Australasian Journal of Philosophy

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/rajp20

Foundations for Knowledge-Based Decision Theories

Zeev Goldschmidt

To cite this article: Zeev Goldschmidt (05 May 2024): Foundations for Knowledge-Based Decision Theories, Australasian Journal of Philosophy, DOI: 10.1080/00048402.2024.2328635

To link to this article: https://doi.org/10.1080/00048402.2024.2328635

Published online: 05 May 2024.

AUSTRALASIAN JOURNAL OF PHILOSOPHY
https://doi.org/10.1080/00048402.2024.2328635

Foundations for Knowledge-Based Decision Theories

Zeev Goldschmidt

London School of Economics and Political Science

Several philosophers have proposed *Knowledge-Based Decision Theories* (KDTs)—theories that require agents to maximize expected utility as yielded by utility and probability functions that depend on the agent's knowledge. Proponents of KDTs argue that such theories are motivated by *Knowledge-Reasons norms* that require agents to act only on reasons that they know. However, no formal derivation of KDTs from Knowledge-Reasons norms has been suggested, and it is not clear how such norms justify the particular ways in which KDTs relate knowledge and rational action. In this paper, I suggest a new axiomatic method for justifying KDTs and providing them with stronger normative foundations. I argue that such theories may be derived from constraints on the relation between knowledge and preference, and that these constraints may be evaluated relative to intuitions regarding practical reasoning. To demonstrate this, I offer a representation theorem for a KDT proposed by Hawthorne and Stanley (2008) and briefly evaluate it through its underlying axioms.

ARTICLE HISTORY Received 28 February 2022; Revised 16 February 2023

KEYWORDS knowledge; Decision Theory; action; rationality; reasons

1. Knowledge-Based Decision Theories

Noa is deliberating whether to drive to work or take the train. What should Noa choose? To answer this question, a standard decision theorist may ask several additional questions: What consequences do these actions yield in different states of the world? What is Noa's utility from these consequences? What are Noa's credences regarding the relevant events? and so forth. The standard decision theorist will not, however, need to ask us: What does Noa know? For them, Noa's rational action is determined by her utility and subjective probability functions, independently of her knowledge. Indeed, knowledge plays no explicit role in standard decision theory and is not represented by its formal apparatus.

In contrast, for some philosophers, Noa's knowledge is indispensable for determining what she ought to do. Such philosophers have proposed decision theories that I will term *Knowledge-Based Decision Theories* (KDTs), in which knowledge plays a foundational role in determining rational action and preference—according to such theories, what one ought to do depends on what one knows. These theories require agents to maximize expected utility in a knowledge-dependent manner—they require that the utility and probability functions that determine this expectation stand in certain relations to the agent's knowledge. According to KDTs, rational action is determined by utility and subjective probability, the latter depend on the agent's knowledge, and therefore rational action is knowledge-dependent.

In standard decision theory,[1] the objects of normative interest are the agent's preferences among alternative courses of action. The theory requires that preferences, represented by a binary relation $\succeq$, stand in a certain relation to a pair of functions—the agent's utility and subjective probability functions: $\langle u, P\rangle$. Specifically, the theory requires that the agent's preference relation order the alternative actions in accordance with their expected utility as yielded by these functions.

In contrast, the objects of normative interest in KDT are not preferences *simpliciter*, but the relation between the agent's knowledge and their preferences. KDT is thus concerned with the agent's preferences given their current knowledge, their preferences given other possible knowledge states, and the ways in which preferences may evolve with changes in knowledge. This notion of knowledge-dependent preference is best represented by multiple preference relations indexed to the agent's different possible knowledge states, rather than a single preference relation: for every knowledge state E, the relation $\succeq_E$ represents the preferences that the agent would have if they were in that state.[2] KDT then requires that for any possible knowledge state E, the agent's knowledge-dependent preferences $\succeq_E$ maximize expected utility with respect to a knowledge-dependent pair of utility and probability functions $\langle u^E, P^E\rangle$.

The space of logically possible KDTs is vast—there are many ways of constraining the relation between knowledge states and knowledge-dependent pairs of utility and probability functions. The literature includes several such theories. Hawthorne and Stanley (2008) tentatively suggest a knowledge-based decision theory that requires expected utility maximization relative to the agent's utility function and a subjective probability function on which known propositions receive probability 1. Another natural KDT replaces subjective probability with Williamson's (2000) *evidential probability*. This theory requires agents to maximize expected utility with respect to their utility function and the evidential probability function conditionalized on their knowledge.[3] Schulz (2017) suggests a similar theory where the set of propositions on which one ought to conditionalize P^E depends on the stakes. For actions with low stakes, evidential probability is conditionalized on known propositions, and Schulz's decision theory converges with the previous suggestion. As the stakes rise, the theory requires conditionalizing evidential probability on propositions for which the agent has some degree of higher-order knowledge.[4]

It is important to distinguish knowledge-based decision theories from decision theoretic necessary conditions for knowledge. Some philosophers have argued that an agent knows that p only if it is rational for them to prefer *as if* p, that is, only if their preferences among some set of available or salient actions are identical to their

[1] The version I have here in mind is Savage's (1972), though much of what I say applies to other versions as well.

[2] In section 3, I associate knowledge states with intersections of known propositions. However, since this involves additional assumptions (discussed there), I leave the notion of a knowledge state formally unanalysed at this stage.

[3] See Hawthorne 2005 and Williamson 2005a for a critical discussion of this KDT.

[4] Moss (2018) presents a knowledge-based decision theory in which rational action is indexed to the agent's *probabilistic* knowledge. The expected utility of actions depends on the agent's utility function and on the properties of their credence function that constitute knowledge. Elga and Rayo (2022) offer a theory on which rational action depends on *accessible* knowledge rather than knowledge *simpliciter*. Though I believe that some of my claims are applicable to these theories as well, I will not discuss them in what follows.

preferences conditional on p (Fantl and McGrath 2002).[5] In contrast to knowledge-based decision theories, this principle does not require any alteration of standard decision theory, rather, it presupposes some such theory and requires that knowledge stand in a certain relation to it. Specifically, given some set of available or salient actions A, standard decision theory determines which propositions p are such that the agent's actual preferences over A are identical to their preferences over A conditional on p. The condition above merely requires that the agent's knowledge be some subset of this set of propositions.[6] It is therefore a constraint on knowledge, not on rational preference.

The plan for the rest of the paper is as follows. In section 2, I argue that KDTs and theories of practical reasoning conjoined with Knowledge-Reasons norms generate constraints on the same relation—the relation between knowledge and preference. In section 3, I present a formal framework in which constraints on this relation may be expressed and from which KDTs may be derived. In section 4, I present a representation theorem for a specific KDT. In section 5, I briefly evaluate the axioms of the theory before drawing some conclusions in section 6.

2. Knowledge-Based Decision Theory and Knowledge-Reasons Norms

Knowledge-based decision theories generate constraints on the relation between knowledge and rational preference. The preferences of an agent with knowledge state E must maximize expected utility as calculated by the knowledge-dependent functions in $\langle u^E, P^E \rangle$. How may proponents of KDTs justify such constraints on the relation between knowledge and rational preference?

Proponents of KDTs often argue that their theories follow from norms governing the relation between knowledge and practical reasoning. Several philosophers have argued that agents should act only on reasons that they know, thus rendering knowledge foundational to practical reasoning. Hawthorne and Stanley (2008) argue that it is appropriate to treat p as a reason for one's action only if one knows that p. In a similar vein, Williamson (2017) argues that one ought to use the premise that p in practical reasoning only if one knows that p.[7] Let us call such norms—norms binding the relation between knowledge and reasons for action (or practical reasoning premises)—*Knowledge-reasons norms* (KR norms).

A proponent of KDT may argue that if KR norms are correct, and all practical reasoning must stem from known propositions, then knowledge should play a foundational role in decision theory. In contrast to standard decision theories that make no reference to knowledge, KDTs place knowledge front and centre, and are plausible explications of the purported centrality of knowledge to practical reasoning. Indeed, Hawthorne and Stanley (2008) and Schulz (2017) argue that their KDTs are motivated by KR norms.

However, while this argument suggests that knowledge should play a central role in decision theory, it fails to justify the precise requirements of knowledge-based decision

[5] Weatherson (2005) argues for such a condition for belief.

[6] Formally, let $\preceq\restriction_{A\times A}$ denote the restriction of the agent's preference relation to A, and $\preceq_p\restriction_{A\times A}$ denote that restriction conditional on p. Then the principle requires that knowledge be a subset of $X = \{p | \preceq\restriction_{A\times A} = \preceq_p\restriction_{A\times A}\}$. Determining X does not require the introduction of KDT; it is determined entirely by standard decision theory.

[7] See also Hyman 1999, Hawthorne 2004, Williamson 2005b, and Stanley 2005 for similar norms.

theories, for example, the requirement to maximize expected utility as calculated when using evidential probability. How do KR norms, such as the requirement to treat only known propositions as reasons, entail anything like knowledge-dependent expected utility maximization? If KR norms are to be used to justify KDTs, then a more careful derivation is required.

Deriving knowledge-based decision theories from KR norms would require elucidating the relation between reasons for action and decision theory. Notice that KR norms and the norms prescribed by KDTs generate constraints on two different relations. The former norms bind the relation between an agent's knowledge and their *potential reasons*—the set of propositions they may appropriately treat as reasons for their actions or preferences[8]—while the latter bind the relation between the agent's knowledge and their preferences. If KDT is to be derived from KR norms, then more must be said regarding the relation between potential reasons and rational preference.

This relation, between potential reasons and rational preference, would be normatively constrained by any adequate theory of practical reasoning. If practical reasoning is a procedure that takes reasons as premises and yields preferences as conclusions, then norms for such reasoning would specify which preferences may be adequately inferred from which sets of reasons. Different theories would place different constraints, but any adequate theory would place *some* constraints on this relation.

Thus, an adequate theory of practical reasoning would specify, for any set of potential reasons, the preferences that may be reached via good practical reasoning from that set of reasons. What is the relation between the set of preferences licensed by our theory of practical reasoning and the set of preferences that are rationally permissible for an agent with those reasons? It seems plausible that our theories of rationality and practical reasoning should be aligned in a way that renders these sets equal. If a preference is rationally permissible for an agent, then it should be possible for them to arrive at it as a conclusion of practical reasoning, and sound practical reasoning should not allow the inference of rationally impermissible preferences. These consistency conditions seem like plausible desiderata for any pair of theories of rationality and practical reasoning.[9] Therefore, an adequate theory of practical reasoning should generate constraints on the relation between potential reasons and rational preference.

Therefore, in conjunction with KR norms—which require that all of the agent's potential reasons be known—any plausible theory of practical reasoning would generate constraints on the relation between the agent's knowledge and their preferences. Since knowledge-based decision theories generate constraints on the same relation,

[8] Reasons for action and reasons for preference appear to be very closely related—if p is a reason for choosing one action over another, then it must also be a reason for preferring one action over another. This is so regardless of whether preference is understood behaviouristically—as hypothetical choice (Arrow 1959), or mentalistically—as all-things-considered betterness judgments (Hausman 2011). On a behaviouristic conception, reasons for preference are just reasons for choosing one action over another when presented with a choice between the two. On a mentalistic interpretation, reasons for preference are reasons for judging one action to be all-things-considered better than another, which presumably would also be reasons for choosing it. Therefore, while KR norms are stated in terms of reasons for action, they may also be thought of in terms of reasons for preference.

[9] The plausibility of these desiderata is quite independent of whether our decision theory is knowledge-based. Decision theory and practical reasoning theory have overlapping subject matters, and therefore accounting for the relation between them is an important task. See Dietrich and List 2013 and Lin 2013 for different accounts of this relation.

perhaps such theories may be derived from KR norms conjoined with norms of practical reasoning.

Proponents of KDTs still must demonstrate how such a derivation is possible. Specifically, for any knowledge state E, it must be shown that the preference relation prescribed by the relevant KDT for E aligns with our intuitions regarding practical reasoning. That is, it must be shown that the preference relation maximizing expected utility as yielded by $\langle u^E, P^E \rangle$ stands in a plausible relation to the potential reasons the agent has in E, given some norms of practical reasoning. The proponent of KDT need not provide a full theory of practical reasoning to demonstrate this. It is sufficient for them to argue that several constraints on the knowledge-preference relation would be required by any adequate theory of practical reasoning, and then demonstrate how these constraints are sufficient for the derivation of KDT.

In the following section I provide a framework in which such constraints may be formalized, and such derivations may be carried out.

3. A Formal Framework for Knowledge-Based Decision Theories

As stated above, the object of normative interest in KDTs is the relation between the agent's knowledge and their preferences. In this section, I propose a formal framework in which constraints on this relation may be formalized, and potentially, knowledge-based decision theories may be derived from them. If the underlying constraints of such a derivation are supported by our intuitions regarding practical reasoning, then the derivation may be leveraged to *justify* the derived theory. In contrast, KDTs may be criticized by demonstrating that they entail constraints on the knowledge-preference relation that do not accord with intuitions regarding practical reasoning.

A formal representation of the knowledge-preference relation requires representing knowledge, preference, and the objects of preference. In representing knowledge, I follow the picture suggested by Hintikka (1962).[10] I introduce a set of possible states Ω at which propositions may be true or false. Propositions (or *events*) correspond to the sets of states at which they are true.[11] The agent's knowledge state is represented by the set of states that are consistent with all of the propositions that they know, that is, the intersection of those propositions. I will refer to this set as the agent's *ken*.[12] It is assumed that knowledge is factive and therefore the agent's ken is never empty.[13] Additionally, it is assumed that the agent knows everything that is true throughout their ken. This implies that knowledge is closed under logical entailment, for the propositions represented in Ω.[14]

[10] For further applications of this framework, see Lewis 1986, Aumann 1999, and Samet 2022.
[11] The negation of a proposition corresponds to its complement. Conjunctions and disjunctions correspond to intersections and unions respectively.
[12] Notice that the term *Ken* is often used to denote the set of known propositions rather than their intersection (Aumann 1999; Samet 2022).
[13] All known propositions are true in at least one state, namely in the actual state.
[14] In contrast to Lewisian possible worlds, states in Ω are assumed to be partial descriptions of possible worlds. They may specify the truth values of some propositions and remain silent about others. Nonetheless, the logical omniscience entailed here is a substantial idealization that may impair the normative validity of conclusions drawn from this formalism. Though I will not examine here how pernicious this idealization is, I suspect that many of the problems it generates are present in standard decision theory as well (e.g., decision theoretic certainty is closed under logical entailment). If so, logical omniscience cannot count against KDT and in favour of standard decision theory.

Kens represent both the agent's knowledge and their ignorance—the agent knows that the actual state is a member of their ken but is ignorant of which one of these possibilities is actual. Therefore, the more states included in the agent's ken the less they know—their knowledge rules out fewer states. Different partitions of the agent's ken represent different ways of describing their ignorance. Consider for example an agent who knows that it snowed in Jerusalem in 2013 (p), does not know whether Vivaldi wrote music for mandolin (q), and does not know whether the Celtics will win the NBA finals this year (r). The agent's knowledge may be represented by positing a set of eight possible states representing the different possible truth-value assignments of these three propositions, as depicted in table 1 below:

Table 1: Set of states for representing the agent's knoedge

	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8
p	*T*	*T*	*T*	*T*	*F*	*F*	*F*	*F*
q	*T*	*T*	*F*	*F*	*T*	*T*	*F*	*F*
r	*T*	*F*	*T*	*F*	*T*	*F*	*T*	*F*

The agent's ken is the set $E = \{\omega_1, \omega_2, \omega_3, \omega_4\}$. p is true in all of the agent's epistemically possible states, thus representing their knowledge that p. Neither q nor r are true throughout the agent's ken, thus representing their ignorance regarding both propositions. Different partitions of E represent different descriptions of the agent's ignorance. For example, the partition $\{\{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}\}$ represents the claim that the agent doesn't know whether Vivaldi wrote music for mandolin, and the partition $\{\{\omega_1, \omega_3\}, \{\omega_2, \omega_4\}\}$ represents the claim that the agent doesn't know whether the Celtics will win the finals.

The framework also allows representing changes in the agent's knowledge, as movements between kens. Learning, or gaining knowledge, is represented by moving from an initial ken to one of its subsets. For example, if the agent comes to know that Vivaldi wrote a mandolin concerto, their ken will change from E to $\{\omega_1, \omega_2\}$, a proper subset of E. In contrast, knowledge loss is represented in the framework by moving from a ken to one of its supersets. For example, if the agent forgets that it snowed in Jerusalem in 2013 their ken will change from E to Ω—this loss of knowledge renders more states epistemically possible, specifically, ones in which it didn't snow in Jerusalem in 2013. Kens may thus be partially ordered with respect to knowledgeability—a ken D is at least as knowledgeable as E if and only if the former is a subset of the latter.

While all possible kens are nonempty sets of states, the converse is not true. Moore-paradoxical propositions of the form 'p and I don't know that p' may be nonempty subsets of Ω but are (arguably) unknowable. Therefore, nonempty subsets of Moore-paradoxical propositions cannot represent possible knowledge states. Notice that since states in Ω are not possible worlds and need not be determinate regarding all matters, Ω need not include Moore-paradoxical propositions. However, such propositions may be practically significant in some decision situations. Such situations and the problems they may generate for KDTs are discussed in section 5.2.

I follow Savage (1972) in representing actions with uncertain consequences—the objects of preference—as functions from states to outcomes. For this, let O denote a finite set of outcomes whose elements represent the possible consequences of actions in as much as they matter to the agent.[15] The states in Ω are assumed to be sufficiently fine-grained to include all contingencies alluded to by the actions of interest. Actions are represented by functions from states to outcomes, and the set of possible actions include all such functions ($\mathbb{A} = O^{\Omega}$). These functions (also termed *acts*) assign each state with the outcome that the represented action will yield if that state is to materialize.

Preferences in Savage's decision theory are represented by a binary relation $\succeq$ over the set of actions $\mathbb{A}$. However, in contrast to Savage's formalism, my aim is to represent the relation between knowledge and preference and not preference *simpliciter*. For this purpose, I introduce a function k from possible kens, into possible preference relations. For any possible ken E, $k(E) = \succeq_E$ represents the manner in which the agent would prefer had they possessed the body of knowledge represented by E (had they known all and only supersets of E). Therefore, k represents the agent's knowledge-dependent preferences—it specifies the agent's preferences for any body of knowledge they may possess.

As stated in the previous sections, knowledge-based decision theories are constraints on the function k. Specifically, they require that for all possible kens E, $\succeq_E$ maximizes expected utility as yielded by u^E, P^E. In addition, if we accept KR norms, then k is a function from potential reasons to preferences and is thus constrained by our conception of practical reasoning. Therefore, constraints on the function k may serve as a nexus between practical reasoning and knowledge-based decision theories, allowing intuitions regarding the former to evaluate the latter.

To generate support for their theories, proponents of KDTs should formulate constraints on the function k that satisfy two conditions. First, the KDT of interest may be derived from the constraints, and second, the constraints are supported by intuitions regarding practical reasoning and the manner in which potential reasons may give rise to preferences. Conversely, one may argue against a KDT by demonstrating how it entails constraints on the knowledge-preference relation that are at odds with practical reasoning intuitions. In the following section, I will formulate a set of constraints that is equivalent to a specific knowledge-based decision theory. The theory may then be evaluated by considering the plausibility of the underlying constraints.

4. A Derivation of KDT

In this section I will utilize the formal framework presented in section 3 to derive a knowledge-based decision theory from a set of axioms—constraints on the function k that do not allude to utility or probability. The derivation allows evaluating the theory through the underlying axioms. If the axioms are plausible and align with our intuitions regarding knowledge, practical reasoning, and preference, then they may serve as a justification for the theory. If they are implausible, they may serve as evidence against it.

Consider the following knowledge-based decision theory: for every possible ken E, the agent should maximize expected utility with respect to the pair u^E, P^E such that:

1. For any two kens E and E', $u^E = u^{E'}$.

[15] Assuming that O is finite simplifies the theorem in the next section (it allows omitting Savage's P7). However, I believe that a corresponding theorem (with an extra axiom) may be proved for infinite outcome sets.

2. For any two kens E and D, if $D \subseteq E$ and $P^E(D) \neq 0$, then for any proposition p, $P^D(p) = P^E(p|D)$.

The proposed theory requires (1) that utility be knowledge-independent and (2) that learning new propositions alters probability by conditionalization. This theory is a natural extension of the one tentatively proposed by Hawthorne and Stanley (2008), which requires that the probability of known propositions always be 1 but is silent about how the probabilities of other propositions should evolve when the agent comes to know a new proposition. Since conditionalization is a well-motivated constraint on this type of evolution, quite independently of KDTs (Diaconis and Zabell 1982; Dietrich, List, and Bradley 2016), the theory examined here is a plausible extension of Hawthorne and Stanley's.

This knowledge-based decision theory is equivalent to seven constraints on the function k that do not allude to utility or subjective probability. To demonstrate this, I will use the following definitions and notational conventions:

Definition 1 Acts f and g *agree* on event A iff $f(\omega) = g(\omega)$ for all $\omega \in A$. For any acts f and g, and event A, let 'f_A^g' denote the act that agrees with f on $\neg A$, and with g on A. Formally: $f_A^g(\omega) = \begin{cases} g(\omega), & \text{if } \omega \in A \\ f(\omega), & \text{if } \omega \in \neg A \end{cases}$.[16]

Definition 2 For any two acts f and g, and any event A, $f \preceq g$ *given* A iff there exists an act h such that $f_{\neg A}^h \preceq g_{\neg A}^h$.

Definition 3 An event A is *B-null* iff for any two acts f and g, $f \preceq_B g$ given A.

Definition 4 The set of *all non-null* kens, denoted $\mathbb{K}$, is the set of all events that are not Ω-null.

Definition 5 For any ken E, the *symmetric* and *asymmetric* parts of $\preceq_E$ are defined as follows: $f \sim_E g$ iff $f \preceq_E g$ and $g \preceq_E f$; $f \prec_E g$ iff $f \preceq_E g$ and $\neg(g \preceq_E f)$.

Definition 6 A binary relation over $\mathbb{A}$, $\preceq$, is *representable* iff there exists a non-constant utility function u: $O \to \mathbb{R}$, unique up to positive linear transformations, and a unique probability function P on Ω such that for all acts f and g: $f \preceq g$ if and only if $\int_\Omega u(f(\omega))dP(\omega) \leq \int_\Omega u(g(\omega))dP(\omega)$. In such a case we say that $\preceq$ is *representable* by $< P, u >$.

Definition 7 If P is a probability function over Ω, and $P(E) > 0$, then the *probability function conditional* on E, denoted P_E,[17] is defined as follows: for all events A,

$P_E(A) = P(A \cap E)/P(E)$.

Definition 8 For any outcome x, the constant act that yields the outcome x on every state of Ω will be denoted by x.

I will now present the axioms that are used in the theorem. The **K**-axioms pertain to different images of k, while the **P**-axioms concern only $\preceq_\Omega$, and are mere applications of Savage's postulates to the relation $\preceq_\Omega$:

[16] I adopt this notational convention from Gilboa (2009: 99).
[17] Notice that in contrast to P_E, P^E denotes the agent's probability when their ken is E. A KDT that requires conditionalizing on knowledge, like the one discussed here, requires that $P^E = P_E$.

K1 For all non-null kens $E \in \mathbb{K}$, $\preceq_E$ is complete and transitive over $\mathbb{A}$.
K2 For all non-null kens $E \in \mathbb{K}$, and for all acts f and g.
1. if there exists a finite partition Π^E of E such that for all $\pi_i \in \Pi^E$, if $\pi_i \in \mathbb{K}$ then $f \preceq_{\pi_i} g$, then $f \preceq_E g$
2. If in addition there exists $\pi_j \in \Pi^E \cap \mathbb{K}$ such that $f \prec_{\pi_j}$ g, then $f \prec_E g$.

K3 For all non-null kens $E \in \mathbb{K}$, and for all outcomes x and y, $\mathrm{x} \preceq_\Omega \mathrm{y}$ if and only if $\mathrm{x} \preceq_E \mathrm{y}$.[18]
P4 For all events A and B, outcomes $x, \bar{x}, y, \bar{y}$ such that $\bar{\mathrm{x}} \prec_\Omega \mathrm{x}$ and $\bar{\mathrm{y}} \prec_\Omega \mathrm{y}$, $\bar{\mathrm{x}}^{\mathrm{x}}_A \preceq_\Omega \bar{\mathrm{x}}^{\mathrm{x}}_B$ if and only if $\bar{\mathrm{y}}^{\mathrm{y}}_A \preceq_\Omega \bar{\mathrm{y}}^{\mathrm{y}}_B$.
P5 $\Omega \in \mathbb{K}$; equivalently, there exist acts f and g such that $f \prec_\Omega g$.
P6 For all acts f and g and outcome x, if $f \prec_\Omega g$ then there exists a finite partition Π^Ω of Ω such that for all $\pi_i \in \Pi^\Omega$, $f^x_{\pi_i} \prec_\Omega g$ and $f \prec_\Omega g^{\mathrm{x}}_{\pi_i}$.
K7 For all non-null kens $E \in \mathbb{K}$ and for all acts f and g, if f and g agree on E, then $f \sim_E g$.

Theorem The following are equivalent:

I. The function k satisfies the seven axioms above.
II. $\preceq_\Omega$ is representable by some $\langle P, u\rangle$, and for all non-null kens $E \in \mathbb{K}$, $\preceq_E$ is representable by $\langle P_E, u\rangle$. [19]

Notice that the derived decision theory is precisely the one presented at the beginning of this section. Utility is knowledge-independent as u is the same function for all kens, and learning affects probability by conditionalization because for all non-null kens E, $P(E) > 0$ and $P^E(\cdot) = P(\cdot|E)$.[20] That is, the knowledge-dependent probability function is generated by conditionalizing P on the agent's ken, when it is non-null.[21] Also, notice that the knowledge-dependent utility and probability pairs are entirely derived; the axioms make no reference to either notion.

The theorem may be used to evaluate the derived knowledge-based decision theory. If the axioms align with intuitions regarding practical reasoning and the constraints it induces on the relation between potential reasons and preferences, then the theorem may be instrumental in supporting the theory. In contrast, if the axioms are counter-intuitive and do not align with the above intuitions then the theorem may be used to undermine the proposed theory. In short, the KDT above is plausible in as much as its underlying axioms are. I now turn to explain and briefly evaluate the axioms.

[18] Following Definition 8, x is the constant act that yields the outcome x.
[19] See appendix for the proof.
[20] It is a consequence of Savage's theorem (see appendix), that all events that are not Ω-null have positive subjective probability. The set of non-null kens is defined as the set of events that are not Ω-null (Def. 4).
[21] Notice that the derived theory allows for knowledge loss. When an agent moves from one ken E_1 to a less knowledgeable ken E_2 with $E_1 \subseteq E_2$ (e.g., by forgetting some knowledge), their subjective probability will shift from $P_1 = P(\cdot|E_1)$ to $P_2 = P(\cdot|E_2)$, and propositions that were known at E_1 and received probability 1 on P_1 may not be known at E_2 and may receive lower probability on P_2. P_2 is not obtained by conditionalizing P_1 on E_2 (that would just give us P_1), rather by conditionalizing P on E_2. That the agent always conditionalizes P, rather than their current probability, on their ken allows accommodating knowledge loss. Williamson's evidential probability also possesses this property which amounts to a failure of the condition he terms *monotonicity* (Williamson 2000: 218). I thank the editor of this journal for pressing me to elucidate this point.

5. Discussion

In this section I will briefly evaluate the axioms of the above theorem. My purpose is not to justify or refute the derived theory, rather it is to demonstrate how excavating the foundations of KDTs may be instrumental in their evaluation. In particular, revealing the axioms enables the application of intuitions regarding practical reasoning and the relation between knowledge and action for the assessment of knowledge-based decision theories.

Before addressing the axioms directly, notice that they constrain only non-null kens (kens in $\mathbb{K}$) and remain silent regarding preferences for null kens. Also, preferences for null kens do not constrain other preferences in any way. In the derived KDT, this property is manifested in the fact that the theory is silent regarding kens E for which $P^{\Omega}(E) = 0$. Is this restriction plausible? Should a knowledge-based decision theory provide verdicts regarding preferences for null kens?

On the one hand, there is a sense in which null kens are unimportant for the agent's reasoning. Since null events receive zero subjective probability, null kens represent knowledge states that the agent considers it maximally unlikely to be in, and thus may arguably discard them from their reasoning process. For example, one may argue that while the agent may reason about their preferences had they known that some coin came up heads, they are excused from doing so for preferences contingent on their knowing that some coin landed on its side. The unlikeliness of the agent ever knowing such a proposition justifies ignoring it in reasoning.

On the other hand, zero-probability events are not necessarily impossible. For example, each possible outcome of tossing a coin infinitely many times should receive zero subjective probability, but all such outcomes are possible, nonetheless.[22] If null events are sometimes possible, then the agent may come to know a null ken, and perhaps KDT should constrain knowledge-based preferences for those cases as well. I will remain neutral regarding this matter and merely point out that if we want normative constraints for null kens, we must construct a stronger KDT than the one derived above.[23]

Before addressing the substantial axioms, let me briefly address axioms P4–P6. As stated above, these axioms are mere applications of Savage's corresponding postulates to the relation $\preceq_{\Omega}$. If these axioms are plausible rationality requirements of preference *simpliciter,* it is reasonable to take them to be plausible for the knowledge-based preference relation $\preceq_{\Omega}$.

5.1 K1: Weak Order

K1 requires that all knowledge-dependent preference relations be transitive and complete. That is, for any ken E, the agent's preference relation given E, $\succeq_E$, should be a weak order. Is K1 a plausible rationality requirement? That very much depends on

[22] See Williamson 2007 and Easwaran 2014 for arguments that differentiate possibility from positive probability.

[23] For example, omitting the 'non-null kens' qualification from the axioms gives rise to a slightly stronger KDT, that is identical to the one in the text on non-null kens but requires complete indifference for null kens. These alternative axioms are slightly more elegant because they do not differentiate between null and non-null kens, but the derived theory has the oddity of requiring total indifference on the event of learning zero-probability propositions.

what one has to say about the normative status of the corresponding axiom in standard decision theory requiring that preferences *simpliciter* be transitive and complete.

K1 is an extension of this corresponding axiom to the knowledge case and may be formulated as the conjunction of the following two claims:

1. Preferences should be transitive and complete.
2. The requirement in 1 is knowledge-independent.

Standard decision theory requires that preferences be transitive and complete, and KDT requires that they be so independently of what the agent knows. That is, according to KDT the standard requirement in 1 must hold in all possible knowledge states.

While K1 is stronger than the corresponding axiom in standard decision theory, it is hard to see how one may reject the former while accepting the latter. If preferences ought to form a weak order but K1 is not normatively required, then there must be some proposition p such that before learning p one's preferences ought to be transitive and complete, but after learning p they need not be. It is not clear what such a proposition would be, and how coming to know it would suffice for waiving the requirements of transitivity and completeness.

While the normative status of transitivity and completeness is controversial (Luce 1956; Temkin 1996; Broome 1999; Chang 2002), K1 does not appear to present any further controversy. K1 is therefore a plausible rationality requirement in as much as transitivity and completeness are plausible conditions in standard decision theory.

5.2 K2: Epistemic Sure Thing Principle

The intuitive appeal of K2 is best demonstrated by the following example from Savage:

> A businessman contemplates buying a certain piece of property. He considers the outcome of the next presidential election relevant. So, to clarify the matter to himself, he asks whether he would buy *if he knew* that the Democratic candidate were going to win, and decides that he would. Similarly, he considers whether he would buy *if he knew* that the Republican candidate were going to win, and again finds that he would. Seeing that he would buy in either event, he decides that he should buy, *even though he does not know which event obtains*. (Savage 1972: 21; my emphasis)

Savage uses the businessman story to motivate his second postulate P2, known as the *Sure Thing Principle*. Though the intuitions motivating the principle in the example concern the relation between knowledge and preference, he does not formulate P2 in epistemic terms because doing so would necessitate the introduction of 'new undefined technical terms referring to knowledge and possibility that would render it mathematically useless without still more postulates governing these terms' (Savage 1972: 22).

In contrast, K2 is an epistemic formulation of the Sure Thing Principle, and it relates directly to the epistemic intuitions solicited by the businessman parable. To see this, let us express the businessman's choice and reasoning formally. For simplicity, let us posit two possible states: ω_1 in which the Democrat is elected and ω_2 in which the Republican is elected, so $\Omega = \{\omega_1, \omega_2\}$. The agent's ken is Ω, as he does not know who will win the election, and he must choose between two available actions: buying the property, denoted b, and not buying it, denoted n. The agent's contemplation involves reasoning about what his preferences would have been had he been more knowledgeable.

Specifically, he considers his preferences under two counterfactual kens—$\{\omega_1\}$ in which he knows that the Democrat won and $\{\omega_2\}$ in which he knows that the Republican won—and concludes that he would prefer to buy in both informational states: $n \prec_{\{\omega_1\}} b$ and $n \prec_{\{\omega_2\}} b$.[24] He then infers, as the Sure Thing Principle requires, that he should have the same preference given his current, less knowledgeable ken as well: $n \prec_{\Omega} b$.

K2 is a generalization of this instance of reasoning. It requires that *whenever* the agent's ken may be partitioned into more knowledgeable kens in which preferences agree (in non-null kens), then the agreeing preferences should be adopted for the current, less knowledgeable state. K2 may be thought of as prohibiting ignorance from affecting preference in certain cases in which it is irrelevant. In many cases, ignorance is extremely relevant to preference. For example, my ignorance regarding whether the Celtics will win the NBA finals next year affects my preferences. If I were to know that they will win, I would accept certain bets that I would currently reject. However, in some cases ignorance is irrelevant to preference. For example, the businessman's ignorance regarding the election results is irrelevant to his preferences regarding the property. Whenever preferences are constant across a partition of the agent's current ken, the extra information in each ken of the partition, is irrelevant. K2 prohibits ignorance from affecting preferences in such cases.[25]

K2 also aligns well with intuitions regarding practical reasoning. Firstly, the businessman's reasoning seems compelling. Secondly, assuming KR norms, K2 prohibits lack of potential reasons from affecting preference when it is irrelevant. The agent has fewer potential reasons when their ken is E, relative to when it is any of the elements of the partition Π^E. However, if preferences are constant across these more knowledgeable kens, then this lack of potential reasons is irrelevant to preference, and $\preceq_E$ should adopt the agreeing parts of the $\preceq_{\pi_i}$ relations.

However, in some cases K2 appears to yield counter-intuitive prescriptions. Consider the following example from Aumann, Hart, and Perry:[26]

> [S]uppose that you consider reading a certain book. If you know it is written by A, you will read it, and if you know it is by B, you will read it. But that does not entail that you will read it if you know that it is by either A or B. Knowing the identity of the writer is important for appreciating the book (for instance, it brings to mind associations to other works by the same author). (Aumann, Hart, and Perry 2005: 10)

Clearly, the knowledge-dependent preferences of the agent in the example are perfectly rational, despite violating K2. This example reveals that K2 fails when knowledge affects the nature of the outcomes of available actions. The agent's knowledge that A wrote the book alters the reading experience because of the associations it induces.

Importantly, standard decision theory gets such cases right. The following table represents the decision problem faced by the above agent:

Table 2: The book decision problem

	$A \cap K(A)$	$A \cap \neg K(A)$	$B \cap K(B)$	$B \cap \neg K(B)$
Read book	*A-associations*	*No associations*	*B-associations*	*No associations*
Watch TV	*TV*	*TV*	*TV*	*TV*

[24] Notice that this entails that neither ken is null.

[25] A similar intuition supports Fleurbaey and Voorhoeve's (2013) *Principle of Full Information*. See also Samet's (2022) principle of *Independence of Irrelevance Knowledge*.

[26] See also the *Boat Race Case* discussed by Fleurbaey and Voorhoeve (2013) and Frick (2013).

A denotes the event that author A wrote the book and B denotes the event that author B did so. $K(A)$ and $K(B)$ denote the events that the agent knows that A is the author, and that they know that B is the author, respectively. The agent must decide between reading the book and watching TV. Reading the book does not dominate watching TV because the agent only prefers the former in states in which they know who wrote the book (*X-associations* are preferred to *TV* but *TV* is preferred to *no associations*). Therefore, the non-epistemic Sure Thing Principle does not require the agent to prefer reading.

The above table allows a more formal statement of the cases in which K2 fails—whenever an action assigns different outcomes to states in $p \cap K(p)$ and in $p \cap \neg K(p)$ for some p.[27] The problem above will arise whenever propositions about the agent's knowledge carry practical import and thus must be included in the formalism to aptly represent the decision problem at hand. As stated in section 3, the Moore-paradoxicality of such propositions (or their complements) generates an additional problem for a KDT decision problem representation. Perhaps KDTs are not applicable in such cases and ought to be restricted to cases in which such propositions are practically unimportant. That is, to cases in which outcomes do not depend on the agent's knowledge.

In sum, while K2 seems to align nicely with intuitions regarding the irrelevance of ignorance, it yields problematic prescriptions in cases in which actions' outcomes depend on the agent's knowledge. This could be used as an argument against K2 and the KDT derived from it, or as an argument for restricting the theory to cases in which this type of knowledge-dependence does not occur.

5.3 K3: Knowledge Independence

K3 requires that preferences over outcomes (constant acts) be independent of the agent's knowledge. For example, if running and swimming are considered outcomes, and I prefer running to swimming given some ken, then I must prefer running to swimming given all other kens as well. This axiom—like Savage's P3—is best understood as a constraint on the specification of outcomes. It requires that outcomes be sufficiently specified to be independent of the agent's knowledge.

To see this, consider the athletic preference above and the proposition p: 'I develop knee problems that make running painful'. It is surely permissible for me to prefer running to swimming when my ken is Ω and I don't know that p, and to prefer the opposite when I know that p. Such cases are evidence of under-specified outcomes rather than violations of K3. In such a case, 'running' should be understood as an uncertain action rather than an outcome. The outcomes in such a decision problem are, 'running with no knee pain', 'running with knee pain', and 'swimming'. With this amendment in place, K3 is not violated: while I prefer running with no knee pain to swimming, I prefer swimming to running with knee pain.

However, since the set of actions over which preferences are defined includes all functions from states to outcomes, it must include the constant act that yields the outcome 'running with no knee pain' in all states, including those in which p is true and running is painful. It seems that there is no real-world action that is aptly

[27] Notice that the existence of this fine-grained partition does not save K2, as the antecedent of K2 is an existential claim and only requires that there be *some* partition that satisfies its condition.

represented by this constant act, but K3 requires that preferences be defined over such acts, nonetheless. However, very similar theoretical oddities are generated in standard decision theory by Savage's P3 (Drèze 1987: 76–81; Bradley 2017: 57–58). Therefore, it is likely that the implausibility of K3 is an artifact of a general decision-theoretic idealization.

5.4 K7: Reliance

K7 requires that agents rely on their knowledge: if the agent knows that two acts will have the same outcome, then they should be indifferent between them. Though initially plausible, this axiom entails that an agent who knows that it's Wednesday today must be indifferent between the following pair of actions:[28]

Table 3: The ice cream decision problem

	It's Wednesday	It's not Wednesday
f	*Get ice cream*	*Die tomorrow morning*
g	*Get ice cream*	*Get ice cream*

Several philosophers have argued that a rational agent need not, and perhaps even should not, be indifferent between f and g (Hawthorne 2005; Weatherson 2012; Greco 2013). Such indifference seems to involve excessive reliance on knowledge. After all, sometimes we mistakenly think that we know that it's Wednesday while in fact it is Thursday. Though there is a sense in which the agent should be indifferent between the actions—as we assumed that they do in fact know that it's Wednesday—there is an important sense in which choosing f is objectionably reckless, or at least not rationally required. If so, K7 may be evidence against the derived KDT.

This problem arises for any KDT that requires giving known propositions a probability of 1. It has led several philosophers to accept pragmatic encroachment on knowledge: the thesis that the agent's knowledge is sensitive to their practical environment—their available actions and the severity of their possible outcomes. On this thesis, the agent's knowledge that it's Wednesday is destroyed by the mere consideration of f.

Williamson (2005a) deals with the problem by distinguishing between rationality conditions posited by decision theory and good cognitive habits. While the former require indifference between f and g, the latter require preferring g. Schulz (2017) proposes a decision theory on which higher stakes require conditionalizing on higher-order knowledge. In our high-stakes case, the mere knowledge that it is Wednesday would be insufficient for indifference, and some degree of higher-order knowledge is required.

I will not adjudicate between these positions. My purpose in this section is not to justify or refute the derived KDT, rather to point out considerations that are relevant

[28] Together with K2, K7 entails that the agent should *strictly prefer h* to g where h yields 'get ice cream and $1' if it's Wednesday and 'die tomorrow morning' otherwise. Most of the literature deals with this type of case, rather than the indifference case. However, the example in the text allows discussing K7 in isolation from K2.

for doing so, and to demonstrate how a representation theorem, like the one presented above, allows evaluating KDTs through their underlying axioms.

6. Conclusion

While knowledge-based decision theories incorporate the plausible idea that knowledge plays an important normative role in determining rational action, their precise normative foundations are unclear. In this paper, I have argued that such foundations may be found in intuitions regarding practical reasoning and the relation between knowledge and action. Under the assumption that reasons for action must be known, the relation between the agent's knowledge and their preferences is the subject matter of both KDTs and theories of practical reasoning. Therefore, KDTs may be evaluated relative to our intuitions regarding practical reasoning.

The formal framework presented above enables the derivation of KDTs from constraints on the knowledge-preference relation that may be evaluated by intuitions regarding practical reasoning. The framework allows utilizing such intuitions to provide KDTs with stronger normative foundations, or to argue against them. The representation theorem and discussion in sections 4 and 5 serve as an example of how this method of evaluation may be employed.

Knowledge is absent from the foundations of standard decision theory. If knowledge is to play a central role in decision theory as proponents of KDTs argue it should, then decision theory must be provided with new foundations. This paper demonstrates one way of doing so.

Acknowledgements

This paper is based on my MA thesis written at the Hebrew University of Jerusalem. I am especially grateful to my supervisors Ittay Nissan-Rozen and David Enoch for their invaluable support. For very helpful conversations and comments on earlier drafts I would like to thank Richard Bradley, Liam Kofi Bright, Christian List, Shahar Zohar, the reviewers and editors of this journal, and audiences at the Hebrew University, the 2019 Knowledge and Decision Conference at the University of Hamburg, and the Choice Group at the London School of Economics.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

ORCID

Zeev Goldschmidt http://orcid.org/0000-0001-5741-2346

References

Arrow, Kenneth J (1959) 'Rational Choice Functions and Orderings', *Economica* **26**: 121–27. doi:10.2307/2550390.

Aumann, Robert J (1999) 'Interactive Epistemology I: Knowledge', *International Journal of Game Theory* **28**: 263–300. doi:10.1007/s001820050111.

Aumann, Robert J, Sergiu Hart, and Motty Perry (2005) *Conditioning and the Sure-Thing Principle*, *Discussion Papers*. https://ratio.huji.ac.il/publications/conditioning-and-sure-thing-principle.

Bradley, Richard (2017) *Decision Theory with a Human Face*. Cambridge University Press.

Broome, John (1999) 'Is Incommensurability Vagueness?', in *Ethics out of Economics*: 123–44. Cambridge University Press.

Chang, Ruth (2002) 'The Possibility of Parity', *Ethics* **112**: 659–88. doi:10.1086/339673.

Diaconis, Persi and Sandy L Zabell (1982) 'Updating Subjective Probability', *Journal of the American Statistical Association* **77**: 822–30. doi:10.2307/2287313.

Dietrich, Franz and Christian List (2013) 'A Reason-Based Theory of Rational Choice', *Noûs* **47**: 104–34. doi:10.1111/j.1468-0068.2011.00840.x.

Dietrich, Franz, Christian List, and Richard Bradley (2016) 'Belief Revision Generalized: A Joint Characterization of Bayes' and Jeffrey's Rules', *Journal of Economic Theory* **162**: 352–71. doi:10.1016/j.jet.2015.11.006.

Drèze, Jacques (1987) *Essays on Economic Decisions under Uncertainty*, *Essays on Economic Decisions under Uncertainty*. Cambridge University Press.

Easwaran, Kenny (2014) 'Regularity and Hyperreal Credences', *Philosophical Review* **123**: 1–41. doi:10.1215/00318108-2366479.

Elga, Adam and Agustín Rayo (2022) 'Fragmentation and Logical Omniscience', *Noûs* **56**: 716–41. doi:10.1111/nous.12381.

Fantl, Jeremy and Matthew McGrath (2002) 'Evidence, Pragmatics, and Justification', *Philosophical Review* **111**: 67–94. doi:10.1215/111-1-67.

Fleurbaey, Marc and Alex Voorhoeve (2013) 'Decide As You Would with Full Information!: An Argument Against Ex Ante Pareto', in Nir Eyal, Samia A Hurst, Ole Frithjof Norheim and Daniel Wikler, eds., *Inequalities in Health: Concepts, Measures, and Ethics*: 113–28. Oxford University Press.

Frick, Johann (2013) 'Uncertainty and Justifiability to Each Person', in Nir Eyal, Samia A Hurst, Ole Frithjof Norheim and Daniel Wikler, eds., *Inequalities in Health: Concepts, Measures, and Ethics*: 129–46. Oxford University Press.

Ghirardato, Paolo (2002) 'Revisiting Savage in a Conditional World', *Economic Theory* **20**: 83–92. doi:10.1007/s001990100188.

Gilboa, Itzhak (2009) *Theory of Decision under Uncertainty*, Econometric Society Monographs. Cambridge University Press.

Greco, Daniel (2013) 'Probability and Prodigality', in Tamar Szabó Gendler and John Hawthorne, eds., *Oxford Studies in Epistemology Vol* **4**: 82–107. Oxford University Press.

Hausman, Daniel M (2011) 'Mistakes about Preferences in the Social Sciences', *Philosophy of the Social Sciences* **41**: 3–25. doi:10.1177/0048393110387885.

Hawthorne, John (2004) *Knowledge and Lotteries*. Clarendon.

Hawthorne, John (2005) 'Knowledge and Evidence', *Philosophy and Phenomenological Research* **70**: 452–58.

Hawthorne, John and Jason Stanley (2008) 'Knowledge and Action', *The Journal of Philosophy* **105**: 571–90.

Hintikka, Jaakko (1962) *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Cornell University Press.

Hyman, John (1999) 'How Knowledge Works', *Philosophical Quarterly* **49**: 433–51. doi:10.1111/1467-9213.00152.

Lewis, David K (1986) *On the Plurality of Worlds*. Basil Blackwell.

Lin, Hanti (2013) 'Foundations of Everyday Practical Reasoning', *Journal of Philosophical Logic* **42**: 831–62. doi:10.1007/s10992-013-9296-0.

Luce, R Duncan (1956) 'Semiorders and a Theory of Utility Discrimination', *Econometrica* **24**: 178–91. doi:10.2307/1905751.

Moss, Sarah (2018) *Probabilistic Knowledge*, *Probabilistic Knowledge*. Oxford University Press.

Samet, Dov (2022) 'The Impossibility of Agreeing to Disagree: An Extension of the Sure-Thing Principle', *Games and Economic Behavior* **132**: 390–99. doi:10.1016/j.geb.2022.01.016.

Savage, Leonard J (1972) *The Foundations of Statistics*. Dover.

Schulz, Moritz (2017) 'Decisions and Higher-Order Knowledge', *Noûs* **51**: 463–83. doi:10.1111/nous.12097.

Stanley, Jason (2005) *Knowledge and Practical Interests*, *Knowledge and Practical Interests*. Clarendon Press.

Temkin, Larry S (1996) 'A Continuum Argument for Intransitivity', *Philosophy and Public Affairs* **25**: 175–210. doi:10.1111/j.1088-4963.1996.tb00039.x.

Weatherson, Brian (2005) 'Can We Do without Pragmatic Encroachment?', *Philosophical Perspectives* **19**: 417–43. doi:10.1111/j.1520-8583.2005.00068.x.
Weatherson, Brian (2012) 'Knowledge, Bets, and Interests', in Jessica Brown and Mikkel Gerken, eds., *Knowledge Ascriptions*: 75–103. Oxford University Press.
Williamson, Timothy (2000) *Knowledge and Its Limits*. Oxford University Press.
Williamson, Timothy (2005a) 'Replies to Commentators', *Philosophy and Phenomenological Research* **70**: 468–91. doi:10.1111/j.1933-1592.2005.tb00542.x.
Williamson, Timothy (2005b) 'Contextualism, Subject-Sensitive Invariantism and Knowledge of Knowledge', *Philosophical Quarterly* **55**: 213–35. doi:10.1111/j.0031-8094.2005.00396.x.
Williamson, Timothy (2007) 'How Probable Is an Infinite Sequence of Heads?', *Analysis* **67**: 173–80. doi:10.1093/analys/67.3.173.
Williamson, Timothy (2017) 'Acting on Knowledge', in J Adam Carter, Emma C Gordon and Benjamin Jarvis, eds., *Knowledge First: Approaches in Epistemology and Mind*: 163–81. Oxford University Press.

Appendix

This appendix includes a proof of the theorem stated in section 4.

Theorem: The following are equivalent:

I. The function k satisfies axioms K1–K7 (see section 4).
II. $\preceq_{\Omega}$ is representable by some $\langle P, u\rangle$, and for all non-null kens $E \in \mathbb{K}$, $\preceq_E$ is representable by $\langle P_E, u\rangle$.

Since the proof relies heavily on Savage's representation theorem, I present it here:

P1 $\preceq$ is complete and transitive.
P2 For all acts f, g, h and $\bar{h}$, and events A, $f^h_{\neg A} \preceq g^h_{\neg A}$ if and only if $f^{\bar{h}}_{\neg A} \preceq g^{\bar{h}}_{\neg A}$.
P3 For all outcomes x and y and non-null[29] events A, $x \preceq y$ if and only if $x \preceq y$ given A.
P4 For all events A and B, outcomes $x, \bar{x}, y, \bar{y}$ such that $\bar{x} \prec x$ and $\bar{y} \prec y$, $\bar{x}^x_A \preceq \bar{x}^x_B$ if and only if $\bar{y}^y_A \preceq \bar{y}^y_B$.
P5 There exist acts f and g such that $f \prec g$.
P6 For all acts f and g and outcome x, if $f \prec g$ then there exists a finite partition Π^{Ω} of Ω such that for all $\pi_i \in \Pi^{\Omega}$, $f^x_{\pi_i} \prec g$ and $f \prec g^x_{\pi_i}$.

Savage's Theorem: The following are equivalent:

- The relation $\preceq$ satisfies the postulates P1–P6.
- $\preceq$ is representable by some $\langle P, u\rangle$.

Furthermore, if $\preceq$ satisfies the postulates P1–P6, then for all non-null events A, the relation $\preceq$ given A is representable by $\langle P_A, u\rangle$.

Proof of Theorem[30]

The proof is composed of two steps that make use of the following proposition:

III. $\preceq_{\Omega}$ satisfies Savage's P1–P6 and COND: for all kens $E \in \mathbb{K}$, $\preceq_E = \preceq_{\Omega}$ given E.

First it will be shown that **I** $\Leftrightarrow$ **III** and then that **III** $\Leftrightarrow$ **II**.

[29] Here null events are defined as in definition 3, with respect to Savage's relation $\preceq$.
[30] The theorem and proof are formally similar to Ghirardato's (2002). However, Ghirardato is not concerned with knowledge, and uses a different axiom instead of my K2.

Part 1: $I \Leftrightarrow III$

First, I assume **I** and prove **III**: that $\preceq_\Omega$ satisfies the axioms P4-P6 is stipulated in **I**.

P1 follows from K1 and P5, because Ω is not Ω-null, and therefore by K1, $\preceq_\Omega$ is complete and transitive. For the rest of the proof, I first prove the following Lemmas:

Lemma 1 If E is E-null then E is Ω-null (from K2 and K7).

Proof. For any acts f and g there is an act h such that $f^h_{\neg E} \preceq_E g^h_{\neg E}$. If $\neg E$ is Ω-null, then by K2 $f^h_{\neg E} \preceq_\Omega g^h_{\neg E}$, and thus E is Ω-null. If $\neg E$ is not Ω-null, then by K7 $f^h_{\neg E} \sim_{\neg E} g^h_{\neg E}$, and by K2 $f^h_{\neg E} \preceq_\Omega g^h_{\neg E}$, and thus E is Ω-null.

Lemma 2 If $E \in \mathbb{K}$ then for all finite partitions Π^E of E, there exists $\pi_i \in \Pi^E$ such that $\pi_i \in \mathbb{K}$ (from K2, K7).

Proof. Assume by way of negation that for some $E \in \mathbb{K}$ there exists a finite partition Π^E such that $\Pi^E \cap \mathbb{K} = \emptyset$. It follows by K2 that for any acts f and g, $f \preceq_E g$, and therefore E is E-null. By Lemma 1, E is also Ω-null contrary to the assumption that $E \in \mathbb{K}$ (definition 4).

Lemma 3 For all events E, if $\neg E \notin \mathbb{K}$ then $E \in \mathbb{K}$ (from K2, P5, K7).

Proof. Follows directly from Lemma 2 and P5, as $\{E, \neg E\}$ is a finite partition of $\Omega \in \mathbb{K}$.

Lemma 4 For all acts f and g and events E:

1. If $f \preceq_\Omega g$ and $\neg E \notin \mathbb{K}$ then $f \preceq_E g$ (from K1, K2, P5, K7).
2. If $f \preceq_\Omega g$ and $f \sim_{\neg E} g$ then $f \preceq_E g$ (from K1, K2, P5, K7).

Proof of 1. Assume by way of negation $\neg(f \preceq_E g)$. From Lemma 3, $E \in \mathbb{K}$ and therefore by K1, $\preceq_E$ is complete thus $\neg(f \preceq_E g) \leftrightarrow f \succ_E g$. It follows by K2 $f \succ_\Omega g$ which (by K1 and P5) is contrary to the assumption that $f \preceq_\Omega g$.

Proof of 2. Assume by way of negation $\neg(f \preceq_E g)$, which is equivalent by K1 to $f \succ_E g$. If $\neg E \notin \mathbb{K}$ the proof follows as in 1. If $\neg E \in \mathbb{K}$ then by K2 it follows that $f \succ_\Omega g$ contrary to the assumption.

Proposition 1 P2 follows from K1, K2, P5, and K7.

Proof. Consider acts f, g, h and $\bar{h}$, and an event E. Assume that $f^h_{\neg E} \preceq_\Omega g^h_{\neg E}$ and prove that $f^{\bar{h}}_{\neg E} \preceq_\Omega g^{\bar{h}}_{\neg E}$. By Lemma 2, $\{E, \neg E\} \cap \mathbb{K} \neq \emptyset$. Therefore, we may consider only the following three situations:

1. If $E \notin \mathbb{K}$: by K7 $f^{\bar{h}}_{\neg E} \sim_{\neg E} g^{\bar{h}}_{\neg E}$, and by K2 $f^{\bar{h}}_{\neg E} \preceq_\Omega g^{\bar{h}}_{\neg E}$.
2. If $\neg E \notin \mathbb{K}$:
   - i By Lemma 4 $f^h_{\neg E} \preceq_A g^h_{\neg E}$.
   - ii By K7 $f^{\bar{h}}_{\neg E} \sim_E f^h_{\neg E}$ and $g^{\bar{h}}_{\neg E} \sim_E g^h_{\neg E}$.
   - iii It follows from i. and ii. and K1 (transitivity of $\preceq_E$) that $f^{\bar{h}}_{\neg E} \preceq_E g^{\bar{h}}_{\neg E}$.
   - iv Therefore, by K2 $f^{\bar{h}}_{\neg E} \preceq_\Omega g^{\bar{h}}_{\neg E}$.
3. If $E \in \mathbb{K}$ and $\neg E \in \mathbb{K}$:
   - i By K7 $f^h_{\neg E} \sim_{\neg E} g^h_{\neg E}$.
   - ii From i. and $f^h_{\neg E} \preceq_\Omega g^h_{\neg E}$ it follows by Lemma 4 $f^h_{\neg E} \preceq_E g^h_{\neg E}$.
   - iii By K7, $f^{\bar{h}}_{\neg E} \sim_E f^h_{\neg E}$ and $g^{\bar{h}}_{\neg E} \sim_E g^h_{\neg E}$.
   - iv It follows from ii. and iii. by K1 (transitivity of $\preceq_E$) that $f^{\bar{h}}_{\neg E} \preceq_E g^{\bar{h}}_{\neg E}$.
   - v By K7 $f^{\bar{h}}_{\neg E} \sim_{\neg E} g^{\bar{h}}_{\neg E}$.
   - vi It follows from iv. and v. by K2 that $f^{\bar{h}}_{\neg E} \preceq_\Omega g^{\bar{h}}_{\neg E}$.

Therefore, in any case, $f^{\bar{h}}_{\neg E} \preceq_\Omega g^{\bar{h}}_{\neg E}$.

Proposition 2 K1, K2, P5, and K7 imply together COND.

Proof. Consider acts f, g and h, and event $E \in \mathbb{K}$. First, assume $f^h_{\neg E} \preceq_\Omega g^h_{\neg E}$, and prove $f \preceq_E g$:

1. If $\neg E \in \mathbb{K}$, we may infer by K7 that since $f^h_{\neg E}$ and $g^h_{\neg E}$ agree on $\neg E$, $f^h_{\neg E} \sim_{\neg E} g^h_{\neg E}$.
2. 1 and the assumption $f^h_{\neg E} \preceq_\Omega g^h_{\neg E}$ imply by Lemma 4 $f^h_{\neg E} \preceq_E g^h_{\neg E}$.
3. If $\neg E \notin \mathbb{K}$, it follows from Lemma 4 that $f^h_{\neg E} \preceq_E g^h_{\neg E}$.
4. In both cases, by K7, $f^h_{\neg E} \sim_E f$ and $g^h_{\neg E} \sim_E g$.
5. 3 and 4 entail by K1 (transitivity of $\preceq_E$) that $f \preceq_E g$.

Assume $f \preceq_E g$, and prove $f^h_{\neg E} \preceq_\Omega g^h_{\neg E}$ (this is equivalent to $f \preceq_\Omega g$ given E for all such h, because as we proved above $\preceq_\Omega$ satisfies P2):

1. $f^h_{\neg E}$ and f agree on E and so do $g^h_{\neg E}$ and g. Therefore, by K7 $f^h_{\neg E} \sim_E f$ and $g^h_{\neg E} \sim_E g$.
2. $f \preceq_E g$ together with 1 entail by K1 (transitivity of $\preceq_E$) that $f^h_{\neg E} \preceq_E g^h_{\neg E}$.
3. If $\neg E \in \mathbb{K}$, since $f^h_{\neg E}$ and $g^h_{\neg E}$ agree on $\neg E$, by K7 $f^h_{\neg E} \sim_{\neg E} g^h_{\neg E}$.
4. 2 and 3 entail by K2 $f^h_{\neg E} \preceq_\Omega g^h_{\neg E}$.
5. If $\neg E \notin \mathbb{K}$ then $f^h_{\neg E} \preceq_\Omega g^h_{\neg E}$ follows from $f^h_{\neg E} \preceq_A g^h_{\neg E}$ by K2.

Proposition 3 K1, K2, K3, P5, and K7 entail that $\preceq_\Omega$ satisfies P3.

Proof. Consider outcomes x and y, and non-Ω-null event E. K3 requires that $x \preceq_\Omega y$ if and only if $x \preceq_E y$. COND (proved above) implies that $x \preceq_E y$ if and only if $x \preceq_\Omega y$ given E. Therefore, $\preceq_\Omega$ satisfies P3.

Second, I assume **III**, and prove **I.**
That $\preceq_\Omega$ satisfies the axioms P4–P6 is stipulated in **III**.

Proposition 4 K1 follows from P1, P2, and COND.

Proof. It follows from P1 and P2 that for all $E \in \mathbb{K}$ the relation $\preceq_\Omega$ given E is transitive and complete (see Savage p. 23), and by COND, $\preceq_\Omega$given $E = \preceq_E$ for all kens E. Therefore, $\preceq_E$ is transitive and complete for all $E \in \mathbb{K}$.

The rest of the proof will make use of Savage's (1972: 24) Theorem 2:

For all events E, and acts f and g:

- If there exists a finite partition Π^E of E such that for all $\pi_i \in \Pi^E$ $f \preceq_\Omega g$ given π_i, then $f \preceq_\Omega g$ given E.
- If in addition there exists $\pi_j \in \Pi^E$ such that $f \prec_\Omega g$ given π_j, then $f \prec_\Omega g$ given E.

Proposition 5 K2 follows from P1, P2, and COND.

Proof. Consider acts f and g, and event $E \in \mathbb{K}$ such that the antecedent of K2 holds: there exists a finite partition Π^E of E such that for all $\pi_i \in \Pi^E$, $\pi_i \in \mathbb{K} \rightarrow f \preceq_{\pi_i} g$. Now we prove $f \preceq_E g$:

1. It follows from COND that for all $\pi_i \in \mathbb{K}$, $f \preceq_\Omega g$ given π_i.
2. Also, for all $\pi_i \notin \mathbb{K}$ $f \preceq_\Omega g$ given π_i (because these π_is are Ω-null).
3. Therefore, for all $\pi_i \in \Pi^E$, $f \preceq_\Omega g$ given π_i, and we may apply theorem 2 to infer $f \preceq_\Omega g$ given E.
4. Applying COND again, it follows that $f \preceq_E g$.

The second part of K2 follows similarly from the second part of theorem 2.

Proposition 6 K7 follows from P1, P2, and COND.

Proof. Consider acts f and g, and event $E \in \mathbb{K}$ such that for all $\omega \in E, f(\omega) = g(\omega)$, and prove $f \sim_E g$:

1. For any act h, $f^h_{\neg E} = g^h_{\neg E}$, and therefore, from the reflexivity of $\preceq_\Omega$, $f^h_{\neg E} \sim_\Omega g^h_{\neg E}$ which is equivalent to $f \sim_\Omega g$ given E.
2. Therefore, applying COND, $f \sim_E g$.

Proposition 7 K3 follows from P3 and COND.

Proof. Consider outcomes x and y, and event $E \in \mathbb{K}$, and prove $x \preceq_\Omega y$ if and only if $x \preceq_E y$.

1. From P3, $x \preceq_\Omega y$ if and only if $x \preceq_\Omega y$ given E.
2. From COND, $x \preceq_\Omega y$ given E if and only if $x \preceq_E y$.
3. Therefore, $x \preceq_\Omega y$. if and only if $x \preceq_E y$.

Part 2: **III** $\Leftrightarrow$ **II**

Recall:

II. $\preceq_{\Omega}$ is representable by some $\langle P, u\rangle$ and for all kens $E \in \mathbb{K}$, $\preceq_E$ is representable by $\langle P_E, u\rangle$.
III. $\preceq_{\Omega}$ satisfies Savage's P1–P6 and COND: for all kens $E \in \mathbb{K}$, $\preceq_E = \preceq_E$ given E.

First let us assume **III** and prove **II**. $\preceq_{\Omega}$ satisfies P1–P6 and therefore by applying Savage's theorem to $\preceq_{\Omega}$. we may infer that $\preceq_{\Omega}$ is representable by me $\langle P, u\rangle$ and for all non-Ω-null events E, $\preceq_{\Omega}$ given E is representable by $\langle P_E, u\rangle$. Together with COND this entails that $\preceq_E$ is also representable by $\langle P_E, u\rangle$ (this follows from the uniqueness up to positive linear ansformation of the representation pair $\langle P_E, u\rangle$).

Now assume **II** and prove **III**. Applying Savage's theorem to $\preceq_{\Omega}$, we may infer that $\preceq_{\Omega}$. satisfies the postulates P1–P6, and that $\preceq_{\Omega}$ given E is represtable by $\langle P_E, u\rangle$ for all non-Ω-null events. Therefore, for all $E \in \mathbb{K}$, the relations $\preceq_{\Omega}$ given E and $\preceq_E$ are representable by the same pair $\langle P_E, u\rangle$. It follows from the uniqueness of representation in Savage's theorem that if two relations are represented by the same probability and utility, then they are identical. Therefore, COND follows.