i-1}$ and $\hat{g}^i \equiv g^i$. We obtain the monotone sequences $\hat{g}^i$ and $\hat{g}^i$. If the sequences do not terminate at some element $g$, then they converge to acts $\hat{g}^*$ and $\check{g}^*$, respectively, which may only differ on null events. By Lemma 1 we have that $\hat{g}^* \sim \check{g}^*$. It follows from monotone continuity that $\hat{g}^* \succ f$ and $h \succ \check{g}^*$, a contradiction. It follows that the sequence terminates at some element $g$ such that $f \succ g \succ h$. \qed

**Lemma 8** (Least Upper Bound Property). For every nonempty subset $I$ of an interval $\mathcal{I}[f, h]$ that is bounded by $g$ with respect to $\succ$, there exists a supremum $g^*$. 35
Proof. By Lemma 5, there exists an interval $\mathcal{J}([\hat{g}^{0}, \tilde{g}^{0}])$ such that for all $x \in I$, $\hat{g}^{0} \preceq x$ and for some $y \in I$, $y \succeq \tilde{g}^{0}$. Define $\mathcal{J}_{1/2}([\hat{g}^{0}, \tilde{g}^{0}])$ as the set of half-distance elements in this interval. Define $g^{i}$ as an arbitrary element of $\mathcal{J}_{1/2}([\hat{g}^{i-1}, \tilde{g}^{i-1}])$. If for all $x \in I$, $g^{i} \succ x$, then define $\hat{g}^{i} \equiv g^{i}$ and $\tilde{g}^{i} \equiv \hat{g}^{i-1}$. If there exists $y \in I$ such that $y \succeq g^{i}$, then define $\hat{g}^{i} \equiv \hat{g}^{i-1}$ and $\tilde{g}^{i} \equiv g^{i}$. We obtain the monotone sequences $\hat{g}^{i}$ and $\tilde{g}^{i}$. The sequences converge to acts $\hat{g}^{*}$ and $\tilde{g}^{*}$, respectively, which may only differ on null events and are thus indifferent by Lemma 1. Clearly, $g^{*} \equiv \hat{g}^{*}$ is a supremum of $I$, since if $g^{*} \succ x \in \mathcal{J}[f, h]$, then there exists some $y \in I$ and $\tilde{g}^{i}$ s.t. $y \preceq \tilde{g}^{i} \succeq x$. □

**Lemma 9** (Intervals are Linear Continuums). If $\mathcal{J}[f', h']$ is nonempty, then its set of indifference curves $\mathcal{J}[f', h']/\sim$ is a linear continuum.

**Proof.** This follows since the relation $\succeq$ has the upper bound property and for all elements $f, h \in \mathcal{J}[f', h']$, there exists an element $g$ such that $f' \succ g \succ h'$. □

**Lemma 10** (Connectedness). The order topology on the set of conditional acts on $E$ is connected.

**Proof.** Suppose we can partition the set of conditional acts into two open sets, $A, B$. Take an arbitrary act from each set. There exists a finite sequence of nonempty intervals between the sets and each interval is a linear continuum. The union of these intervals forms a linear continuum $U$ as well. But then by continuity $A \cap U$ and $B \cap U$ partition the linear continuum, a contradiction. □

Using Connectedness, we can now obtain an additive representation $U(f_{E}g_{E}) = u_{E}(f_{E}) + u_{E}(g_{E})$. From here, we use the remaining axioms to obtain a representation of the form $U(f) = \sum_{a \in O} h(f^{-1}(a))U(a) + h(f^{-1}(a))$

We call any partition $\mathcal{P}_{E}$ of $E$ that has at least three elements and is a subset of $E$ an admissible partition of $E$.

**Lemma 11** (Additive Representation). For every event $E$ and admissible partition $\mathcal{P}_{E}$, there exists a representation $U_{\mathcal{P}_{E}} : \prod_{E' \in \mathcal{P}_{E}} A_{E'} \to \mathbb{R}$ of $\succeq_{E}$

$$U_{\mathcal{P}_{E}}(f_{E}) = \sum_{E' \in \mathcal{P}_{E}} u_{\mathcal{P}_{E}}(f_{E'}, E').$$

(32)

**Proof.** Choose any admissible partition $\mathcal{P}_{E}$. Note that for every element $E'$ of $\mathcal{P}_{E}$ the set of acts $\{f_{E'} : E' \to \mathbb{R}\}$ is connected under the $\succeq_{E'}$-order topology. We now assume that $a', \beta', \text{etc.}$ can only occur on $E' \in \mathcal{P}_{E}$, $a'', \beta''$, etc. can only occur on $E'' \neq E'$ and so on. Since outcomes $a' \sim a''$ can be substituted
for another by our monotonicity assumption and since we have countably
many indifferent outcomes for each outcome, this is without loss of generality.
By the information-neutral sure-thing principle, we have jointly independent
preferences over the product space $\prod_{E' \in \mathcal{P}_E} \mathcal{A}_{E'}$. Since preferences over these are
continuous in the order topology, it follows by Wakker (1988b) that an additive
representation of the form
\[ U_{\mathcal{P}_E}(f_E) = \sum_{E' \in \mathcal{P}_E} u_{\mathcal{P}_E}(f_{E'}, E') \quad (33) \]
exists.

**Lemma 12 (Outcome Additive Representation).** For every event $E$, there exists a
representation $U_E : \mathcal{A}_E \to \mathbb{R}$ of $\succeq_E$ of the form:
\[ U_E(f_E) = \sum_{\alpha \in \mathcal{X}} u_E(\alpha, f_E^{-1}(\alpha)) \quad (34) \]

**Proof.** Let $\mathcal{P}'_E$ be a refinement of $\mathcal{P}_E$. Then $U_{\mathcal{P}_E}$ and $U_{\mathcal{P}'_E}$ must be affine transfor-
mations of another by the uniqueness of additive representations over $\prod_{\mathcal{P}'_E} \mathcal{A}_{E'}$. We may assume without loss of generality (by simply applying this affine trans-
formation to one of the representations), that they are identity transformations of another on the domain $\prod_{\mathcal{P}'_E} \mathcal{A}_{E'}$. Thus, for all $f$, such that the decision maker
is informed about $\mathcal{P}'_E$, we have that $U_{\mathcal{P}_E}(f_E) = U_{\mathcal{P}'_E}(f_E)$. It follows that we may
also choose $u_{\mathcal{P}_E}(f_{E''} g_{E'''} E'' \cup E''') = u_{\mathcal{P}_E}(f_{E''}, E'') + u_{\mathcal{P}_E}(g_{E''}, E''')$ for arbitrary
disjoint $E''$, $E'''$. Choosing $U_{\mathcal{P}_E}(\gamma_E)$ and $U_{\mathcal{P}_E}(\gamma_E)$ therefore uniquely fixes the
scale and utility representation of all refinements of $\mathcal{P}_E$. Moreover, since all
other partitions $\mathcal{P}'_E$ share a common refinement with $\mathcal{P}_E$, this indeed fixes the
scale of all admissible partitions of $E$.

We now argue that for arbitrary partitions $\mathcal{P}_E$ and $\mathcal{P}'_E$, $U_{\mathcal{P}_E}(f_E) \geq U_{\mathcal{P}'_E}(g_E)$
if and only if $f_E \succeq_E g_E$. Since for any $F \subset E$ and all $f_F (g_F)$, there exists an
outcome $\delta (e)$ such that $f_F \sim_E \delta_F (g_F \sim_E e_F)$, $f_F \succeq_E g_F$ iff $\delta_F \succeq_E e_F$ we have
consistency between $U_{\mathcal{P}_E}$ and $U_{\mathcal{P}'_E}$ if there are at least two common partition
elements $E', E''$ via the representation $U_{\{E', E'' \in \mathcal{P}_E \}}(f_E)$. By continuity,
for sufficiently small $E'$, $E''$, we can refine $\mathcal{P}_E$ without changing the set of indifference curves much. Call this refinement $\overline{\mathcal{P}_E}$. We can therefore ensure
that any two partitions are consistent with another since we can choose $E'$ and
$E''$ such that the partitions $\mathcal{P}_E, \overline{\mathcal{P}_E}, \mathcal{P}'_E, \overline{\mathcal{P}'_E}$ are all consistent with another on an
arbitrarily large utility interval.
We can therefore define the following utility representation on all acts with at least three distinct outcomes:

\[ U_E(f_E) = \sum_{\alpha \in X} v_E(\alpha, f_E^{-1}(\alpha)) \quad (35) \]

\[ v_E(\alpha, E') = u_{\mathcal{P}_E}(\alpha_{E'}, E') \text{ for some } \mathcal{P}_E \ni E' \quad (36) \]

To extend this representation to all acts, we consider disjoint monotone sequences \( E^k \to \emptyset \) and \( F^k \to \emptyset \). By continuity, the utility of the outcome \( \alpha \) must be the limit of \( U_S(\gamma E, \delta F, \alpha) \). The utility of \( \alpha F \beta \) must be the limit of \( U_S(\gamma E, \alpha F \beta) \) if \( F \) is disjoint from \( E \).

**Lemma 13** (Monotone Additive Representation). For an arbitrary partition \( \mathcal{P}_E \), there exists a function \( v_{\mathcal{P}_E} \) such that:

\[ U_E(f_E) = \sum_{E' \in \mathcal{P}_E} v_{\mathcal{P}_E}(U_E(f_{E'}), E') \quad (37) \]

**Proof.** Note that under the normalizations employed in the previous lemma, \( U_E(f_E) = U_{\mathcal{P}_E}(f_E) \) if \( \mathcal{P}_E \) is coarser than the events the individual is informed about, \( f_E^{-1}(X) \). The desired result then follows from the existence of an additive representation \( U_{\mathcal{P}_E} \) and that each \( u_{\mathcal{P}_E} \) must be an increasing function of \( U_E \).

**Lemma 14** (Affinity). For all \( E' \subset E \in \mathcal{E} \) \( v_{\mathcal{P}_E}(x, E') = A_{E'\mid E}x + B_{E'\mid E} \)

**Proof.** Note that since we have normalized \( u_{\mathcal{P}_E} \) and \( u_{\mathcal{P}_E}' \), such that they are identical on identical sub-acts, we know that \( v_{\mathcal{P}_E} \) does not depend on the choice of partition but only on \( E \). By the uniqueness of additive representations, we then have that \( v_{\mathcal{P}_E} \) must be affine in its first argument.

**Lemma 15** (Multiplicativity). \( A_{E'\mid E}A_{E'\mid E} = A_{E'\mid E} \).

**Proof.** Consider the utility of the following act.

\[ U_E(\alpha_{E'} \beta_{E'-E''} \gamma_{E-E''}) = A_{E'\mid E} U_{E''}(\alpha_{E''}) + A_{E'-E''\mid E} U_{E'-E''}(\beta_{E'-E''}) + A_{E-E''\mid E} U_{E-E''}(\gamma_{E-E''}) + B_{E'\mid E} + B_{E'-E''\mid E} \quad (38) \]

\[ = A_{E'\mid E} \left( A_{E''\mid E} U_{E''}(\alpha_{E''}) + A_{E'-E''\mid E} U_{E'-E''}(\beta_{E'-E''}) + B_{E''\mid E} + B_{E'-E''\mid E} \right) \]

\[ + A_{E-E''\mid E} U_{E-E''}(\gamma_{E-E''}) + B_{E'\mid E} + B_{E'-E''\mid E} \quad (39) \]
Thus, \( \alpha \) is the unique \( \gamma \), for some \( 1. \) For some \( \beta \), \( \gamma \) disjoint from \( \beta, \gamma \):

\[ A_{E|S} + A_{F|S} = u_{S-E-F}(f_{S-E-F}) - u_{S-E-F}(g_{S-E-F}) \]

\[ \left( A_{E|S} + A_{F|S} \right) \left( U_E(\gamma) - U_E(\beta) \right) = u_{S-E-F}(f_{S-E-F}) - u_{S-E-F}(g_{S-E-F}) \]

\[ \Rightarrow U(\gamma \gamma_{EF}) = U(\beta \beta'_{EF}) \]

\[ \Rightarrow A_{EUF|S} (U_{EUF}(\gamma_{EUF}) - U_{EUF}(\beta_{EUF})) = u_{S-E-F}(f_{S-E-F}) - u_{S-E-F}(g_{S-E-F}) \]

\[ \Rightarrow A_{EUF|S} = u_{S-E-F}(f_{S-E-F}) - u_{S-E-F}(g_{S-E-F}) \]

Thus, \( A_{EUF|S} = A_{E|S} + A_{F|S} \). It is straightforward to show that \( A_{E|S} = 0 \) and without loss of generality \( A_{S|S} = 1 \). It follows that \( A_{E|S} = \mu(E|S) \) is the unique probability representation of \( \succeq^* \).

We now show that under our normalization, \( U_E(\alpha_E) = U_F(\alpha_F) \). For some
acts $f, g$, we have:

$$\left( A_{E|S} + A_{F|S} \right) \left( U_S(\gamma_E) - U_E(\alpha_E) \right) = u_{S-E-F}(f_{S-E-F}) - u_{S-E-F}(g_{S-E-F})$$  \hspace{1cm} (50)

$$\Leftrightarrow U(\gamma_E \gamma_f) = U(\alpha_E \alpha'_F g)$$  \hspace{1cm} (51)

$$\Leftrightarrow U(\gamma_{EUF}) = U(\alpha_{EUF} g)$$  \hspace{1cm} (52)

$$\Leftrightarrow A_{E|S} (U_{EUF}(\gamma_{EUF}) - U_{EUF}(\alpha_{EUF})) = u_{S-E-F}(f_{S-E-F}) - u_{S-E-F}(g_{S-E-F})$$  \hspace{1cm} (53)

Thus,

$$A_{E|S} (U_{EUF}(\alpha_{EUF}) - U_E(\alpha_E)) = -A_{F|S} (U_{EUF}(\alpha_{EUF}) - U_F(\alpha_F))$$  \hspace{1cm} (54)

Since the LHS and RHS must be of the same sign and all events $E \cup F$ can be partitioned into two equally likely events $E$ and $F$, it follows for all events we must have that $U_E(\alpha_E) = U_F(\alpha_F) = U_{EUF}(\alpha_{EUF}) = U_S(\alpha) = U(\alpha)$. We thus obtain the representation:

$$U_E(f_E) = \sum_{\alpha} \mu(f_E^{-1}(\alpha))U(\alpha) + B_{f^{-1}(\alpha)|S}$$  \hspace{1cm} (55)

Defining $h(E) = B_{E|S}$ yields the desired representation. Lastly, we show that $h$ has special uniqueness properties.

**Lemma 16.** The function $h$ in the subjective knowledge utility representation is unique up to an additions of a finite measure, i.e., if $h' = h + m$ where $m$ is a finite measure on the event algebra, then $h'$ represents the same knowledge preferences as $h$. 

Proof of Proposition 3

Proof. By the uniqueness of probabilities and utility representations over lotteries, we must have that $\mu_1 = \mu_2$ and $u_1$ is an affine transformation of $u_2$. Moreover, $D$ is invariant under affine transformations of the utilities. Without loss of generality, we therefore assume that the transformation is the identity and thus $u_1 = u_2$. 

C Proof of Proposition 3

Proof. By the uniqueness of probabilities and utility representations over lotteries, we must have that $\mu^1 = \mu^2$ and $u^1$ is an affine transformation of $u^2$. Moreover, $D$ is invariant under affine transformations of the utilities. Without loss of generality, we therefore assume that the transformation is the identity and thus $u^1 = u^2$. 

41
We show that $3 \Rightarrow 2$. If $D$ is subadditive, then

$$D(E) + D(F) \geq D(E \cup F)$$

and thus if $\beta_E \beta_F \alpha \succsim^2 \gamma_{E \cup F} \alpha$, then $\beta_E \beta_F \alpha \succsim^1 \gamma_{E \cup F} \alpha$.

2 $\Rightarrow$ 1 follows straightforward from transitivity and identical preferences over outcomes.

1 $\Rightarrow$ 3 follows from:

$$D(E) + D(F) > D(E \cup F)$$

where the equivalence between the first and second line follows from the definition of the knowledge equivalent, and the remaining steps follow from the assumption that $u^1 = u^2$.

\[ \Box \]

### D Proof of Proposition 4

**Proof.** It is straightforward to show that an entropic knowledge utility representation has an elasticity of curiosity of zero and that a constant elasticity of curiosity representation has a constant elasticity of curiosity. We prove the reverse implication. If $el(p, q) = c$, then

$$el(p, q) = \frac{p^2 \left( \frac{h(pq)q + h'(p(1-q))(1-q) - h'(p)}{p} - \frac{h(pq) + h(p(1-q)) - h(p)}{p^2} \right)}{h(pq) + h(p(1-q) - h(p))} = c$$

(76)
and after reordering terms
\[ p \left( h'(pq)q + h'(p(1-q))(1-q) - h'(p) \right) = (1 + c) \left( h(pq) + h(p(1-q)) - h(p) \right) \]  
(77)

Defining \( k(p) = p \cdot h'(p) - (1 + c)h(p) \), we obtain:

\[ k(pq) + k(p(1-q)) = k(p) \]  
(78)

Substituting \( pq = x \) and \( p(1-q) = y \), we have Cauchy’s functional equation:

\[ k(x) + k(y) = k(x + y) \]  
(79)

with the solution \( k(x) = A \cdot x \). Thus, \( A \cdot x = x \cdot h'(x) - (1 + c)h(x) \) which is a linear differential equation. The integration factor of this differential equation is \( x^{-(1+c)} \). The solution is then:

\[ h(x) = \begin{cases} A \cdot x + Cx^{1+c}, & c \neq 0 \\ A \cdot x \ln x + Cx & c = 0. \end{cases} \]  
(80)

Since we may remove terms that are constant or linear in probability without changing the functional form of the representation, we obtain the desired representations. \( \square \)

## E Proof of Theorem 2

### E.1 Sufficiency Proof

**Proof.** Since we do not assume outcome solvability, we cannot use Theorem 1 to prove this result. However, the principle of indifference of information allows us to proceed in a more standard manner than in the proof of Theorem 1. We divide the proof into the following steps. First, we show that \( \succsim \) under our axioms induces a qualitative probability \( \succsim^* \). Since our event space has no atoms, we obtain a quantitative probability. Next we show that the probability distribution over outcomes determines the preference. Using our separability and monotonicity conditions we obtain a utility representation that is additively separable in outcomes. Lastly, we use learning independence to separate the utility function into an expected utility and the value of information.
**Definition 25** (Qualitative Probability). $\succeq^*$ is a qualitative probability if it is a complete and transitive relation on $\mathcal{E}$ and for any events $E, F, G \in \mathcal{E}$ such that $(E \cup F) \cap G = \emptyset$,

\begin{align*}
S & \succeq^* \emptyset \\
E \cup G & \succeq^* E \\
E \succeq^* F & \Leftrightarrow E \cup G \succeq^* F \cup G.
\end{align*}

**Conjecture 1** (Existence of a Qualitative Probability). $\succeq^*$ is a qualitative probability.

We first prove several lemmas about the properties of $\succeq^*$.

**Lemma 17** (Completeness). For all $E, F \in \mathcal{E}$, $E \succeq^* F$, $F \succeq^* E$, or $E \sim^* F$.

*Proof.* Completeness is guaranteed since for any two events $E, F \in \mathcal{E}$, we can write $E = E - F \cup (E \cap F)$ and $F = F - E \cup (E \cap F)$. Completeness of $\succeq$ together with (7) then guarantee that either $E \succeq^* F$, or $F \succeq^*$, or both. \qed

**Definition 26** (Event Solvability). $\succeq$ fulfills Event Solvability if whenever $E \succeq^* F$ and $E$ and $F$ are disjoint, then there exists an event $E' \subseteq E$ such that $E' \sim^* F$.

In other words, if $\gamma_{E\beta F} \succeq_{E \cup F} \beta_{E \gamma F}$ for $\gamma \succ \beta$, then there exists a subevent $E'$ of the event $E$ such that $\gamma_{E'\beta F} \sim \beta_{E' \gamma F}$.

**Lemma 18.** If there are no atoms and continuity holds, then Event Solvability holds.

*Proof.* The proof is straightforward. We iteratively split the larger event and add or remove subevents to obtain a converging sequence of events. This sequence can be split into sequences of events that are too large and those that are too small. Continuity then guarantees that the associated acts characterizing the likelihood relation converge.

With no-atom $\mathcal{E}$, the proof is trivial if either $E \sim^* F$ or $F \sim^* \emptyset$. Hence, assume $E \succ^* F \succ^* \emptyset$.

Because $\mathcal{E}$ has no atoms, for every event $A \succ^* \emptyset$, $\exists A' \subset A$ such that $\emptyset \prec^* A' \prec^* A$. Furthermore, we can find a partition $P$ of $E \{E_1, E_2, ..., E_n\}$ such that for all $j$, $F - E_j \succeq^* E_j$ and that for all $j$, $\emptyset \prec^* E_j \prec^* F$ because $F \succ \emptyset$. The former works because of Theorem 4 of Villegas (1964). Suppose the latter does not work, then there is an atom $E'$ in $E$ such that $E' \succeq^* F \succ^* \emptyset$, a contradiction. By construction, we can find $i < n$ such that $L_1 \equiv \bigcup_{1 \leq j \leq i} E_j \succeq^* F$ and that $U_1 \equiv L_1 \cup E_{i+1} \succeq^* F$. 

44
Similarly, we can again find a non-null partition \( P^{(2)} \) of \( E_{i+1} \), \( \{ E_1^{(2)}, E_2^{(2)}, \ldots, E^{(2)}_{n_2} \} \), such that for every \( j \) \( (1) \not\prec E_j^{(2)} \prec E_{i+1} \) \( (2) \) \( L_1 \cup E_j^{(2)} \prec F \) \( (3) \) \( E_{i+1} - E_j^{(2)} \succeq E_j^{(2)} \). It follows that there exists again \( i^{(2)} \) such that \( L_2 \equiv L_1 \cup \bigcup_{1 \leq j \leq i^{(2)}} E_j^{(2)} \succeq F \) and that \( U_2 \equiv L_2 \cup E_{i^{(2)}+1}^{(2)} \succeq F \).

We repeat by finding such a partition \( P_{k+1} \) for \( E_{(k)+1}^{(k)} \) and obtain two sequence of events \( \{ L_k \}, \{ U_k \} \). By construction, \( E' \equiv \bigcup_{k=1}^{\infty} L_k = \bigcap_{k=1}^{\infty} U_k \) and \( L_k \succeq F \succeq U_k \). The former implies \( E' \in \mathcal{E} \) (since \( \mathcal{E} \) is a \( \sigma \)-algebra) whereas the latter implies \( E' \sim F \) by continuity.

Lemma 19 (Nontriviality of Likelihood). \( S \not\succ \emptyset \)

Proof. Nontriviality of \( \succeq \) ensures that \( \gamma_S \beta_\emptyset \alpha = \gamma \succ \beta = \beta_S \gamma_\emptyset \alpha \) and therefore \( S \not\succ \emptyset \).

Lemma 20 (Subset Consistency). \( E \subseteq F \Rightarrow F \succeq E \)

Proof. If \( E \subseteq F \), then \( E \cup G = F \) where \( E \cap G = \emptyset \). We need to show that if \( \gamma \succ \beta \), then,

\[
\gamma_{E-F} \beta_{E-F} \alpha \succeq \beta_{E-F} \gamma_{E-F} \alpha
\]

(84)

\[
\iff \gamma_G \alpha \succeq \beta_G \alpha
\]

(85)

which is guaranteed by information-neutral monotonicity and \( \gamma \succ \beta \).

Lemma 21 (Expansion Consistency). \( E \succeq F \leftrightarrow E \cup G \succeq F \cup G \) whenever \( (E \cup F) \cap G = \emptyset \).

The result follows since the definition of \( E \succeq F \) is the same as the definition of \( E \cup G \succeq F \cup G \);

\[
E \succeq F
\]

(86)

\[
\iff \gamma_{E-F} \beta_{E-F} \alpha \succeq \beta_{E-F} \gamma_{E-F} \alpha
\]

(87)

\[
\iff \gamma_{(E \cup G)-(F \cup G)} \beta_{(F \cup G)-(E \cup G)} \alpha \succeq \beta_{(E \cup G)-(F \cup G)} \gamma_{(F \cup G)-(E \cup G)} \alpha
\]

(88)

\[
\iff E \cup G \succeq F \cup G.
\]

(89)

Lemma 22 (Disjoint Measurability). If \( G \sim H \) and \( E, F, G, \) and \( H \) are mutually disjoint, then \( E \succeq F \leftrightarrow E \cup G \succeq F \cup H \).
Proof.

\[
\gamma E \beta F \gamma F \gamma G \delta \preceq \beta E \gamma F \gamma G \delta.
\]

(90)

\[
\Leftrightarrow \gamma E \beta F \gamma G \delta \succeq \beta E \gamma F \gamma G \delta.
\]

(91)

\[
\Leftrightarrow \gamma E \beta F \gamma G \delta \succeq \beta E \gamma F \gamma G \delta.
\]

(92)

\[
\Leftrightarrow \gamma E \beta F \gamma G \delta \succeq \beta E \gamma F \gamma G \delta.
\]

(93)

\[
\Leftrightarrow \gamma E \beta F \gamma G \delta \succeq \beta E \gamma F \gamma G \delta.
\]

(94)

\[
\Leftrightarrow \gamma E \beta F \gamma G \delta \succeq \beta E \gamma F \gamma G \delta.
\]

(95)

While the first, second, forth and fifth equivalence comes from information-neutral sure-thing principle, the third holds thanks to the principle of indifference of information.

\[\square\]

Lemma 23 (Transitivity). \(E \succeq^* F \) and \(F \succeq^* G \) implies \(E \succeq^* G \).

We prove this in several steps.

Lemma 24 (Disjoint Transitivity). If \(E, F, G \) are mutually disjoint events, then \(E \succeq^* F \) and \(F \succeq^* G \) implies \(E \succeq^* G \).

Proof. Let \(\gamma > \beta > \alpha \) be given. Since \(E \succeq^* F \), we have \(\gamma E \beta F \gamma G \delta \succeq \beta E \gamma F \gamma G \delta\). Since \(F \succeq^* G \), we have \(\beta E \gamma F \gamma G \delta \succeq \beta E \gamma F \gamma G \delta\). By \(E \succeq^* F \) again, \(\alpha E \beta F \gamma G \delta \succeq \alpha E \beta F \gamma G \delta\). Hence we conclude that, \(\gamma E \beta F \gamma G \delta \succeq \alpha E \beta F \gamma G \delta\), that is \(E \succeq^* G \). \[\square\]

Lemma 25 (Subset Transitivity).

1. Suppose \(F \subseteq E \) and \(F \succeq^* G \), then \(E \succeq^* G \).

2. Suppose \(G \subseteq F \) and \(E \succeq^* F \), then \(E \succeq^* G \).

Proof.

1. Solve for \(G \) on \(F \), obtain \(E \supseteq F' \sim^* G \). If \(G > E \), then \(\beta E \gamma G \delta \sim \beta E \gamma G \delta\) which holds if and only if (By learning independence) \(\beta'_{E-F} \beta F \gamma G \delta \sim \gamma'_{E-F} \gamma F \beta G \delta\). Since \(F' \sim G \), we have that \(\beta'_{E-F} \beta F \gamma G \delta \sim \beta'_{E-F} \gamma F \beta G \delta\) violating monotonicity.

2. By Event Solvability and \(E \succeq^* F \), we can solve for \(F \) on \(E \), getting us \(E \supseteq E' \sim^* F \). Suppose for contradiction that \(G \succ^* E \); that is, \(\gamma G \beta F \delta \succ \gamma E \beta F \delta\). This is identical to the following expression: \(\gamma E \gamma F \beta E \delta \succ \gamma'_{E-F} \beta G \delta \). Since \(E' \sim^* F \) and they are disjoint, we can switch
\[ \gamma \text{ and } \beta \text{ on } E' \cup F. \text{ This gives us } \emptyset = (G - F) \succ^* E - E' \succ^* \emptyset, \text{ a contradiction.} \]

\[ \square \]

**Lemma 26** (Limited Transitivity 1). If \((E \cup F) \cap G = \emptyset\), then \(E \succ^* F\) and \(F \succ^* G\) imply \(E \succ^* G\).

*Proof.* If \(E \cap F \succ G\), then we have the result by Subset Transitivity 1. Otherwise we can solve \(E \cap F\) into \(G\), calling it \(G'\). By Disjoint Measurability \(F - E \succ G - G'\). Now we solve \(G - G'\) into \(F - E\), calling it \(F' \sim G - G'\) and via Subset Transitivity 1 have that \(F - E \succ G - G'\). By Disjoint Transitivity \(E - F \succ^* G - G'\). By Disjoint Measurability \(E \succ G\).

\[ \square \]

**Lemma 27** (Limited Transitivity 2). If \(E \cap G = \emptyset\), then \(E \succ^* F\) and \(F \succ^* G\) imply \(E \succ^* G\).

*Proof.* Since \(E \succ^* F\), \(E - F \succ^* F - E \supseteq F \cap G\). By Subset Transitivity \(E - F \succ^* F \cap G\). We can therefore solve for \(E' \sim^* F \cap G\) such that \(E' \subseteq E - F\). By Disjoint Measurability, \(E - F \succ F - E\) if and only if \((E - F) - E' \succ F - G\). By Limited Transitivity 1, \((E - F) - E' \succ^* G - F\) follows from \(F - G \succ G - F\) and \((E - F) - E' \succ F - G\). Moreover, by Disjoint Measurability and Subset Transitivity, \(E \succ G\) if \((E - F) - E' \succ^* G - F\).

\[ \square \]

**Lemma 28** (Limited Transitivity 3). If \((F - E) - G = \emptyset\) and \(E \cap F \cap G = \emptyset\), then \(E \succ^* F\) and \(F \succ^* G\) imply \(E \succ^* G\).

*Proof.* Solve \(G - F\) into \(F - G\), naming the set \(F'\). Also, \(E \succ^* F\) is equivalent to \(E - F \succ^* F - E\). By Disjoint Measurability, we can add \(F'\) to the LHS and \(G - F\) to the RHS, yielding \((E - F) \cup F' \succ G\). By Subset Transitivity, \(E \succ G\).

\[ \square \]

**Lemma 29** (Limited Transitivity 4). If \(E \cap F \cap G = \emptyset\), then \(E \succ^* F\) and \(F \succ^* G\) imply \(E \succ^* G\).

*Proof.* If \(G \cap E \succ (F - G) - E\), then we can solve for a \(G'\) in \(G \cap E\) such that \(G' \sim (F - G) - E\). Then \(E \succ G\) if \(E - G' \succ (F \cap E) \cup (F \cap G)\). Moreover, \(F \succ G\) iff \((F \cap E) \cup (F \cap G) \preceq G - G'\). Applying Limited Transitivity 3, we have that \(E - G' \succ G - G'\) and by Expansion consistency \(E \succ G\). If \((F - G) - E \succ^* E \cap G\), then we can solve for an \(F'\) in \((F - G) - E\). Then, \(E \succ F\) if and only if \(E - G \succ F - F'\) and \(F \succ G\) if and only if \(F - F' \succ G - E\). By Limited Transitivity 2 it follows that \(E - G \succ^* G - E\).

\[ \square \]
Lemma 30 (Transitivity). If $\succsim^*$ is transitive on events with $E \cap F \cap G = \emptyset$, then it is transitive.

Proof. By Expansion consistency, for any $X, Y \in \{E, F, G\}$, we have that $X \succsim Y$ if and only if $X - (E \cap F \cap G) \succsim Y - (E \cap F \cap G)$. Thus, proving Limited Transitivity 4 is equivalent to proving Transitivity. \qed

Lemma 31. $\succsim^*$ has a quantitative probability representation.

Proof. This follows from Villegas (1964) since we have a qualitative probability without atoms. \qed

Lemma 32 (Event Swapping). Suppose $\{E_1, \ldots, E_k\}$ and $\{F_1, \ldots, F_k\}$ are partitions of $S$ such that for all $i$, $E_i \sim^* F_i$. Then for all lists of outcomes $\alpha_1, \ldots, \alpha_k$,

$$(E_1 : \alpha_1, \ldots, E_k : \alpha_k) \sim (F_1 : \alpha_1, \ldots, F_k : \alpha_k)$$ \hspace{1cm} (96)

Proof. If $G \subseteq E_1$ and $H \subseteq E_2$ and $G \sim^* H$, then by the principle of indifference of information we have that

$$(E_1 : \alpha_1, E_2 : \alpha_2, \overline{E_1 \cup E_2} : \gamma) \sim ((E_1 - G) \cup H : \alpha_1, (E_2 - H) \cup G : \alpha_2, \overline{E_1 \cup E_2} : \gamma).$$ \hspace{1cm} (97)

By the information-neutral sure-thing principle

$$(E_1 : \alpha_1, E_2 : \alpha_2, \ldots, E_k : \alpha_k) \sim ((E_1 - G) \cup H : \alpha_1, (E_2 - H) \cup G : \alpha_2, \ldots, E_k : \alpha_k).$$ \hspace{1cm} (98)

From this result, the lemma is almost obvious as we can change parts of events on which outcomes occur as long as the probabilities are maintained. We iteratively apply this result to construct the indifference in (96). For this, choose $H_i = F_1 \cap E_i$ and $G_i$ as an equiprobable event contained in $E_1 - \bigcup_{l=2}^{i-1} H_l$ which...
exists by Event Solvability. By (98) we have that:

\((E_1 : \alpha_1, \ldots, E_k : \alpha_k)\)

\(\sim ((E_1 - G_2) \cup H_2 : \alpha_1, (E_2 - H_2) \cup G_2, E_3 : \alpha_3, \ldots, E_k : \alpha_k)\)

\(\sim \ldots\)

\(\sim \left( \left( E_1 - \bigcup_{i=2}^{k} G_i \right) \cup \bigcup_{i=2}^{k} H_i, (E_2 - H_2) \cup G_2, \ldots, (E_k - H_k) \cup G_k \right)\)

\(= (F_1, (E_2 - H_2) \cup G_2, \ldots, (E_k - H_k) \cup G_k).\) \hspace{1cm} (99)

Note, that our choice of \(E_1\) and \(F_1\) was arbitrary, we can therefore repeat the above argument for all events. Since \(F_i \cap F_j = \emptyset\) for all \(i \neq j\), all previously changed events will not be changed by applying (99) to another event. Applying (99) to all events yields the desired indifference (96).

\[\square\]

**Lemma 33** (Epimorphism to Mixture Space). Let \(\Delta X\) be the space of finite support probability measures over outcomes. Let \(\mu\) be the quantitative probability representation of \(\succ^*\). There exists a function \(\phi : A \to \Delta X\) and a unique relation \(\succeq^+\) such that for all \(\alpha \in X\) and all \(a, b \in A\):

\[a \succeq b \iff \phi(a) \succ^+ \phi(b)\] \hspace{1cm} (100)

\[\mu(a^{-1}(\alpha)) = (\phi(a))(\alpha)\] \hspace{1cm} (101)

**Proof.** By Event Swapping, any two acts are indifferent that induce the same probability measure on the outcomes. If \(\phi\) maps every act to this probability measure, then (100) trivially holds. \[\square\]

**Lemma 34** (Mixture Space Completeness). \(\succeq^+\) is complete and transitive.

**Proof.** Completeness and transitivity follow directly from completeness and transitivity of \(\succeq\). \[\square\]

We employ the monotone subsequence theorem adapted to simple probability measures.

**Lemma 35** (Monotone Subsequence Theorem). If \(\mu^k \to \mu\) of simple probability measures on \(X\), then for any outcome \(\alpha \in \text{supp}(\mu)\), the sequence \(\{\mu^k\}\) has a convergent subsequence \(\{\bar{\mu}^k\} \to \mu\) such that \(\bar{\mu}^k(\alpha) \geq \bar{\mu}^l(\alpha)\) for all \(l \geq k\) or \(\bar{\mu}^k(\alpha) \leq \bar{\mu}^l(\alpha)\) for all \(l \geq k\).

49
Proof. Let $\alpha$ be given. We first prove that every sequence $\{\mu_k\}$ has a monotonic subsequence. We call the $m$-th term of the sequence $\mu^m$ a peak if for all $n \geq m$ we have that $\mu^n(\alpha) \geq \mu^m(\alpha)$.

— Case 1: Suppose the sequence $\{\mu_k\}$ has infinitely many peaks. We use these peaks to form a subsequence and denote it $\{\bar{\mu}_k\}$. Since they are peaks, we have $\bar{\mu}_1(\alpha) \geq \bar{\mu}_2(\alpha) \geq \bar{\mu}_3(\alpha) \geq \cdots$ and thus we have a monotonic decreasing sequence.

— Case 2: Suppose the sequence $\{\mu_k\}$ has only finitely many peaks and denote it $\mu_{k1}, \mu_{k2}, \ldots, \mu_{kn}$. Let $\bar{\mu}_1 = \mu_{kn+1}$. Since $\bar{\mu}_1$ is not a peak we can find $\bar{\mu}_2$ such that $\bar{\mu}_1(\alpha) \leq \bar{\mu}_2(\alpha)$. Continue this process, we have a monotonic increasing sequence.

Since every subsequence of a convergent sequence converges to the same limit, we have the desired result. 

Lemma 36 (Mixture Space Continuity). $\preceq^+$ is continuous (closed weakly upper and lower sets)

Proof. Consider sequences $\mu^k \rightarrow \mu$ and $\nu^k \rightarrow \nu$ of simple probability measures on $X$. By Lemma 35, for any outcome $\alpha \in \text{supp}(\mu)$ the sequence $\{\mu_k\}$ has a convergent subsequence $\{\bar{\mu}_k\} \rightarrow \mu$ such that $\bar{\mu}_k(\alpha) \geq \bar{\mu}_l(\alpha)$ for all $l \geq k$ or $\bar{\mu}_k(\alpha) \leq \bar{\mu}_l(\alpha)$ for all $l \geq k$. We can therefore convert $\mu^k$ into a sequence that monotonically changes the probabilities of the outcomes in the support of $\mu$. Next, we convert the sequence of measures into a sequence of acts. Let $f$ be an arbitrary element of $\phi^{-1}(\mu)$. We then choose a sequence of $f^k \in \phi^{-1}(\mu^k)$ such that it converges monotonically to $f$. Proceeding the same way for $\nu^k \rightarrow \nu$, we obtain a monotonically convergent sequence $g_k \rightarrow g$. We then have that if for all $k$, $\mu^k \succeq^+ \nu^k$, then also $f^k \succeq g^k$. It follows by continuity that $f \succeq g$ and therefore $\mu \succeq^+ \nu$. 

Lemma 37 (Additive Representation). $\preceq^+$ can be represented by

$$U(\mu) = \sum_{\alpha \in \text{supp}(\mu)} v(\mu(\alpha), U(\alpha)).$$

Proof. By the information-neutral sure-thing principle, for a state space partition $\{E, F, G\}$ we have that preferences are separable on conditional acts $f_E, f_F$, and $f_G$ for disjoint outcome domains. Since these conditional acts straightforwardly map into restrictions of $\mu$ to the subsets of outcomes $\mu|_{\text{supp}(f_E)}, \mu|_{\text{supp}(f_F)}$, and...
by a standard argument, it follows that there exists an additive representation of the form

$$U(\mu) = u(\mu_{|\text{supp}(f_E)}) + v(\mu_{|\text{supp}(f_F)}) + w(\mu_{|\text{supp}(f_G)})$$  \hspace{1cm} (103)

Choosing any other partition of outcomes on $f_E$ and $f_F$ yields another representation of the above form that must be a monotone transformation of $U$ since only the probabilities of outcomes matter, not the event on which they arise. It is straightforward to show that by the uniqueness of additive representations, the two representations must be affine transformations of another. It follows that $u$ (and $v$ and $w$ in a similar manner) is additively separable in the outcomes. We obtain a representation:

$$U(\mu) = \sum_{\alpha \in \text{supp}(\mu)} V(\mu(\alpha), \alpha).$$  \hspace{1cm} (104)

By information-neutral monotonicity, $V$ must be increasing in $U(\alpha)$. The desired representation follows.

We now finish the proof of the theorem. By learning independence, it follows after cancelling terms that

$$v(\mu(\alpha), U(\alpha)) + v(\mu(\alpha'), U(\alpha)) - v(\mu(\alpha) + \mu(\alpha'), U(\alpha))$$

$$= v(\mu(\alpha), U(\gamma)) + v(\mu(\alpha'), U(\gamma)) - v(\mu(\alpha) + \mu(\alpha'), U(\gamma))$$  \hspace{1cm} (105)

This is a Pexider functional equation with the solution:

$$v(\mu(\alpha), U(\alpha)) + v(\mu(\alpha), U(\gamma)) = A(U(\alpha), U(\gamma))\mu(\alpha) + B(U(\alpha), U(\gamma)).$$  \hspace{1cm} (107)

Since $\gamma$ was arbitrarily chosen, it follows that $v(\mu(\alpha), U(\alpha)) = A(U(\alpha))\mu(\alpha) + B(U(\alpha)) + C(\mu(\alpha))$. Noting that continuity requires that $v(0, U(\alpha)) = 0$, we must have that $B(U(\alpha)) = 0$. The desired representation follows.

\paragraph{E.2 Necessity Proof}  

\textbf{Lemma 38} (Necessity). \textit{In a probabilistic knowledge utility representation, the Axioms 2, 3, 4, 6, 7, 8 and weak order hold.}
Proof. It is clear that weak order holds. Axiom 2 holds because

\[
U(f_{E}k) \geq U(g_{E}k) \quad (108)
\]

\[
\Leftrightarrow \sum_{a \in f(E)} u(a)\mu(f^{-1}(a)) - h(\mu(f^{-1}(a))) + \sum_{a \in k(E)} u(a)\mu(k^{-1}(a)) - h(\mu(k^{-1}(a))) \quad (109)
\]

\[
\geq \sum_{a \in g(E)} u(a)\mu(f^{-1}(a)) - h(\mu(f^{-1}(a))) + \sum_{a \in k(E)} u(a)\mu(k^{-1}(a)) - h(\mu(k^{-1}(a))) \quad (110)
\]

\[
\Leftrightarrow \sum_{a \in f(E)} u(a)\mu(f^{-1}(a)) - h(\mu(f^{-1}(a))) + \sum_{a \in l(E)} u(a)\mu(k^{-1}(a)) - h(\mu(k^{-1}(a))) \quad (111)
\]

\[
\geq \sum_{a \in g(E)} u(a)\mu(f^{-1}(a)) - h(\mu(f^{-1}(a))) + \sum_{a \in l(E)} u(a)\mu(k^{-1}(a)) - h(\mu(k^{-1}(a))) \quad (112)
\]

\[
\Leftrightarrow U(f_{E}l) \geq U(g_{E}l) \quad (113)
\]

Axiom 3 holds because

\[
U(\gamma_{E}k) \geq U(\beta_{E}k) \quad (114)
\]

\[
\Leftrightarrow u(\gamma)\mu(E) - h(\mu(E)) + \sum_{a \in k(E)} u(a)\mu(k^{-1}(a)) - \sum_{a \in k(E)} h(\mu(k^{-1}(a))) \quad (115)
\]

\[
\geq u(\beta)\mu(E) - h(\mu(E)) + \sum_{a \in k(E)} u(a)\mu(k^{-1}(a)) - \sum_{a \in k(E)} h(\mu(k^{-1}(a))) \quad (116)
\]

\[
\Leftrightarrow u(\gamma) \geq u(\beta) \quad (117)
\]

\[
\Leftrightarrow \gamma \succeq \beta \quad (118)
\]

Axiom 4 holds because

\[
U(\gamma_{E}\beta_{E}a) \geq U(\beta_{E}\gamma_{E}a) \quad (119)
\]

\[
\Leftrightarrow \mu(E)u(\gamma) + \mu(F)u(\beta) \geq \mu(E)u(\beta) + \mu(F)u(\gamma) \quad (120)
\]

\[
\Leftrightarrow \mu(E) \geq \mu(F) \quad (121)
\]

\[
\Leftrightarrow \mu(E)u(\gamma') + \mu(F)u(\beta') \geq \mu(E)u(\beta') + \mu(F)u(\gamma') \quad (122)
\]

\[
\Leftrightarrow U(\gamma'_{E}\beta'_{E}a) \geq U(\beta'_{E}\gamma'_{E}a) \quad (123)
\]

For Axiom 6, \(f^k \rightarrow f\) means that for each \(a\) there is a monotone sequence of events \(E^k \rightarrow E = f^{-1}(a)\). From this we obtain a sequence of probabilities
$\rho^k(\alpha) \rightarrow \mu(E)$. Now define the utility contribution of $\alpha$ $U^k_\alpha = u(\alpha)\rho^k(\alpha) - h(\rho^k(\alpha)) \rightarrow u(\alpha)\mu(E) - h(\mu(E))$. But if for all $\alpha$ these sequences converge, then the sum converges as well. Axiom 7 holds because

$$U(\alpha_E \alpha_F f) \geq U(\delta_E \delta_F g) \quad (124)$$

$$\Leftrightarrow \mu(E \cup F)u(\alpha) + \sum_{\beta \in f(E \cap F)} u(\beta)\mu(f^{-1}(\beta)) - \sum_{\beta \in f(E \cap F)} h(\mu(f^{-1}(\beta))) \quad (125)$$

$$\geq \mu(E \cup F)u(\delta) + \sum_{\beta \in g(E \cap F)} u(\beta)\mu(g^{-1}(\beta)) - \sum_{\beta \in g(E \cap F)} h(\mu(g^{-1}(\beta))) \quad (126)$$

$$\Leftrightarrow \mu(E)u(\alpha) + \mu(F)u(\alpha) + \sum_{\beta \in f(E \cap F)} u(\beta)\mu(f^{-1}(\beta)) - \sum_{\beta \in f(E \cap F)} h(\mu(f^{-1}(\beta))) \quad (127)$$

$$\geq \mu(E)u(\delta) + \mu(F)u(\delta) + \sum_{\beta \in g(E \cap F)} u(\beta)\mu(g^{-1}(\beta)) - \sum_{\beta \in g(E \cap F)} h(\mu(g^{-1}(\beta))) \quad (128)$$

$$\Leftrightarrow \mu(E)u(\alpha) + \mu(F)u(\alpha') + \sum_{\beta \in f(E \cap F)} u(\beta)\mu(f^{-1}(\beta)) - \sum_{\beta \in f(E \cap F)} h(\mu(f^{-1}(\beta))) \quad (129)$$

$$\geq \mu(E)u(\delta) + \mu(F)u(\delta') + \sum_{\beta \in g(E \cap F)} u(\beta)\mu(g^{-1}(\beta)) - \sum_{\beta \in g(E \cap F)} h(\mu(g^{-1}(\beta))) \quad (130)$$

$$\Leftrightarrow U(\alpha_E \alpha'_F f) \geq U(\delta_E \delta'_F g) \quad (131)$$

Axiom 8 holds because

$$U(\alpha_E \alpha_F \alpha_G \beta_H \gamma) = U(\beta_E \alpha_F \alpha_G \beta_H \gamma) \quad (132)$$

$$\Leftrightarrow \mu(E \cup G)u(\alpha) + \mu(F \cup H)u(\beta) - h(\mu(E \cup G)) - h(\mu(F \cup H)) \quad (133)$$

$$= \mu(E \cup H)u(\beta) + \mu(F \cup G)u(\alpha) - h(\mu(E \cup H)) - h(\mu(F \cup G)) \quad (134)$$

$$\Leftrightarrow \mu(E)u(\alpha) + \mu(G)u(\alpha) + \mu(F)u(\beta) + \mu(H)u(\beta) \quad (135)$$

$$= \mu(E)u(\beta) + \mu(H)u(\beta) + \mu(F)u(\alpha) + \mu(G)u(\alpha) \quad (136)$$

$$\Leftrightarrow \mu(E) = \mu(F) \quad (137)$$
Proof of Theorem 3

Proof. The result that 3 \implies 2 is trivial.

We prove 2 \implies 1:

Using outcome solvability, we can show by a simple induction proof that if stationarity I (II) holds for \( k = 2 \), then stationarity I (stationarity I with equiprobable events) holds for arbitrary \( k \in \mathbb{N} \). Naturally, stationarity I for \( k > 2 \) implies stationarity 1 for \( k = 2 \). What is left to show is that stationarity I for equiprobable events and arbitrary \( k \) implies stationarity II for \( k = 2 \).

Suppose \( a_E \beta_F e \preceq \gamma_E \delta_F e \) and \( E|E \cup F \sim^+ G|G \cup H \). Without loss of generality, assume \( E \succ^* F \). If \( \mu(E|E \cup F) \) is rational, we can find an equiprobable partition of \( E \cup F \) into \( k \) events such that \( E = \bigcup_{i=1}^{k} I_i^* \) and \( F = \bigcup_{i=1}^{k} L_i^* \). Then by Learning Independence, \( a'_1 a''_2 \ldots a''_{k} \delta'_{k+1} \delta''_{k+2} \ldots \epsilon \preceq \gamma'_1 \gamma''_2 \ldots \delta'_{k+1} \delta''_{k+2} \ldots \epsilon \). By stationarity I for equiprobable events, this is equivalent to: \( a'_1 a''_2 \ldots a''_{k} \delta'_{k+1} \delta''_{k+2} \ldots \epsilon \preceq \gamma'_1 \gamma''_2 \ldots \delta'_{k+1} \delta''_{k+2} \ldots \epsilon \) for an equiprobable partition of \( G \cup H \) fulfilling \( G = \bigcup_{i=1}^{k} I_i \) and \( H = \bigcup_{i=1}^{k} H_i \).

By Learning Independence, \( a_C \beta_H e \preceq \gamma_C \delta_E e \). The case in which \( \mu(E|E \cup F) \) is not rational follows from the continuity of the representation.

1 \implies 3: It is straightforward to derive that the stationarity axioms imply the principle of indifference of information and thus a probabilistic knowledge representation. Let \( \psi(f_F) \) map a conditional act \( f_F \) to an act \( \psi(f_F) \) such that the conditional probabilities of outcomes remain unchanged. Under stationarity I and a probabilistic knowledge utility representation,

\[
\sum_{\alpha} \frac{\mu(f^{-1}(\alpha))}{\mu(F)} u(\alpha) + h \left( \frac{\mu(f^{-1}(\alpha))}{\mu(F)} \right) \geq \sum_{\alpha} \frac{\mu(g^{-1}(\alpha))}{\mu(F)} u(\alpha) + h \left( \frac{\mu(g^{-1}(\alpha))}{\mu(F)} \right)
\]

\[
\iff \quad \psi(f_F) \preceq \psi(g_F)
\]

\[
\quad \iff f_F \preceq f_F g_F
\]

\[
\quad \iff f_F \beta \preceq g_F \beta
\]

\[
\quad \iff \sum_{\alpha} \mu(f^{-1}(\alpha)) u(\alpha) + h \left( \mu(f^{-1}(\alpha)) \right) \geq \sum_{\alpha} \mu(g^{-1}(\alpha)) u(\alpha) + h \left( \mu(g^{-1}(\alpha)) \right)
\]

Thus, \( U(f_F|\alpha) = T(\psi(f_F)) \) for a continuous, monotone transformation. It follows from the uniqueness of additive representations that \( T \) is affine. It is
straightforward to show then that:

\[ U(\alpha E \beta \gamma) = \mu(E) U(\alpha) + h(\mu(E)) + h(1 - \mu(\overline{E})) + (1 - \mu(E)) \left( \frac{\mu(F)}{1 - \mu(E)} U(\beta) + h\left( \frac{\mu(F)}{1 - \mu(E)} \right) \right) + h(\mu(E)) + h(1 - \mu(\overline{F})) \left( \frac{\mu(\overline{E} \cup F)}{1 - \mu(E)} U(\gamma) + h\left( \frac{\mu(\overline{E} \cup F)}{1 - \mu(E)} \right) \right) \]  

(143)

Canceling terms, we obtain the fundamental equation of information:

\[ h(\mu(\alpha)) + h(1 - \mu(\alpha))(1 - \mu(\alpha)) \left( h\left( \frac{\mu(\beta)}{1 - \mu(\alpha)} \right) + h\left( \frac{\mu(\gamma)}{1 - \mu(\alpha)} \right) \right) = h(\mu(\beta)) + h(1 - \mu(\beta))(1 - \mu(\beta)) \left( h\left( \frac{\mu(\alpha)}{1 - \mu(\beta)} \right) + h\left( \frac{\mu(\gamma)}{1 - \mu(\beta)} \right) \right) \]  

(145)

The continuous solution to this functional equation is \( h(\mu(\alpha)) = \mu(\alpha) \ln(\mu(\alpha)) \). The desired representation follows.

\[ \square \]

G Proof of Theorem 4

Proof. From the existence of a subjective knowledge utility representation we can find for the evaluation of all conditional acts given an arbitrary event \( E \) a representation of the form:

\[ U_E(a_E) = \sum_{\alpha} \mu_E(a_E^{-1}(\alpha)) u_E(\alpha) + h_E(a_E^{-1}(\alpha)) \]  

(146)

Since the expected utility component always has the richness of a continuum, these representations are unique up to affine transformations. By stationarity I* (with respect to partitions fulfilling \( \mu(E_1 \cup \ldots \cup E_k) = \mu(G_1 \cup \ldots \cup G_k) \)) it is straightforward that \( h_E(E) = h_E(\mu(E), \mu_E) \) for a suitably defined function \( h_E \).

\[ U_E(a_E) = \sum_{\alpha} \mu_E(a_E^{-1}(\alpha)) u_E(\alpha) + h_E(\mu(a_E^{-1}(\alpha)), \mu_E(a_E^{-1}(\alpha))) \]  

(147)
By stationarity $t^*$ with respect to partitions with $\mu(E_1 \cup \ldots \cup E_k) \neq \mu(G_1 \cup \ldots \cup G_k)$ the representations $U_E$ and $U_G$ must be affine transformations of another if $\mu_E^V = \mu_G^V$. If we normalize $u_E = u_G = u$, then this reduces to an additive transformation. Thus, whenever $a_E$ and $a_G$ are corresponding conditional acts in outcomes and knowledge partition gained, then

$$U_E(a_E) = \sum_{\alpha} \mu_E(a_E^{-1}(\alpha)) u(\alpha) + h_E(\mu(a_E^{-1}(\alpha)), \mu_{a_E^{-1}(\alpha)}^V) + t_E$$

$$= U_G(a_G) = \sum_{\alpha} \mu_G(a_G^{-1}(\alpha)) u(\alpha) + h_G(\mu(a_G^{-1}(\alpha)), \mu_{a_G^{-1}(\alpha)}^V) + t_G. \quad (148)$$

Cancelling the identical instrumental preferences,

$$U_E(a_E) = \sum_{\alpha} h_E(\mu(a_E^{-1}(\alpha)), \mu_{a_E^{-1}(\alpha)}^V) + t_E$$

$$= U_G(a_G) = \sum_{\alpha} h_G(\mu(a_G^{-1}(\alpha)), \mu_{a_G^{-1}(\alpha)}^V) + t_G. \quad (149)$$

By Lemma 16, we can then assume that $h_E = h_G = h_{\mu_E}$. By the information-adjusted sure-thing principle and the uniqueness of additive representations, the representation $U_E(a_E)$ must also be an affine transformation of $U_{E \cup F}(a_E \gamma_F)$ for all $a_E$ disjoint from $\gamma_F$ if $E \cap F = \emptyset$. Thus,

$$U_E(a_E) \quad (150)$$

$$= \sum_{\alpha} \mu_E(a_E^{-1}(\alpha)) u(\alpha) + h_{\mu_E}(\mu(a_E^{-1}(\alpha)), \mu_{a_E^{-1}(\alpha)}^V) + t_E \quad (151)$$

$$= A(E, \gamma, F) U_{E \cup F}(a_E \gamma_F) + B(E, \gamma, F) \quad (152)$$

$$= A(E, \gamma, F)(\mu_{E \cup F}(F) u(\gamma) + h_{\mu_{E \cup F}}(\mu_{E \cup F}(F), \mu_F^V) \quad (153)$$

$$+ \sum_{\alpha} \mu_{E \cup F}(E) \mu_E(a_E^{-1}(\alpha)) u(\alpha) + h_{\mu_{E \cup F}}(\mu_{E \cup F}(E) \mu_E(a_E^{-1}(\alpha)), \mu_{a_E^{-1}(\alpha)}^V) + t_{E \cup F})$$

$$+ B(E, \gamma, F). \quad (154)$$

It is not without loss of generality to assume an arbitrary function $A$. This is because we already chose the affine transformation between $U_E$ and $U_G$ and other arbitrary events $G'$ with $\mu_E^V = \mu_G^V = \mu_G^V$ and we chose $u$ to be identical in all representations.

Since $U_E$ and $U_G$ only differ by an additive transformation, we can identify that $A(E, \gamma, F) = \frac{1}{\mu_{E \cup F}(E)}$. Moreover, for convenience we substitute $B(E, \gamma, F) = B(\mu(E), \mu_E^V, \mu(F), \mu_F^V) + t_E - A(E, \gamma)(-t_{E \cup F} + \mu(E) u(\gamma) + h_{\mu_{E \cup F}}(\mu_{E \cup F}(F), \mu_F^V)).$
We then obtain:

\[
\sum_{a} h_{\mu_E}(\mu(a^{-1}(a)), \mu_E) = \sum_{a} \frac{1}{\mu_{EUF}(E)} h_{\mu_E}(\mu_{EUF}(E)\mu(a^{-1}(a)), \mu_E) + \bar{B}(\mu(E), \mu_E, \mu(F), \mu_F)
\]  

(155)

Thus, \( \bar{B}(\mu(E), \mu_E, \mu(F), \mu_F) = F(\mu_E) + G(\mu_{EUF}(E), \mu_E) \). From the case in which \( \mu_E(a^{-1}(a)) = 1 \) for some \( a \), we have:

\[
h_{\mu_E}(1, \mu_E) = \frac{1}{\mu_{EUF}(E)} h_{\mu_E}(\mu_{EUF}(E), \mu_E) + F(\mu_E) + G(\mu_{EUF}(E), \mu_E)
\]

(156)

and therefore \( h_{\mu_E}(\mu_{EUF}(E), \mu_E) = \mu_{EUF}(E)(f_{\mu_E}(\mu(E)) + g(\mu_E)) \). Next, we consider uninformative conditional acts such that:

\[
U_E(a_E) = \sum_{a} \mu_E(a^{-1}(a))u(a) + h_{\mu_E}(\mu(a^{-1}(a)), \mu_E) + t_E
\]

(157)

For some arbitrary uninformative events \( F \) and \( G \) we determine the knowledge equivalent:

\[
U_{FUG}(\alpha_F \alpha_G) = u(\alpha) + h_{\mu_E}(\mu_{FUG}(F), \mu_E) + h_{\mu_E}(\mu_{FUG}(G), \mu_E) + t_{FUG}
\]

\[
= u(\gamma) + h_{\mu_E}(1, \mu_E) + t_{FUG}
\]

(158)

This can be substituted into \( U_{EUFUG}(\alpha_E \gamma_{FUG}) \) to obtain:

\[
U_{EUFUG}(\alpha_E \gamma_{FUG}) = u(\alpha) + h_{\mu_E}(\mu_{EUFUG}(E), \mu_E) + h_{\mu_E}(1 - \mu_{EUFUG}(E), \mu_E) + t_{EUFUG}
\]

\[
+ \mu_{EUFUG}(F \cup G) \left( h_{\mu_E}(\mu_{FUG}(F), \mu_E) + h_{\mu_E}(\mu_{FUG}(G), \mu_E) - h_{\mu_E}(1, \mu_E) \right)
\]

(159)

Defining \( H(p, \mu^V) = h_{\mu^V}(p, \mu^V) + h(1 - p, \mu^V) - g(\mu^V) = pf_{\mu^V}(p) + (1 - p)f_{\mu^V}(1 - p) \) and repeating the above steps for a knowledge equivalent of \( E \)
and $G$ yields the generalized fundamental equation of information:

$$
H(\mu_{E|FUG}(F), \mu_{E}^{V})
+ (1 - \mu_{E|FUG}(F)) \left( H(\frac{\mu_{E|FUG}(E)}{1 - \mu_{E|FUG}(F)}, \mu_{E}^{V}) - h_{\mu_{E}^{V}}(1, \mu_{E}^{V}) \right)
= H(\mu_{E|FUG}(E), \mu_{E}^{V})
+ (1 - \mu_{E|FUG}(E)) \left( H(\frac{\mu_{E|FUG}(F)}{1 - \mu_{E|FUG}(E)}, \mu_{E}^{V}) - h_{\mu_{E}^{V}}(1, \mu_{E}^{V}) \right). \quad (160)
$$

The general solution to this functional equation is (up to the uniqueness properties of Lemma 16) $pf_{\mu^{V}}(p) = c(\mu^{V})(p \ln p) + pd(\mu^{V}) + e(\mu^{V})$. In the specific case above, we can substitute this solution to obtain:

$$
(1 - \mu_{E|FUG}(F)) \left( e(\mu_{E}^{V}) \right) = (1 - \mu_{E|FUG}(E)) \left( e(\mu_{E}^{V}) \right) \quad (161)
$$

and therefore $e(\mu^{V}) = 0$.

Thus,

$$
U_{E}(a_{E}) = \sum_{\alpha} \mu_{E}(a_{E}^{-1}(\alpha))(\mu_{E}(\alpha) + c(\mu_{E}^{V}) \ln(\mu_{E}(a_{E}^{-1}(\alpha)))) + d(\mu_{E}) + g(\mu_{a_{E}^{-1}(\alpha)}) + t_{E}
$$

(162)

We now consider an arbitrary informative conditional act $a_{E}$. Let $\gamma$ be the knowledge equivalent of distinguishing events $G$ and $F$. Solving $U_{FUG}(\gamma_{FUG}) = U_{FUG}(\alpha_{FUG}a_{\gamma}^{G})$ for $u(\gamma)$ and inserting the result into $U_{E|FUG}(a_{E}\gamma_{FUG}) = U_{E|FUG}(a_{E}\alpha_{FUG}^{G})$ yields after simplifying:

$$
\left( c(\mu_{E|FUG}) - c(\mu_{FUG}) \right)(\mu_{E|FUG}(F) \ln \mu_{E|FUG}(F) + \mu_{E|FUG}(G) \ln \mu_{E|FUG}(G) - \mu_{E|FUG}(F \cup G) \ln \mu_{E|FUG}(F \cup G)) = 0 \quad (163)
$$

and therefore $c(\mu_{E|FUG}) = c(\mu_{FUG})$. Lastly, we choose $t_{E}$ such that $u(\gamma) = U_{E}(\gamma_{E})$ which implies $g(\mu_{E}^{V}) = -d(\mu_{E}^{V})$ and obtain the desired representation.

\[
\Box
\]

**References**


