Define normative theory

Relating to an evaluative standard For other uses, see Normative (disambiguation). This article needs additional citations for verification. Please help improve this article by adding citations to reliable sources. Unsourced material may be challenged and removed.Find sources: "Normative" – news · newspapers · books · scholar · JSTOR (April 2009) (Learn how and when to remove this template message) Normative generally means relating to an evaluative standard. Normativity is the phenomenon in human societies of designating some actions or outcomes as good or desirable or permissible and others as bad or undesirable or impermissible. A norm in this normative sense means a standard for evaluating or making judgments about behavior or outcomes. Normative is sometimes also used, somewhat confusingly, to mean relating to a descriptive standard: doing what is normally done or what most others are expected to do in practice. In this sense a norm is not evaluative, a basis for judging behavior or outcomes; it is simply a fact or observation about behavior or outcomes, without judgment. Many researchers in this[clarification needed] field try to restrict the use of the term normative to the evaluative sense and refer to the description of behavior and outcomes as positive, descriptive, predictive, or empirical.[1][2] Normative has specialised meanings in different academic disciplines such as philosophy, social sciences, and law. In most contexts, normative means 'relating to an evaluation or value judgment.' Normative propositions tend to evaluate some object or some course of action. Normative content differs from descriptive content.[3] One of the major developments in analytic philosophy has seen the reach of normativity spread to virtually all corners of the field, from ethics and the philosophy of action, to epistemology, metaphysics, and the philosophy of science. Saul Kripke famously showed that rules (including mathematical rules, such as the repetition of a decimal pattern) are normative in an important respect.[4][5] Though philosophers disagree about how normativity should be understood, it has become increasingly common to understand normative claims as claims about reasons.[4] As Derek Parfit explains: We can have reasons to believe something, to do something, to have some desire or aim, and to have many other attitudes and emotions, such as fear, regret, and hope. Reasons are given by facts, such as the fact that someone's finger-prints are on some gun, or that calling an ambulance would save someone's life. It is hard to explain the concept of a reason, or what the phrase 'a reason' means. Facts give us reasons, we might say, when they count in favour of our having some attitude, or our acting in some way. But 'counts in favour of' means roughly 'gives a reason for'. The concept of a reason is best explained by example. One example is the thought that we always have a reason to want to avoid being in agony.[6] Philosophy Main article: Ethics In philosophy, normative theory aims to make moral judgements on events, focusing on preserving something they deem as morally good, or prevent a change for the worse.[7] The theory has its origins in Greece.[8] Normative statements make claims about how institutions should or ought to be designed, how to value them, which things are good or bad, and which actions are right or wrong.[9] Normative claims are usually contrasted with positive (i.e. descriptive, explanatory, or constative) claims when describing types of theories, beliefs, or propositions. Positive statements are (purportedly) factual, empirical statements that attempt to describe reality.[citation needed] For example, "children should eat vegetables", and "those who would sacrifice liberty for security deserve neither" are normative claims. On the other hand, "vegetables contain a relatively high proportion of vitamins", and "a common consequence of sacrificing liberty for security is a loss of both" are positive claims. Whether a statement is normative is logically independent of whether it is verified, verifiable, or popularly held. There are several schools of thought regarding the status of normative statements and whether they can be rationally discussed or defended. Among these schools are the tradition of practical reason extending from Aristotle through Kant to Habermas, which asserts that they can, and the tradition of emotivism, which maintains that they are merely expressions of emotions and have no cognitive content. There is large debate in philosophy surrounding the normative and whether you can get a normative statement from an empirical one (ie whether you can get an 'ought' from an 'is', or a 'value' from a 'fact'). Aristotle is one scholar who believed that you could in fact get an ought from an is. He believed that the universe was teleological and that everything in it has a purpose. To explain why something is a certain way, Aristotle believed you could simply say that it is trying to be what it ought to be.[10] On the contrary, David Hume believed you cannot get an ought from an is because no matter how much you think something ought to be a certain way it will not change the way it is. Despite this, Hume used empirical experimental methods whilst looking at the normative. Similar to this was Kames, who also used the study of facts and objective to discover a correct system of morals.[11] The assumption that 'is' can lead to 'ought' is an important component of the philosophy of Roy Bhaskar.[12] Normative statements and norms, as well as their meanings, are an integral part of human life. They are fundamental for prioritizing goals and organizing and planning. Thought, belief, emotion, and action are the basis of much ethical and political discourse; indeed, normativity is arguably the key feature distinguishing ethical and political discourse from other discourses (such as natural science). [citation needed] Much modern moral/ethical philosophy takes as its starting point the apparent variance between peoples and cultures regarding the ways they define what is considered to be appropriate/desirable/praiseworthy/valuable/good etc. (In other words, variance in how individuals, groups and societies define what is in accordance with their normative standards.) This has led philosophers such as A.J. Ayer and J.L. Mackie (for different reasons and in different ways) to cast doubt on the meaningfulness of normative statements. However, other philosophers, such as Christine Korsgaard, have argued for a source of normative value which is independent of individuals' subjective morality and which consequently attains (a lesser or greater degree of) objectivity.[13] Social sciences In the social sciences, the term "normative" has broadly the same meaning as its usage in philosophy, but may also relate, in a sociological context, to the role of cultural 'norms'; the shared values or institutions that structural functionalists regard as constitutive of the social structure and social cohesion. These values and units of socialization thus act to encourage or enforce social activity and outcomes that ought to (with respect to the norms implicit in those structures) occur, while discouraging or preventing social activity that ought not occur. That is, they promote social activity that is socially valued (see philosophy above). While there are always anomalies in social activity (typically described as "crime" or anti-social behaviour, see also normality (behavior)) the normative effects of popularly endorsed beliefs (such as "family values" or "common sense") push most social activity towards a generally homogeneous set. From such reasoning, however, functionalism shares an affinity with ideological conservatism. Normative economics deals with questions of what sort of economic policies should be pursued, in order to achieve desired (that is, valued) economic outcomes. International relations In the academic discipline of International relations, Smith, Baylis & Owens in the Introduction to their 2008 [14] book make the case that the normative position or normative theory is to make the world a better place and that this theoretical worldview aims to do so by being aware of implicit assumptions and explicit assumptions that constitute a non-normative position, and align or position the normative towards the loci of other key socio-political theories such as political liberalism, Marxism, political constructivism, political realism, political idealism and political globalization. Law See also: Normative jurisprudence In law, as an academic discipline, the term "normative" is used to describe the way something ought to be done according to a value position. As such, normative arguments can be conflicting, insofar as different values can be inconsistent with one another. For example, from one normative value position the purpose of the criminal process may be to repress crime. From another value position, the purpose of the criminal justice system could be to protect individuals from the moral harm of wrongful conviction. Standards documents Normative elements are defined in International Organization for Standardization Directives Part 2 as "elements that describe the scope of the document, and which set out provisions". Provisions include "requirements", "recommendations" and "statements". "Statements" include permissions, possibilities and capabilities. A "requirement" is an "expression in the content of a document conveying criteria to be fulfilled if compliance with the document is to be claimed and from which no deviation is permitted." It is not necessary to comply with recommendations and statements in order to comply with the standard; it is necessary to comply only with the requirements (that are denoted by the verbal form "shall"). There is much confusion between "normative" and "requirement", however the ISO terminology is supported by national standards bodies worldwide and is the legitimate description of these terms in the context of standards documents. In standards terminology still used by some organisations, "normative" means "considered to be a prescriptive part of the standard". It characterises that part of the standard which describes what ought (see philosophy above) to be done within the application of that standard. It is implicit that application of that standard will result in a valuable outcome (ibid.). For example, many standards have an introduction, preface, or summary that is considered non-normative, as well as a main body that is considered normative. "Compliance" is defined as "complies with the normative sections of the standard"; an object that complies with the normative sections but not the non-normative sections of a standard is still considered to be in compliance. Normative = prescriptive = how to comply Informative = descriptive = help with conceptual understanding Typically, normative is contrasted with informative (referring to the standard's descriptive, explanatory or positive content). Informative data is supplemental information such as additional guidance, supplemental recommendations, tutorials, commentary as well as background, history, development, and relationship with other elements. Informative data is not a requirement and doesn't compel compliance. See also Look up normative in Wiktionary, the free dictionary. Conformity Decision theory Economics Hypothesis Is-ought problem Linguistic prescription Norm (philosophy) Normative economics Normative ethics Normative science Philosophy of law Political science Scientific method Value References

^ Bicchieri, Cristina (2005). The Grammar of Society:The Nature and Dynamics of Social Norms. Cambridge University Press. ISBN 978-0521574907. ^ Bicchieri, Cristina (2017). Norms in the Wild: How to Diagnose, Measure, and Change Social Norms. Oxford University Press. ISBN 9780190622053. ^ Jarvis., Thomson, Judith (2008). Normativity. Chicago, Ill.: Open Court. ISBN 9780812696585. OCLC 227918828. ^ a b Thomas, Scanlon (2014). Being realistic about reasons (First ed.). Oxford. ISBN 9780199678488. OCLC 862091562. ^ 1940-, Kripke, Saul A. (1982). Wittgenstein on rules and private language : an elementary exposition. Cambridge, Mass.: Harvard University Press. ISBN 0674954009. OCLC 7998796.CS1 maint: numeric names: authors list (link) ^ Parfit, Derek (2011). On what matters Volume one. Edited and Introduced by Samuel Scheffler. Oxford: Oxford University Press. ISBN 9780191576706. OCLC 744616054. ^ Strauss, Leo (2017-09-08), "What Is Political Philosophy? The Problem of Political Philosophy", Behavioralism in Political Science, Routledge, pp. 93–108, ISBN 978-1-351-31436-7, retrieved 2020-12-14 ^ Bauböck, Rainer, "Normative political theory and empirical research", Approaches and Methodologies in the Social Sciences, Cambridge: Cambridge University Press, pp. 40–60, ISBN 978-0-511-80193-8, retrieved 2020-12-14 ^ Leftwich, Adrian (2004). What is politics?: the activity and its study. Oxford: Polity. ^ Gray, JW (July 19, 2011). "The Is/Ought Gap: How Do We Get "Ought" from "Is?"". Ethical realism. Retrieved December 14, 2020. ^ Shaver, Robert. "Hume's Moral Theory?" History of Philosophy Quarterly, vol. 12, no. 3, 1995, pp. 317–331., www.jstor.org/stable/27744669. Accessed 14 Dec. 2020. ^ Leigh Price (2019) Introduction to the special issue: normativity, Journal of Critical Realism, 18:3, 221-238 [1] ^ Korsgaard, C. (1992). "The Sources of Normativity" (PDF). The Tanner Lectures on Human Value. ^ The Globalization of World Politics: An introduction to international relations, New York, Oxford University Press ISBN 9780199297771, Fourth edition, pp.2-13 Further reading Canguilhem, Georges, The Normal and the Pathological, ISBN 0-942299-59-0. Andreas Dorschel, 'Is there any normative claim internal to stating facts?', in: Communication & Cognition XXI (1988), no. 1, pp. 5–16. Retrieved from " First published Fri Aug 8, 2014; substantive revision Thu Aug 15, 2019 The concept of expected utility is best illustrated by example. Suppose I am planning a long walk, and need to decide whether to bring my umbrella. I would rather not tote the umbrella on a sunny day, but I would rather face rain with the umbrella than without it. There are two acts available to me: taking my umbrella, and leaving it at home. Which of these acts should I choose? This informal problem description can be recast, slightly more formally, in terms of three sorts of entities. First, there are outcomes—objects of non-instrumental preferences. In the example, we might distinguish three outcomes: either I end up dry and unencumbered; I end up dry and encumbered by an unwieldy umbrella; or I end up wet. Second, there are states—things outside the decision-maker's control which influence the outcome of the decision. In the example, there are two states: either it is raining, or it is not. Finally, there are acts—objects of the decision-maker's instrumental preferences, and in some sense, things that she can do. In the example, there are two acts: I may either bring the umbrella; or leave it at home. Expected utility theory provides a way of ranking the acts according to how choiceworthy they are: the higher the expected utility, the better it is to choose the act. (It is therefore best to choose the act with the highest expected utility—or one of them, in the event that several acts are tied.) Following general convention, I will make the following assumptions about the relationships between acts, states, and outcomes. States, acts, and outcomes are propositions, i.e., sets of possibilities. There is a maximal set of possibilities, \(\Omega\), of which each state, act, or outcome is a subset. The set of acts, the set of states, and the set of outcomes are all partitions on \(\Omega\). In other words, acts and states are individuated so that every possibility in \(\Omega\) is one where exactly one state obtains, the agent performs exactly one act, and exactly one outcome ensues. Acts and states are logically independent, so that no state rules out the performance of any act. I will assume for the moment that, given a state of the world, each act has exactly one possible outcome. (Section 1.1 briefly discusses how one might weaken this assumption.) So the example of the umbrella can be depicted in the following matrix, where each column corresponds to a state of the world; each row corresponds to an act; and each entry corresponds to the outcome that results when the act is performed in the state of the world. states it rains it does not rain acts take umbrella encumbered, dry encumbered, dry leave umbrella wet free, dry Having set up the basic framework, I can now rigorously define expected utility. The expected utility of an act \(A\) (for instance, taking my umbrella) depends on two features of the problem: The value of each outcome, measured by a real number called a utility. The probability of each outcome conditional on \(A\). Given these three pieces of information, \(A\)'s expected utility is defined as: \[EU(A) = \sum_{o \in O} P_{A}(o) U(o) \] where \(O\) is is the set of outcomes, \(P_{A}(o)\) is the probability of outcome \(o\) conditional on \(A\), and \(U(o)\) is the utility of \(o\). The next two subsections will unpack the conditional probability function \(P_A\) and the utility function \(U\). 1.1 Conditional Probabilities The term \(P_{A}(o)\) represents the probability of \(o\) given \(A\)—roughly, how likely it is that outcome \(o\) will occur, on the supposition that the agent chooses act \(A\). (For the axioms of probability, see the entry on interpretations of probability.) To understand what this means, we must answer two questions. First, which interpretation of probability is appropriate? And second, what does it mean to assign a probability on the supposition that the agent chooses act \(A\)? Expected utility theorists often interpret probability as measuring individual degree of belief, so that a proposition \(E\) is likely (for an agent) to the extent that that agent is confident of \(E\) (see, for instance, Ramsey 1926, Savage 1972, Jeffrey 1983). But nothing in the formalism of expected utility theory forces this interpretation on us. We could instead interpret probabilities as objective chances (as in von Neumann and Morgenstern 1944), or as the degrees of belief that are warranted by the evidence, if we thought these were a better guide to rational action. (See the entry on interpretations of probability for discussion of these and other options.) What is it to have a probability on the supposition that the agent chooses \(A\)? Here, there are two basic types of answer, corresponding to evidential decision theory and causal decision theory. According to evidential decision theory, endorsed by Jeffrey (1983), the relevant suppositional probability \(P_{A}(o)\) is the conditional probability \(P(o \mid A)\), defined as the ratio of two unconditional probabilities: \(P(A \amp o) / P(A)\). Against Jeffrey's definition of expected utility, Spohn (1977) and Levi (1991) object that a decision-maker should not assign probabilities to the very acts under deliberation: when freely deciding whether to perform an act \(A\), you shouldn't take into account your beliefs about whether you will perform \(A\). If Spohn and Levi are right, then Jeffrey's ratio is undefined (since its denominator is undefined). Nozick (1969) raises another objection: Jeffrey's definition gives strange results in the Newcomb Problem. A predictor hands you a closed box, containing either $0 or $1 million, and offers you an open box, containing an additional $1,000. You can either refuse the open box ("one-box") or take the open box ("two-box"). But there's a catch: the predictor has predicted your choice beforehand, and all her predictions are 90% accurate. In other words, the probability that you one-box, given that she predicts you one-box, is 90%, and the probability that you two-box, given that she predicts you two-box, is 90%. Finally, the contents of the closed box depend on the prediction: if the predictor thought you would two-box, she put nothing in the closed box, while if she thought you would one-box, she put $1 million in the closed box. The matrix for your decision looks like this: states $1 million in closed box $0 in closed box acts one-box $1,000,000 $0 two-box $1,001,000 $1,000 Two-boxing dominates one-boxing: in every state, two-boxing yields a better outcome. Yet on Jeffrey's definition of conditional probability, one-boxing has a higher expected utility than two-boxing. There is a high conditional probability of finding $1 million is in the closed box, given that you one-box, so one-boxing has a high expected utility. Likewise, there is a high conditional probability of finding nothing in the closed box, given that you two-box, so two-boxing has a low expected utility. Causal decision theory is an alternative proposal that gets around these problems. It does not require (but still permits) acts to have probabilities, and it recommends two-boxing in the Newcomb problem. Causal decision theory comes in many varieties, but I'll consider a representative version proposed by Savage (1972), which calculates \(P_{A}(o)\) by summing the probabilities of states that, when combined with the act \(A\), lead to the outcome \(o\). Let \(f_{A,s}(o)\) be a of outcomes, which maps \(o\) to 1 if \(o\) results from performing \(A\) in state s, maps \(o\) to 0 otherwise. Then \[P_{A}(o) = \sum_{s\in S}P(s)f_{A,s}(o) \] On Savage's proposal, two-boxing comes out with a higher expected utility than one-boxing. This result holds no matter which probabilities you assign to the states prior to your decision. Let \(x\) be the probability you assign to the state that the closed box contains $1 million. According to Savage, the expected utilities of one-boxing and two-boxing, respectively, are: \[x {\cdot} U({$1,000,000}) + (1 - x) {\cdot} U($0) \] and \[x {\cdot} U({$1,001,000}) + (1 - x) {\cdot} U({$1,000}) \] As long as the larger monetary amounts are assigned strictly larger utilities, the second sum (the utility of two-boxing) is guaranteed to be larger than the first (the utility of one-boxing). Savage assumes that each act and state are enough to uniquely determine an outcome. But there are cases where this assumption breaks down. Suppose you offer to sell me the following gamble: you will toss a coin; if the coin lands heads, I win $100; and if the coin lands tails, I lose $100. But I refuse the gamble, and the coin is never tossed. There is no outcome that would have resulted, had the coin been tossed—I might have won $100, and I might have lost $100. We can generalze Savage's proposal by letting \(f_{A,s}\) be a probability function that maps outcomes to real numbers in the \([0, 1]\) interval. Lewis (1981), Skyrms (1980), and Sobel (1994) equate \(f_{A,s}\) with the objective chance that \(o\) would be the outcome if state \(s\) obtained and the agent chose action \(A\). In some cases—most famously the Newcomb problem—the Jeffrey definition and the Savage definition of expected utility come apart. But whenever the following two conditions are satisfied, they agree. Acts are probabilistically independent of states. In formal terms, for all acts \(A\) and states \(s\), \[P(s) = P(s \mid A) = \frac{P(s \amp A)}{P(A)}. \] (This is the condition that is violated in the Newcomb problem.) For all outcomes \(o\), acts \(A\), and states \(s\), \(f_{A,s}(o)\) is equal to the conditional probability of \(o\) given \(A\) and \(s\); in formal terms, \[f_{A,s}(o) = P(o \mid A \amp s) = \frac{P(o \amp A \amp s)}{P(A \amp s)}.\] (The need for this condition arises when acts and states fail to uniquely determine an outcome; see Lewis 1981.) 1.2 Outcome Utilities The term \(U(o)\) represents the utility of the outcome \(o\)—roughly, how valuable \(o\) is. Formally, \(U\) is a function that assigns a real number to each of the outcomes. (The units associated with \(U\) are typically called utiles, so that if \(U(o) = 2\), we say that \(o\) is worth 2 utiles.) The greater the utility, the more valuable the outcome. What kind of value is measured in utiles? Utiles are typically not taken to be units of currency, like dollars, pounds, or yen. Bernoulli (1738) argued that money and other goods have diminishing marginal utility: as an agent gets richer, every successive dollar (or gold watch, or apple) is less valuable to her than the last. He gives the following example: It makes rational sense for a rich man, but not for a pauper, to pay 9,000 ducats in exchange for a lottery ticket that yields a 50% chance at 20,000 ducats and a 50% chance at nothing. Since the lottery gives the two men the same chance at each monetary prize, the prizes must have different values depending on whether the player is poor or rich. Classic utilitarians such as Bentham (1789), Mill (1861), and Sidgwick (1907) interpreted utility as a measure of pleasure or happiness. For these authors, to say \(A\) has greater utility than \(B\) (for an agent or a group of agents) is to say that \(A\) results in more pleasure or happiness than \(B\) (for that agent or group of agents). One objection to this interpretation of utility is that there may not be a single good (or indeed any good) which rationality requires us to seek. But if we understand "utility" broadly enough to include all potentially desirable ends—pleasure, knowledge, friendship, health and so on—it's not clear that there is a unique correct way to make the tradeoffs between different goods so that each outcome receives a utility. There may be no good answer to the question of whether the life of an ascetic monk contains more or less good than the life of a happy libertine—but assigning utilities to these options forces us to compare them. Contemporary decision theorists typically interpret utility as a measure of preference, so that to say that \(A\) has greater utility than \(B\) (for an agent) is simply to say that the agent prefers \(A\) to \(B\). It is crucial to this approach that preferences hold not just between outcomes (such as amounts of pleasure, or combinations of pleasure and knowledge), but also between uncertain prospects (such as a lottery that pays $1 million dollars if a particular coin lands heads, and results in an hour of painful electric shocks if the coin lands tails). Section 2 of this article addresses the formal relationship between preference and choice in detail. Expected utility theory does not require that preferences be selfish or self-interested. Someone can prefer giving money to charity over spending the money on lavish dinners, or prefer sacrificing his own life over allowing his child to die. Sen (1977) suggests that each person's psychology is best represented using three rankings: one representing the person's narrow self-interest, a second representing the person's self-interest construed more broadly to account for feelings of sympathy (e.g., suffering when watching another person suffer), and a third representing the person's commitments, which may require her to act against her self-interest broadly construed. Broome (1991) interprets utilities as measuring comparisons of objective betterness and worseness, rather than personal preferences: to say that \(A\) has a greater utility than \(B\) is to say that \(A\) is objectively better than \(B\), or that a rational person would prefer \(A\) to \(B\). Just as there is nothing in the formalism of probability theory that requires us to use subjective rather than objective probabilities, so there is nothing in the formalism of expected utility theory that requires us to use subjective rather than objective values. Those who interpret utilities in terms of personal preference face a special challenge: the so-called problem of interpersonal utility comparisons. When making decisions about how to distribute shared resources, we often want to know if our acts would make Alice better off than Bob—and if so, how much better off. But if utility is a measure of individual preference, there is no clear, meaningful way of making these comparisons. Alice's utilities are constituted by Alice's preferences, Bob's utilities are constituted by Bob's preferences, and there are no preferences spanning Alice and Bob. We can't assume that Alice's utility 10 is equivalent to Bob's utility 10, any more than we can assume that getting an A grade in differential equations is equivalent to getting an A grade in basket weaving. Now is a good time to consider which features of the utility function carry meaningful information. Comparisons are informative: if \(U(o_1) \gt U(o_2)\) (for a person), then \(o_1\) is better than (or preferred to) \(o_2\). But it is not only comparisons that are informative—the utility function must carry other information, if expected utility theory is to give meaningful results. To see why, consider the umbrella example again. This time, I've filled in a probability for each state, and a utility for each outcome. states it rains \((P = 0.6)\) it does not rain \((P = 0.4)\) acts take umbrella encumbered, dry \((U = 5)\) encumbered, dry \((U = 5)\) leave umbrella wet \((U = 0)\) free, dry \((U =10)\) The expected utility of taking the umbrella is \[\begin{align} EU(\take)&= P_{\take}(\encumbered, \dry) \cdot 5 \\ & \quad + P_{\take}(\wet) \cdot 0 \\ & \quad + P_{\take}(\free, dry) \cdot 10 \\ &=5 \end{align} \] while the expected utility of leaving the umbrella is \[\begin{align} EU(\leave)&= P_{\leave}(\encumbered, \dry) \cdot 5 \\ & \quad + P_{\leave}(\wet) \cdot 0 \\ & \quad + P_{\leave}(\free, dry) \cdot 10 \\ &=4 \end{align} \] Since \(EU(\take) \gt EU(\leave)\), expected utility theory tells me that taking the umbrella is better than leaving it. But now, suppose we change the utilities of the outcomes: instead of using \(U\), we use \(U'\). states it rains \((P=0.6)\) it does not rain \((P=0.4)\) acts take umbrella encumbered, dry \((U'=4)\) encumbered, dry \((U'=4)\) leave umbrella wet \((U'=2)\) free, dry \((U'=8)\) The new expected utility of taking the umbrella is \[\begin{align} EU'(\take)&= P_{\take}(\encumbered, \dry) \cdot 4 \\ & \quad + P_{\take}(\wet) \cdot 2 \\ & \quad + P_{\take}(\free, dry) \cdot 8 \\ &= 4 \end{align} \] while the new expected utility of leaving the umbrella is \[\begin{align} EU'(\leave)&= P_{\leave}(\encumbered, \dry) \cdot 4 \\ & \quad + P_{\leave}(\wet) \cdot 2 \\ & \quad + P_{\leave}(\free, dry) \cdot 8 \\ &= 4.4 \end{align} \] Since \(EU'(\take) \lt EU'(\leave)\), expected utility theory tells me that leaving the umbrella is better than taking it. The utility functions \(U\) and \(U'\) rank the outcomes in exactly the same way: free, dry is best; encumbered, dry ranks in the middle; and wet is worst. Yet expected utility theory gives different advice in the two versions of the problem. So there must be some substantive difference between preferences appropriately described by \(U\), and preferences appropriately described by \(U'\). Otherwise, expected utility theory is fickle, and liable to change its advice when fed different descriptions of the same problem. When do two utility functions represent the same basic state of affairs? Measurement theory answers the question by characterizing the allowable transformations of a utility function—ways of changing it that leave all of its meaningful features intact. If we characterize the allowable transformations of a utility function, we have thereby specified which of its features are meaningful. Defenders of expected utility theory typically require that utility be measured by a linear scale, where the allowable transformations are all and only the positive linear transformations, i.e., functions \(f\) of the form \[f(U(o))=x {\cdot} U(o)+y \] for real numbers \(x \gt 0\) and \(y\). Positive linear transformations of outcome utilities will never affect the verdicts of expected utility theory: if \(A\) has greater expected utility than \(B\) where utility is measured by function \(U\), then \(A\) will also have greater expected utility than \(B\) where utility is measured by any positive linear transformation of \(U\). 2. Arguments for Expected Utility Theory Why choose acts that maximize expected utility? One possible answer is that expected utility theory is rational bedrock—that means-end rationality essentially involves maximizing expected utility. For those who find this answer unsatisfying, however, there are two further sources of justification. First, there are long-run arguments, which rely on evidence that expected-utility maximization is a profitable policy in the long term. Second, there are arguments based on representation theorems, which suggest that certain rational constraints on preference entail that all rational agents maximize expected utility. 2.1 Long-Run Arguments One reason for maximizing expected utility is that it makes for good policy in the long run. Feller (1968) gives a version of this argument. He relies on two mathematical facts about probabilities: the strong and weak laws of large numbers. Both these facts concern sequences of independent, identically distributed trials—the sort of setup that results from repeatedly betting the same way on a sequence of roulette spins or craps games. Both the weak and strong laws of large numbers say, roughly, that over the long run, the average amount of utility gained per trial is overwhelmingly likely to be close to the expected value of an individual trial. The weak law of large numbers states that where each trial has an expected value of \(\mu\), for any arbitrarily small real numbers \(\epsilon \gt 0\) and \(\delta \gt 0\), there is some finite number of trials \(n\), such that for all \(m\) greater than or equal to \(n\), with probability at least \(1-\delta\), the gambler's average gains for the first \(m\) trials will fall within \(\epsilon\) of \(\mu\). In other words, in a long run of similar gamble, the average gain per trial is highly likely to become arbitrarily close to the gamble's expected value within a finite amount of time. So in the finite long run, the average value associated with a gamble is overwhelmingly likely to be close to its expected value. The strong law of large numbers states that where each trial has an expected value of \(\mu\), for any arbitrarily small real number \(\epsilon \gt 0\), as the number of trials increases, the probability that the gambler's average winnings per trial fall within \(\epsilon\) of \(\mu\) converges to 1. In other words, as the number of repetitions of a gamble approaches infinity, the average gain per trial will become arbitrarily close to the gamble's expected value with probability 1. So in the long run, the average value associated with a gamble is virtually certain to equal its expected value. There are several objections to these long run arguments. First, many decisions cannot be repeated over indefinitely many similar trials. Decisions about which career to pursue, whom to marry, and where to live, for instance, are made at best a small finite number of times. Furthermore, where these decisions are made more than once, different trials involve different possible outcomes, with different probabilities. It is not clear why long-run considerations about repeated gambles should bear on these single-case choices. Second, the argument relies on two independence assumptions, one or both of which may fail. One assumption holds that the probabilities of the different trials are independent. This is true of casino gambles, but not true of other choices where we wish to use decision theory—e.g., choices about medical treatment. My remaining sick after one course of antibiotics makes it more likely I will remain sick after the next course, since it increases the chance that antibiotic-resistant bacteria will spread through my body. The argument also requires that the utilities of different trials be independent, so that winning a prize on one trial makes the same contribution to the decision-maker's overall utility no matter what she wins on other trials. But this assumption is violated in many real-world cases. Due to the diminishing marginal utility of money, winning $10 million on ten games of roulette is not worth ten times as much as winning $1 million on one game of roulette. A third problem is that the strong and weak laws of large numbers are modally weak. Neither law entails that if a gamble were repeated indefinitely (under the appropriate assumptions), the average utility gain per trial would be close to the game's expected utility. They establish only that the average utility gain per trial would with high probability be close to the game's expected utility. But high probability—even probability 1—is not certainty. (Standard probability theory rejects Cournot's Principle, which says events with low or zero probability will not happen. But see Shafer (2005) for a defense of Cournot's Principle.) For any sequence of independent, identically distributed trials, it is possible for the average utility payoff per trial to diverge arbitrarily far from the expected utility of an individual trial. 2.2 Representation Theorems A second type of argument for expected utility theory relies on so-called representation theorems. We follow Zynda's (2000) formulation of this argument—slightly modified to reflect the role of utilities as well as probabilities. The argument has three premises: The Rationality Condition. The axioms of expected utility theory are the axioms of rational preference. Representability. If a person's preferences obey the axioms of expected utility theory, then she can be represented as having degrees of belief that obey the laws of the probability calculus [and a utility function such that she prefers acts with higher expected utility]. The Reality Condition. If a person can be represented as having degrees of belief that obey the probability calculus [and a utility function such that she prefers acts with higher expected utility], then the person really has degrees of belief that obey the laws of the probability calculus [and really does prefer acts with higher expected utility]. These premises entail the following conclusion. If a person [fails to prefer acts with higher expected utility], then that person violates at least one of the axioms of rational preference. If the premises are true, the argument shows that there is something wrong with people whose preferences are at odds with expected utility theory—they violate the axioms of rational preference. Let us consider each of the premises in greater detail, beginning with the key premise, Representability. A probability function and a utility function together represent a set of preferences just in case the following formula holds for all values of \(A\) and \(B\) in the domain of the preference relation \[EU(A) \gt EU(B) \text{ if and only if } A \text{ is preferred to } B. \] Mathematical proofs of Representability are called representation theorems. Section 2.1 surveys three of the most influential representation theorems, each of which relies on a different set of axioms. No matter which set of axioms we use, the Rationality Condition is controversial. In some cases, preferences that seem rationally permissible—perhaps even rationally required—violate the axioms of expected utility theory. Section 3 discusses such cases in detail. The Reality Condition is also controversial. Hampton (1994), Zynda (2000), and Meacham and Weisberg (2011) all point out that to be representable using a probability and utility function is not to have a probability and utility function. After all, an agent who can be represented as an expected utility maximizer with degrees of belief that obey the probability calculus, can also be represented as someone who fails to maximize expected utility with degrees of belief that violate the probability calculus. Why think the expected utility representation is the right one? There are several options. Perhaps the defender of representation theorems can stipulate that what it is to have particular degrees of belief and utilities is just to have the corresponding preferences. The main challenge for defenders of this response is to explain why representations in terms of expected utility are explanatorily useful, and why they are better than alternative representations. Or perhaps probabilities and utilities are a good cleaned-up theoretical substitutes for our folk notions of belief and desire—precise scientific substitutes for our folk concepts. Meacham and Weisberg challenge this response, arguing that probabilities and utilities are poor stand-ins for our folk notions. A third possibility, suggested by Zynda, is that facts about degrees of belief are made true independently of the agent's preferences, and provide a principled way to restrict the range of acceptable representations. The challenge for defenders of this type of response is to specify what these additional facts are. I now turn to consider three influential representation theorems. These representation theorems differ from each other in three of philosophically significant ways. First, different representation theorems disagree about the objects of preference and utility. Are they repeatable? Must they be wholly within the agent's control Second, representation theorems differ in their treatment of probability. They disagree about which entities have probabilities, and about whether the same objects can have both probabilities and utilities. Third, while every representation theorem proves that for a suitable preference ordering, there exist a probability and utility function representing the preference ordering, they differ how unique this probability and utility function are. In other words, they differ as to which transformations of the probability and utility functions are allowable. 2.2.1 Ramsey The idea of a representation theorem for expected utility dates back to Ramsey (1926). (His sketch of a representation theorem is subsequently filled in by Bradley (2004) and Elliott (2017).) Ramsey assumes that preferences are defined over a domain of gambles, which yield one prize on the condition that a proposition \(P\) is true, and a different prize on the condition that \(P\) is false. (Examples of gambles: you receive a onesie if you're having a baby and a bottle of scotch otherwise; you receive twenty dollars if Bojack wins the Kentucky Derby and lose a dollar otherwise.) Ramsey calls a proposition ethically neutral when "two possible worlds differing only in regard to [its truth] are always of equal value". For an ethically neutral proposition, probability 1/2 can be defined in terms of preference: such a proposition has probability 1/2 just in case you are indifferent as to which side of it you bet on. (So if Bojack wins the Kentucky Derby is an ethically neutral proposition, it has probability 1/2 just in case you are indifferent between winning twenty dollars if it's true and losing a dollar otherwise, and winning twenty dollars if it's false and losing a dollar otherwise.) By positing an ethically neutral proposition with probability 1/2, together with a rich space of prizes, Ramsey defines numerical utilities for prizes. (The rough idea is that if you are indifferent between receiving a middling prize \(m\) for certain, and a gamble that yields a better prize \(b\) if the ethically neutral proposition is true and a worse prize \(w\) if it is falls, then the utility of \(m\) is halfway between the utilities of \(b\) and \(w\).) Using these numerical utilities, he then exploits the definition of expected utility to define probabilities for all other propositions. The rough idea is to exploit the richness of the space of prizes, which ensures that for any gamble \(g\) that yields better prize \(b\) if \(E\) is true and worse prize \(w\) if \(E\) is false, the agent is indifferent between \(g\) and some middling prize \(m\). This means that \(EU(g) = EU(m)\). Using some algebra, plus the fact that \(EU(g) = P(E)U(b) + (1-P(E))U(w)\), Ramsey shows that \[P(E) = \frac{(1 - U(m)}{(U(b) - U(w))} \] 2.2.2 Von Neumann and Morgenstern Von Neumann and Morgenstern (1944) claim that preferences are defined over a domain of lotteries. Some of these lotteries are constant, and yield a single prize with certainty. (Prizes might include a banana, a million dollars, a million dollars' worth of debt, death, or a new car.) Lotteries can also have other lotteries as prizes, so that one can have a lottery with a 40% chance of yielding a banana, and a 60% chance of yielding a 50-50 gamble between a million dollars and death.) The domain of lotteries is closed under a mixing operation, so that if \(L\) and \(L'\) are lotteries and \(x\) is a real number in the \([0, 1]\) interval, then there is a lottery \(x L + (1-x) L'\) that yields \(L\) with probability \(x\) and \(L'\) with probability \(1-x\). They show that every preference relation obeying certain axioms can be represented by the probabilities used to define the lotteries, together with a utility function which is unique up to positive linear transformation. 2.2.3 Savage Instead of taking probabilities for granted, as von Neumann and Morgenstern do, Savage (1972) defines them in terms of preferences over acts. Savage posits three separate domains. Probability attaches to events, which we can think of as disjunctions of states, while utility and intrinsic preference attach to outcomes. Expected utility and non-intrinsic preference attach to acts. For Savage, acts, states, and outcomes must satisfy certain constraints. Acts must be wholly under the agent's control (so publishing my paper in Mind is not an act, since it depends partly on the editor's decision, which I do not control). Outcomes must have the same utility regardless of which state obtains (so "I win a fancy car" is not an outcome, since the utility of the fancy car will be greater in states where the person I most want to impress wishes I had a fancy car, and less in states where I lose my driver's license). No state can rule out the performance of any act, and an act and a state together must determine an outcome with certainty. For each outcome \(o\), there is a constant act which yields \(o\) in every state. (Thus, if world peace is an outcome, there is an act that results in world peace, no matter what the state of the world.) Finally, he assumes for any two acts \(A\) and \(B\) and any event \(E\), there is a mixed act \(A_E \amp B_{\sim E}\) that yields the same outcome as \(A\) if \(E\) is true, and the same outcome as \(B\) otherwise. (Thus, if world peace and the end of the world are both outcomes, then there is a mixed act that results in world peace if a certain coin lands heads, and the end of the world otherwise.) Savage postulates a preference relation over acts, and gives axioms governing that preference relation. He then defines subjective probabilities, or degrees of belief, in terms of preferences. The key move is to define an "at least as likely as" relation between events; I paraphrase here. Suppose \(A\) and \(B\) are constant acts such that \(A\) is preferred to \(B\). Then \(E\) is at least as likely as \(F\) just in case the agent either prefers \(A_E \amp B_{\sim E}\) (the act that yields \(A\) if \(E\) obtains, and \(B\) otherwise) to \(A_F \amp B_{\sim F}\) (the act that yields \(A\) if \(F\) obtains, and \(B\) otherwise), or else is indifferent between \(A_E \amp B_{\sim E}\) and \(A_F \amp B_{\sim F}\). The thought behind the definition is that the agent considers \(E\) at least as likely as \(F\) just in case she would not rather bet on \(F\) than on \(E\)). Savage then gives axioms constraining rational preference, and shows that any set of preferences satisfying those axioms yields an "at least as likely" relation that can be uniquely represented by a probability function. In other words, there is one and only one probability function \(P\) such that for all \(E\) and \(F\), \(P(E) \ge P(F)\) if and only if \(E\) is at least as likely as \(F\). Every preference relation obeying Savage's axioms is represented by this probability function \(P\), together with a utility function which is unique up to positive linear transformation. Savage's representation theorem gives strong results: starting with a preference ordering alone, we can find a single probability function, and a narrow class of utility functions, which represent that preference ordering. The downside, however, is that Savage has to build implausibly strong assumptions about the domain of acts. Luce and Suppes (1965) point out that Savage's constant acts are implausible. (Recall that constant acts yield the same outcome and the same amount of value in every state.) Take some very good outcome—total bliss for everyone. Is there really a constant act that has this outcome in every possible state, including states where the human race is wiped out by a meteor? Savage's reliance on a rich space of mixed acts is also problematic. Savage has had to assume that any two outcomes and any event, there is a mixed act that yields the first outcome if the event occurs, and the second outcome otherwise? Is there really an act that yields total bliss if everyone is killed by an antibiotic-resistant plague, and total misery otherwise? Luce and Krantz (1971) suggest ways of reformulating Savage's representation theorem that weaken these assumptions, but Joyce (1999) argues that even on the weakened assumptions, the domain of acts remains implausibly rich. 2.2.4 Bolker and Jeffrey Bolker (1966) proves a general representation theorem about mathematical expectations, which Jeffrey (1983) uses as the basis for a philosophical account of expected utility theory. Bolker's theorem assumes a single domain of propositions, which are objects of preference, utility, and probability alike. Thus, the proposition that it will rain today has a utility, as well as a probability. Jeffrey interprets this utility as the proposition's news value—a measure of how happy or disappointed I would be to learn that the proposition was true. By convention, he sets the value of the necessary proposition at 0—the necessary proposition is no news at all! Likewise, the proposition that I take my umbrella to work, which is an act, has a probability as well as a utility. Jeffrey interprets this to mean that I have degrees of belief about what I will do. Bolker gives axioms constraining preference, and shows that any preferences satisfying his axioms can be represented by a probability measure \(P\) and a utility measure \(U\). However, Bolker's axioms do not ensure that \(P\) is unique, or that \(U\) is unique up to positive linear transformation. Nor do they allow us to define comparative probability in terms of preference. Instead, where \(P\) and \(U\) jointly represent a preference ordering, Bolker shows that the pair \(\langle P, U \rangle\) is unique up to a fractional linear transformation. In technical terms, where \(U\) is a utility function normalized so that \(U(\Omega) = 0\), \(\inf\) is the greatest lower bound of the values assigned by \(U\), \(\sup\) is the least upper bound of the values assigned by by \(U\), and \(\lambda\) is a parameter falling between \(-1/\inf\) and \(-1/sup\), the fractional linear transformation \(\langle P_{\lambda}, U_{\lambda} \rangle\) of \(\langle P, U \rangle\) corresponding to \(\lambda\) is given by: \[\begin{align} P_{\lambda} &= P(x)(1 + \lambda U(x)) \\ U_{\lambda} &= U(x)((1+\lambda)/(1 + \lambda U(x)) \end{align} \] Notice that fractional linear transformations of a probability-utility pair can disagree with the original pair about which propositions are likelier than which others. Joyce (1999) shows that with additional resources, Bolker's theorem can be modified to pin down a unique \(P\), and a \(U\) that is unique up to positive linear transformation. We need only supplement the preference ordering with a primitive "more likely than" relation, governed by its own set of axioms, and linked to belief by several additional axioms. Joyce modifies Bolker's result to show that given these additional axioms, the "more likely than" relation is represented by a unique \(P\), and the preference ordering is represented by \(P\) together with a utility function that is unique up to positive linear transformation. 2.2.5 Summary Together, these four representation theorems above can be summed up in the following table. Theorem Objects ofpreference Order ofconstruction Allowabletransformations:probability Allowabletransformations: utility Ramsey gambles preference → utility → probability identity positive linear von Neumann/Morgenstern lotteries (preference & probability) → utility N/A positive linear Savage acts preference → probability → utility identity positive linear Jeffrey/Bolker propositions preference → (probability & utility) — fractional linear — Notice that the order of construction differs between theorems: Ramsey constructs a representation of probability using utility, while von Neumann and Morgenstern begin with probabilities and construct a representation of utility. Thus, although the arrows represent a mathematical relationship of representation, they cannot represent a metaphysical relationship of grounding. The Reality Condition needs to be justified independently of any representation theorem. Suitably structured ordinal probabilities (the relations picked out by "at least as likely as", "more likely than", and "equally likely") stand in one-to-one correspondence with the cardinal probability functions. Finally, the grey line from preferences to ordinal probabilities indicates that every probability function satisfying Savage's axioms is represented by a unique cardinal probability—but this result does not hold for Jeffrey's axioms. Notice that it is often possible to follow the arrows in circles—from preference to ordinal probability, from ordinal probability to cardinal probability, from cardinal probability and preference to expected utility, and from expected utility back to preference. Thus, although the arrows represent a mathematical relationship of representation, they do not represent a metaphysical relationship of grounding. This fact drives home the importance of independently justifying the Reality Condition—representation theorems cannot justify expected utility theory without additional assumptions. 3. Obections to Expected Utility Theory 3.1 Maximizing Expected Utility is Impossible Ought implies can, but is it humanly possible to maximize expected utility? March and Simon (1958) point out that in order to compute expected utilities, an agent needs a dauntingly complex understanding of the available acts, the possible outcomes, and the values of those outcomes, and that choosing the best act is much more demanding than choosing an act that is merely good enough. Similar points appear in Lindblom (1959), Feldman (2006), and Smith (2010). McGee (1991) argues that maximizing expected utility is not mathematically possible even for an ideal computer with limitless memory. In order to maximize expected utility, we would have to accept any bet we were offered on the truths of arithmetic, and reject any bet we were offered on false sentences in the language of arithmetic. But arithmetic is undecidable, so no Turing machine can determine whether a given arithmetical sentence is true or false. One response to these difficulties is the bounded rationality approach, which aims to replace expected utility theory with some more tractable rules. Another is to argue that the demands of expected utility theory are more tractable than they appear (Burch-Brown 2014; see also Greaves 2016), or that the relevant "ought implies can" principle is false (Srinivasan 2015). 3.2 Maximizing Expected Utility is Irrational A variety of authors have given examples in which expected utility theory seems to give the wrong prescriptions. Sections 3.2.1 and 3.2.2 discuss examples where rationality seems to permit preferences inconsistent with expected utility theory. These examples suggest that maximizing expected utility is not necessary for rationality. Section 3.2.3 discusses examples where expected utility theory permits preferences that seem irrational. These examples suggest that maximizing expected utility is not sufficient for rationality. Section 3.2.4 discusses an example where expected utility theory requires preferences that seem rationally forbidden—a challenge to both the necessity and the sufficiency of expected utility for rationality. 3.2.1 Counterexamples Involving Transitivity and Completeness Expected utility theory implies that the structure of preferences mirrors the structure of the greater-than relation between real numbers. Thus, according to expected utility theory, preferences must be transitive: If \(A\) is preferred to \(B\) (so that \(U(A) \gt U(B)\)), and \(B\) is preferred to \(C\) (so that \(U(B) \gt U(C)\)), then \(A\) must be preferred to \(C\) (since it must be that \(U(A) \gt U(C)\)). Likewise, preferences must be complete: for any two options, either one must be preferred to the other, or the agent must be indifferent between them (since of their two utilities, either one must be greater or the two must be equal). But there are cases where rationality seems to permit (or perhaps even require) failures of transitivity and failures of completeness. An example of preferences that are not transitive, but nonetheless seem rationally permissible, is Quinn's puzzle of the self-torturer (1990). The self-torturer is hooked up to a machine with a dial with settings labeled 0 to 1,000, where setting 0 does nothing, and each successive setting delivers a slightly more powerful electric shock. Setting 0 is painless, while setting 1,000 causes excruciating agony, but the difference between any two adjacent settings is so small as to be imperceptible. The dial is fitted with a ratchet, so that it can be turned up but never down. Suppose that at each setting, the self-torturer is offered $10,000 to move up to the next, so that for tolerating setting \(n\), he receives a payoff of \(n {\cdot} {$10,000}\). It is permissible for the self-torturer to prefer setting \(n+1\) to setting \(n\) for each \(n\) between 0 and 999 (since the difference in pain is imperceptible, while the difference in monetary payoffs is significant), but not to prefer setting 1,000 to setting 0 (since the pain of setting 1,000 may be so unbearable that no amount of money will make up for it. It also seems rationally permissible to have incomplete preferences. For some pairs of actions, an agent may have no considered view about which she prefers. Consider Jane, an electrician who has never given much thought to becoming a professional singer or a professional astronaut. (Perhaps both of these options are infeasible, or perhaps she considers both of them much worse than her steady job as an electrician). It is false that Jane prefers becoming a singer to becoming an astronaut, and it is false that she prefers becoming an astronaut to becoming a singer. But it is also false that she is indifferent between becoming a singer and becoming an astronaut. She prefers becoming a singer and receiving a $100 bonus to becoming a singer, and if she were indifferent between becoming a singer and becoming an astronaut, she would be rationally compelled to prefer being a singer and receiving a $100 bonus to becoming an astronaut. There is one key difference between the two examples considered above. Jane's preferences can be extended, by adding new preferences without removing any of the ones she has, in a way that lets us represent her as an expected utility maximizer. On the other hand, there is no way of extended the self-torturer's preferences so that he can be represented as an expected utility maximizer. Some of his preferences would have to be altered. One popular response to incomplete preferences is to claim that, while rational preferences need not satisfy the axioms of a given representation theorem (see section 2.2), it must be possible to extend them so that they satisfy the axioms. From this weaker requirement on preferences—that they be extendible to a preference ordering that satisfies the relevant axioms—one can prove the existence halves of the relevant representation theorems. However, one can no longer establish that each preference ordering has a representation which is unique up to allowable transformations. No such response is available in the case of the self-torturer, whose preferences cannot be extended to satisfy the axioms of expected utility theory. See the entry on preferences for a more extended discussion of the self-torturer case. 3.2.2 Counterexamples Involving Independence Allais (1953) and Ellsberg (1961) propose examples of preferences that cannot be represented by an expected utility function, but that nonetheless seem rational. Both examples involve violations of Savage's Independence axiom: Independence. Suppose that \(A\) and \(A^*\) are two acts that produce the same outcomes in the event that \(E\) is false. Then, for any act \(B\), one must have \(A\) is preferred to \(A^*\) if and only if \(A_E \amp B_{\sim E}\) is preferred to \(A^*_E \amp B_{\sim E}\) The agent is indifferent between \(A\) and \(A^*\) if and only if she is indifferent between \(A_E \amp B_{\sim E}\) and \(A^*_E \amp B_{\sim E}\) In other words, if two acts have the same consequences whenever \(E\) is false, then the agent's preferences between those two acts should depend only on their consequences when \(E\) is true. On Savage's definition of expected utility, expected utility theory entails Independence. And on Jeffrey's definition, expected utility theory entails Independence in the presence of the assumption that the states are probabilistically independent of the acts. The first counterexample, the Allais Paradox, involves two separate decision problems in which a ticket with a number between 1 and 100 is drawn at random. In the first problem, the agent must choose between these two lotteries: In the second decision problem, the agent must choose between these two lotteries: It seems reasonable to prefer \(A\) (which offers a sure $100 million) to \(B\) (where the added 10% chance at $500 million is more than offset by the risk of getting nothing). It also seems reasonable to prefer \(D\) (an 10% chance at a $500 million prize) to \(C\) (a slightly larger 11% chance at a much smaller $100 million prize). But together, these preferences (call them the Allais preferences) violate Independence. Lotteries \(A\) and \(C\) yield the same $100 million prize for tickets 12–100. They can be converted into lotteries \(B\) and \(D\) by replacing this $100 million prize with $0. Because they violate Independence, the Allais preferences are incompatible with expected utility theory. This incompatibility does not require any assumptions about the relative utilities of the $0, the $100 million, and the $500 million. Where $500 million has utility \(x\), $100 million has utility \(y\), and $0 has utility \(z\), the expected utilities of the lotteries are as follows. \[\begin{align} EU(A) &= 0.11y + 0.89y \\ EU(B) &= 0.10x + 0.01z + 0.89y \\ EU(C) &= 0.11y + 0.89z \\ EU(D) &= 0.10x + 0.01z + 0.89z \end{align} \] It is easy to see that the condition under which \(EU(A) \gt EU(B)\) is exactly the same as the condition under which \(EU(C) \gt EU(D)\): both inequalities obtain just in case \(0.11y \gt 0.10x + 0.01z\) The Ellsberg Paradox also involves two decision problems that generate a violation of the sure-thing principle. In each of them, a ball is drawn from an urn containing 30 red balls, and 60 balls that are either white or yellow in unknown proportions. In the first decision problem, the agent must choose between the following lotteries: In the second decision problem, the agent must choose between the following lotteries: It seems reasonable to prefer \(R\) to \(W\), but at the same time prefer \(WY\) to \(RY\). (Call this combination of preferences the Ellsberg preferences.) Like the Allais preferences, the Ellsberg preferences violate Independence. Lotteries \(W\) and \(R\) yield a $100 loss if a yellow ball is drawn; they can be converted to lotteries \(RY\) and \(WY\) simply by replacing this $100 loss with a sure $100 gain. Because they violate independence, the Ellsberg preferences are incompatible with expected utility theory. Again, this incompatibility does not require any assumptions about the relative utilities of winning $100 and losing $100. Nor do we need any assumptions about where between 0 and 1/3 the probability of drawing a yellow ball falls. Where winning $100 has utility \(w\) and losing $100 has utility \(l\), \[\begin{align} EU(R) &= \tfrac{1}{3} w + P(W)l + P(Y)l \\ EU(W) &= \tfrac{1}{3} l + P(W)w + P(Y)l \\ EU(RY)&= \tfrac{1}{3} w + P(W)l + P(Y)w \\ EU(WY)&= \tfrac{1}{3} l + P(W)w + P(Y)w \end{align} \] It is easy to see that the condition in which \(EU(R) \gt EU(W)\) is exactly the same as the condition under which \(EU(RY) \gt EU(WY)\): both inequalities obtain just in case \(1/3\,w + P(W)l \gt 1/3\,l + P(W)w\). There are three notable responses to the Allais and Ellsberg paradoxes. First, one might follow Savage (101 ff) and Raiffa (1968, 80–86), and defend expected utility theory on the grounds that the Allais and Ellsberg preferences are irrational. Second, one might follow Buchak (2013) and claim that the Allais and Ellsberg preferences are rationally permissible, so that expected utility theory fails as a normative theory of rationality. Buchak develops an a more permissive theory of rationality, with an extra parameter representing the decision-maker's attitude toward risk. This risk parameter interacts with the utilities of outcomes and their conditional probabilities on acts to determine the values of acts. One setting of the risk parameter yields expected utility theory as a special case, but other, "risk-averse" settings rationalise the Allais preferences. Third, one might follow Loomes and Sugden (1986), Weirich (1986), and Pope (1995) and argue that the outcomes in the Allais and Ellsberg paradoxes can be re-described to accommodate the Allais and Ellsberg preferences. The alleged conflict between the Allais and Ellsberg preferences on the one hand, and expected utility theory on the other, was based on the assumption that a given sum of money has the same utility no matter how it is obtained. Some authors challenge this assumption. Loomes and Sugden suggest that in addition to monetary amounts, the outcomes of the gambles include feelings of disappointment (or elation) at getting less (or more) than expected. Pope distinguishes "post-outcome" feelings of elation or disappointment from "pre-outcome" feelings of excitement, fear, boredom, or safety, and points out that both may affect outcome utilities. Weirich suggests that the value of a monetary sum depends partly on the risks that went into obtaining it, irrespective of the gambler's feelings, so that (for instance) $100 million as the result of a sure bet is more than $100 million from a gamble that might have paid nothing. Broome (1991) raises a worry about this re-description solution. Any preferences can be justified by re-describing the space of outcomes, thus rendering the axioms of expected utility theory devoid of content. Broome rebuts this objection by suggesting an additional constraint on preference: if \(A\) is preferred to \(B\), then \(A\) and \(B\) must differ in some way that justifies preferring one to the other. An expected utility theorist can then count the Allais and Ellsberg preferences as rational if, and only if, there is a non-monetary difference that justifies placing outcomes of equal monetary value at different spots in one's preference ordering. 3.2.3 Counterexamples Involving Probability 0 Events Above, we've seen purported examples of rational preferences that violate expected utility theory. There are also purported examples of irrational preferences that satisfy expected utility theory. On a typical understanding of expected utility theory, when two acts are tied for having the highest expected utility, agents are required to be indifferent between them. Skyrms (1980, p. 74) points out that this view lets us derive strange conclusions about events with probability 0. For instance, suppose you are about to throw a point-sized dart at a round dartboard. Classical probability theory countenances situations in which the dart has probability 0 of hitting any particular point. You offer me the following lousy deal: if the dart hits the board at its exact center, then you will charge me $100; otherwise, no money will change hands. My decision problem can be captured with the following matrix: states hit center \((P=0)\) miss center \((P=1)\) acts accept deal \(-100\) \(0\) refuse deal \(0\) \(0\) Expected utility theory says that it is permissible for me to accept the deal—accepting has expected utility of 0. (This is so on both the Jeffrey definition and the Savage definition, if we assume that how the dart lands is probabilistically independent of how you bet.) But common sense says it is not permissible for me to accept the deal. Refusing weakly dominates accepting: it yields a better outcome in some states, and a worse outcome in no state. Skyrms suggests augmenting the laws of classical probability with an extra requirement that only impossibilities are assigned probability 0. Easwaran (2014) argues that we should instead reject the view that expected utility theory commands indifference between acts with equal expected utility. Instead, expected utility theory is not a complete theory of rationality: when two acts have the same expected utility, it does not tell us which to prefer. We can use non-expected-utility considerations like weak dominance as tiebreakers. 3.2.4 Counterexamples Involving Unbounded Utility A utility function \(U\) is bounded above if there is a limit to how good things can be according to \(U\), or more formally, if there is some least natural number \(sup\) such that for every \(A\) in \(U\)'s domain, \(U(A) \le sup\). Likewise, \(U\) is bounded below if there is a limit to how bad things can be according to \(U\), or more formally, if there is some greatest natural number \(inf\) such that for every \(A\) in \(U\)'s domain, \(U(A) \ge inf\). Expected utility theory can run into trouble when utility functions are unbounded above, below, or both. One problematic example is the St. Petersburg game, originally published by Bernoulli. Suppose that a coin is tossed until it lands tails for the first time. If it lands tails on the first toss, you win $2; if it lands tails on the second toss, you win $4; if it lands tails on the third toss, you win $8, and if it lands tails on the \(n\)th toss, you win $\(2^n\). Assuming each dollar is worth one utile, the expected value of the St Petersburg game is \[(\tfrac{1}{2} \cdot 2) + (\tfrac{1}{4} \cdot 4) + (\tfrac{1}{8} \cdot 8) + \cdots + (\tfrac{1}{2^n} \cdot 2^n) + \cdots \] or \[1 + 1 + 1 + \cdots = \infty \] It turns out that this sum diverges; the St Petersburg game has infinite expected utility. Thus, according to expected utility theory, you should prefer the opportunity to play the St Petersburg game to any finite sum of money, no matter how large. Furthermore, since an infinite expected utility multiplied by any nonzero chance is still infinite, anything that has a positive probability of yielding the St Petersburg game has infinite expected utility. Thus, according to expected utility theory, you should prefer any chance at playing the St Petersburg game, however slim, to any finite sum of money, however large. Nover and Hájek (2004) argue that in addition to the St. Petersburg game, which has infinite expected utility, there are other infinitary games whose expected utilities are undefined, even though rationality mandates certain preferences among them. One response to these problematic infinitary games is to argue that the decision problems themselves are ill-posed (Jeffrey (1983, 154); another is to adopt a modified version of expected utility theory that agrees with its verdicts in the ordinary case, but yields intuitively reasonable verdicts about the infinitary games (Thalos and Richardson 2013) (Fine 2008) (Colyvan 2006, 2008) (Easwaran 2008). 4. Applications 4.1 Economics and Public Policy In the 1940s and 50s, expected utility theory gained currency in the US for its potential to provide a mechanism that would explain the behavior of macro-economic variables. As it became apparent that expected utility theory did not accurately predict the behaviors of real people, its proponents instead advanced the view that it might serve instead as a theory of how rational people should respond to uncertainty (see Herfeld 2017). Expected utility theory has a variety of applications in public policy. In welfare economics, Harsanyi (1953) reasons from expected utility theory to the claim that the most socially just arrangement is the one that maximizes total welfare distributed across a society society. The theory of expected utility also has more direct applications. Howard (1980) introduces the concept of a micromort, or a one-in-a-million chance of death, and uses expected utility calculations to gauge which mortality risks are acceptable. In health policy, quality-adjusted life years, or QALYs, are measures of the expected utilities of different health interventions used to guide health policy (see Weinstein et al 2009). McAskill (2015) uses expected utility theory to address the central question of effective altruism: "How can I do the most good?" (Utilities in these applications are most naturally interpreted as measuring something like happiness or wellbeing, rather than subjective preference satisfaction for an individual agent.) Another area where expected utility theory finds applications is in insurance sales. Like casinos, insurance companies take on calculated risks with the aim of long-term financial gain, and must take into account the chance of going broke in the short run. 4.2 Ethics Utilitarians, along with their descendants contemporary consequentialists, hold that the rightness or wrongness of an act is determined by the moral goodness or badness of its consequences. Some consequentialists, such as (Railton 1984), interpret this to mean that we ought to do whatever will in fact have the best consequences. But it is difficult—perhaps impossible—to know the long-term consequences of our acts (Lenman 2000, Howard-Snyder 2007). In light of this observation, Jackson (1991) argues that the right act is the one with the greatest expected moral value, not the one that will in fact yield the best consequences. As Jackson notes, the expected moral value of an act depends on which probability function we work with. Jackson argues that, while every probability function is associated with an "ought", the "ought" that matters most to action is the one associated with the decision-maker's degrees of belief at the time of action. Other authors claim priority for other "oughts": Mason (2013) favors the probability function that is most reasonable for the agent to adopt in response to her evidence, given her epistemic limitations, while Oddie and Menzies (1992) favor the objective chance function as a measure of objective rightness. (They appeal to a more complicated probability function to define a notion of "subjective rightness" for decisionmakers who are ignorant of the objective chances.) Still others (Smart 1973, Timmons 2002) argue that even if that we ought to do whatever will have the best consequences, expected utility theory can play the role of a decision procedure when we are uncertain what consequences our acts will have. Feldman (2006) objects that expected utility calculations are horribly impractical. In most real life decisions, the steps required to compute expected utilities are beyond our ken: listing the possible outcomes of our acts, assigning each outcome a utility and a conditional probability given each act, and performing the arithmetic necessary to expected utility calculations. The expected-utility-maximizing version of consequentialism is not strictly speaking a theory of rational choice. It is a theory of moral choice, but whether rationality requires us to do what is morally best is up for debate. 4.3 Epistemology Expected utility theory can be used to address practical questions in epistemology. One such question is when to accept a hypothesis. In typical cases, the evidence is logically compatible with multiple hypotheses, including hypotheses to which it lends little inductive support. Furthermore, scientists do not typically accept only those hypotheses that are most probable given their data. When is a hypothesis likely enough to deserve acceptance? Bayesians, such as Maher (1993), suggest that this decision be made on expected utility grounds. Whether to accept a hypothesis is a decision problem, with acceptance and rejection as acts. It can be captured by the following decision matrix: states hypothesis is true hypothesis is false acts accept correctly accept erroneously accept reject erroneously reject correctly reject On Savage's definition, the expected utility of accepting the hypothesis is determined by the probability of the hypothesis, together with the utilities of each of the four outcomes. (We can expect Jeffrey's definition to agree with Savage's on the plausible assumption that, given the evidence in our possession, the hypothesis is probabilistically independent of whether we accept or reject it.) Here, the utilities can be understood as purely epistemic values, since it is epistemically valuable to believe interesting truths, and to reject falsehoods. Critics of the Bayesian approach, such as Mayo (1996), object that scientific hypotheses cannot sensibly be given probabilities. Mayo argues that in order to assign a useful probability to an event, we need statistical evidence about the frequencies of similar events. But scientific hypotheses are either true once and for all, or false once and for all—there is no population of worlds like ours from which we can meaningfully draw statistics. Nor can we use subjective probabilities for scientific purposes, since this would be unacceptably arbitrary. Therefore, the expected utilities of acceptance and rejection are undefined, and we ought to use the methods of traditional statistics, which rely on comparing the probabilities of our evidence conditional on each of the hypotheses. Expected utility theory also provides guidance about when to gather evidence. Good (1967) argues on expected utility grounds that it is always rational to gather evidence before acting, provided that evidence is free of cost. The act with the highest expected utility after the extra evidence is in will always be always at least as good as the act with the highest expected utility beforehand. In epistemic decision theory, expected utilities are used to assess belief states as rational or irrational. If we think of belief formation as a mental act, facts about the contents of the agent's beliefs as events, and closeness to truth as a desirable feature of outcomes, then we can use expected utility theory to evaluate degrees of belief in terms of their expected closeness to truth. The entry on epistemic utility arguments for probabilism includes an overview of expected utility arguments for a variety of epistemic norms, including conditionalization and the Principal Principle. 4.4 Law Kaplan (1968), argues that expected utility considerations can be used to fix a standard of proof in legal trials. A jury deciding whether to acquit or convict faces the following decision problem: states guilty innocent acts convict true conviction false conviction acquit false acquittal true acquittal Kaplan shows that $EU(convict) > EU(acquit)$ whenever $$P(guilty) > \frac{1}{1+ \frac{U(\mathrm{true~conviction})-U(\mathrm{false~acquittal})}{U(\mathrm{true~acquittal})-U(\mathrm{false~conviction})}}$$ Qualitatively, this means that the standard of proof increases as the disutility of convicting an innocent person ($U(\mathrm{true~conviction})-U(\mathrm{false~acquittal})$) increases, or as the disutility of acquitting a guilty person ($U(\mathrm{true~acquittal})-U(\mathrm{false~conviction})$) decreases. Critics of this decision-theoretic approach, such as Laudan (2006), argue that it's difficult or impossible to bridge the gap between the evidence admissible in court, and the real probability of the defendant's guilt. The probability guilt depends on three factors: the distribution of apparent guilt among the genuinely guilty, the distribution of apparent guilt among the genuinely innocent, and the ratio of genuinely guilty to genuinely innocent defendants who go to trial (see Bell 1987). Obstacles to calculating any of these factors will block the inference from a judge or jury's perception of apparent guilt to a true probability of guilt.