

UNCERTAINTY AND PROBABILITY WITHIN UTILITARIAN THEORY

– Jonathan Baron –

Abstract. Probability is a central concept in utilitarian moral theory, almost impossible to do without. I attempt to clarify the role of probability, so that we can be clear about what we are aiming for when we apply utilitarian theory to real cases. I point out the close relationship between utilitarianism and expected-utility theory, a normative standard for individual decision-making. I then argue that the distinction between “ambiguity” and risk is a matter of perception. We do not need this distinction in the theory itself. In order to make this argument I rely on the personalist theory of probability, and I try to show that, within this theory, we do not need to give up completely on the idea that a “true probability” (other than 0 or 1) exists. Finally, I discuss several examples of applied utilitarianism, emphasizing the role of probability in each example: reasonable doubt (in law), the precautionary principle in risk regulation, charity, climate change, and voting.

Keywords: utilitarianism, ambiguity, probability, expected utility.

1. Introduction

Utilitarianism is one approach to moral theory. It provides a normative model of decision making, by which the best choice among the options on the table is an option that maximizes the expected utility of the overall outcome for everyone affected. Utility is a numerical measure of “good” (as a quantity, as in “doing more good”). It assumes that the relevant measure of overall good is the sum of good across individuals; so each person is treated independently of everyone else. It requires that utility differences (e.g., those between choice options, for a particular person) are interpersonally comparable (in theory, even though such comparison may be difficult in practice).

The result is that it makes sense to ask whether the harm to some people is compensated by the benefit to others. An important consequence of the theory is this: if the chosen option is not one that is optimal according to the theory, then some people are being harmed in a way that is not compensated by a benefit to others, and the harm comes from someone’s failure to choose the best option. Any moral theory that prescribes such a sub-optimal option will thus lead to harm that cannot be justified by compensating benefit and thus requires some alternative justification. In practice, it often happens that, if we simply ask ourselves which

options are optimal in this sense, the answer is obvious, and this step in is often part of our moral reasoning even if we then go on to ask whether the utilitarian optimum is ethical in some other sense.

The assumptions behind utilitarian theory have been defended elsewhere by me and others.¹ Although many scholars criticize utilitarianism, they often do so by making arguments that have already been answered.² I will not defend it further in this short article. My purpose here is to try to clarify the role of probability within utilitarian theory.³ The points I make may apply to other moral theories as well. If we can clarify what we are aiming for when we try to follow some theory, then we may be less confused and more likely to succeed, even when we do nothing more than keep the theory in mind when we think as we normally do.

Utilitarianism relies heavily on probability and has done so since the outset.⁴ In this regard, it often differs from both intuition and some other moral approaches, but the arguments for the use of probability seem strong to me, as I shall explain. A lack of appreciation of probability can lead people to reject any attempt to think in utilitarian terms at all.

Utilitarianism and many other moral theories are about decisions, choices among options. Utilitarianism and some other theories focus (at least in part) on the outcomes of options. Outcomes may appear to be certain, but they usually are not. Certainty itself is an illusion, because outcomes can be unpacked into other outcomes that are less certain. You may think that “I win \$100” is a certain outcome, but the receipt of the money is only the beginning of a chain of events. You can spend the money on a restaurant dinner that turns out to be wonderful or that gives you food poisoning. You can invest in an investment that gains or loses. And so on. The description of an outcome as certain is just a label for a series of events that includes uncertainty. Probability is a way of thinking about uncertainty.

2. Expected-utility theory (EUT)

Utilitarianism has close ties with expected-utility theory (EUT), which is a normative model (a standard) for the trade-off between utility and probability of outcomes.⁵ EUT says that the evaluation of options should correspond to their expected utility (EU), which is the sum of the utilities of the possible consequences of

¹ Baron (1993, 1996, 2004, 2006, 2008); Hare (1963, 1981); Kaplow, Shavell (2002, 2004).

² Especially, Hare (1981).

³ Baron (1996) discusses utility, another central concept of the theory.

⁴ Bentham (1843/1948).

⁵ Baron (2008).

the option, after multiplying each of these utilities by the respective probability of the consequence. In a simple example where all the outcomes are monetary, suppose you have a choice between options A and B, where option A is a gamble in which you could get \$16 with probability .5 or \$4 with probability .5, and option B is \$9.50 for sure. Suppose that the utility of money is the square root of the amount. In this case, the EU of A is $.5\sqrt{16} + .5\sqrt{4}$, or 3, and the EU of B is a little more than 3, so you should choose B, even though the expected outcome of A (\$10, the average of \$16 and \$4) is greater than that of B (\$9.50).⁶

EUT implies several principles, which (arguably) follow logically from the idea of analyzing decisions into outcomes (which have utility), options (which you control) and unknown states of the world (which you do not control, and which, together with options, determine the outcome).⁷ Probability is a property of the states of the world. One important principle (the sure-thing principle of Savage⁸), for example, holds roughly this: if some state of the world leads to the same outcome (in terms of everything you care about) regardless of which option is chosen, then the nature of that outcome should not affect your choice. Another principle (transitivity) holds that, if option A is at least as good as B, and B is at least as good as C, then A is at least as good as C. Alternatives to EUT will violate at least one of these basic assumptions, which seem to follow from what we mean by states, options, outcomes, and “better.”

EUT has a close relationship to utilitarianism. Consider a case in which an identical choice affects many people with the same outcomes (with their utilities) and probabilities for each person affected. An example could be a medical treatment, which has a high probability of curing a serious illness but a very low probability of causing death. (We assume that the utility of the illness relative to death is the same for everyone.) In one situation, an individual could decide whether to get the treatment according to EUT. In another situation, a policy maker could decide whether to recommend the treatment or not to everyone with the illness in question (assuming that the recommendation would be accepted). For a 1000 people, a 5% probability of death means that we can expect 50 people to die. More generally, probabilities in the EUT analysis turn into numbers in the utilitarian analysis. If the EU of the treatment is positive (better than no-treatment) for an

⁶ This example illustrates another common assumption often made in practical contexts: the utility of money is a concave function of the amount, on the average if not in every case. This assumption implies that, other things equal, we can increase total utility by taking from the rich and giving to the poor. Money has more utility for the poor.

⁷ Baron (2006, 2008).

⁸ Savage (1954).

individual, then the total utility must be likewise positive for the group when we *sum* the utilities of the outcomes across the individuals affected, as utilitarianism prescribes. If EUT is the normative theory for individuals, then, at least in situations like this (and by extension in other, more complex, situations), utilitarianism must be the normative theory for the group. A decision made for the group on the basis of any theory that gives a different answer from that of utilitarianism could thus make every member of the group worse off, as determined by EUT.⁹

Note that probability is relevant to utilitarianism even when the probabilities are not only the same for everyone but also correlated, so that the proposition in question is either true for everyone or false for everyone (e.g., some event either happens or doesn't happen). If everyone is affected in the same way by a decision, then it should not matter whether we view the decision from the perspective of the individual (EUT) or that of the group (utilitarianism).

3. What is probability? Risk vs. ambiguity

A source of unease about the usual of probability, often raised as an objection to utilitarianism, is that probability is not always meaningful in some cases that we would still say involve "uncertainty." The problem is not unique to utilitarianism. It applies just as well to individual decision making based on EUT when nobody else is considered. The particular version of probability theory that I advocate is a solution to this problem, but I do not claim that it is the only possible solution. Still, both EUT and utilitarianism require some concept of probability to be applicable. If probability is meaningless for some cases that involve uncertainty, then both theories are limited.

I do not claim that probability can always be calculated or determined exactly. What is critical is that it makes conceptual sense.

Many writers make a distinction between risk and uncertainty, with uncertainty also called "ambiguity" or "deep uncertainty." Intuitively, the distinction seems to matter. Which would you choose? In option A, you can draw a ball from an urn with 50 black balls and 50 red balls. If you guess the color correctly before you draw, you win \$100. In option B, you can draw from an urn with 100 balls, each red or black, but with an unknown number of each type. Again, if you guess the color correctly, you win \$100. Most people intuitively favor A, even if the payoff is somewhat greater for B.¹⁰ It feels as if the probability of winning in A is "known" to be .5, while the probability in B is ambiguous, or unknown, or deeply

⁹ Kaplow, Shavell (2002, 2004) extend this kind of argument to a number of cases.

¹⁰ Baron, Frisch (1994) review the evidence.

uncertain. We tend to avoid action when we feel that we don't know something that we ought to know.¹¹ The avoidance of ambiguity also leads to violation of the sure-thing principle.¹²

The intuition that ambiguity matters is a matter of perception, the feeling that some information is missing. We can change the perception without changing the situation. For example, we can remind you that the ball in urn A will be drawn from the top layer of balls, and you have no idea how many balls are in that layer. Or we can make a small change in case B, in which we flip a coin after drawing the ball in order to determine whether your guess or the opposite wins, in which case it now becomes easier to think that the probability of winning is indeed .5.

More generally, people tend to think of probability as two different concepts, which Fox and Ülkümen call "aleatory" and "epistemic" probability.¹³ Aleatory probabilities are those that seem to arise from mechanisms that produce different frequencies of outcomes in ways that are understood and repeatable, like urns with known proportions, unbiased coins, and selection of individuals from a population. (The probability that a randomly selected American will have a birthday in February seems aleatory.) Epistemic probabilities are those that arise from ignorance, from missing information that may become available. These typically concern the probabilities of unique events, such as a military invasion of a particular country, the outcome of a particular election, or whether string theory is true.

Both types of probability may be subsumed under a single interpretation, which Savage calls "personal" probability.¹⁴ By this account, probability is a number assigned to a personal degree of belief. It is strongly related to decision making. If I want to know your degree of belief that some party will win the next election, I can ask you whether you would prefer to bet on a win by that party or on a red ball being drawn from an urn with 70% red balls. If you are indifferent, and if you assign a probability of .7 to the red ball, then I can infer that you assign the same probability to the election. (If I ask you directly for the probability of win-

¹¹ Ritov, Baron (1990).

¹² Ellsberg (1961). An urn has 30 red balls, and 60 balls that are either blue or yellow. You don't know how many are blue, or yellow. One ball will be drawn. You have a choice between option X and Y. With X you get \$100 if a red ball is drawn. With Y, \$100 if it is blue. Most people choose X because they feel that they "know what the probability is." Now you have a choice between V and W. With V, you get \$100 if the ball is red or yellow. With W, \$100 if it is blue or yellow. Most people now choose W because they "know that they have a 2/3 chance to win." Note that X and Y have the same outcome for yellow (a state of the world), \$0. And V and W have the same outcome for yellow. The sure-thing principle says that the outcome shouldn't matter. So you should either choose X and V, or W and Y. Or you should be everywhere indifferent.

¹³ Fox, Ülkümen (2011).

¹⁴ Savage (1954).

ning the election, you may not give 70% exactly; you may distort probability judgments in various ways.) If you are unwilling to bet according to your degree of belief, you are acting in a way that is inconsistent with your own goals and preferences. If you like money, you can wind up with less of it by betting on outcomes you consider less likely to win.

In the case of epistemic probability, people often have different degrees of belief, and even people with the same degree of belief can assign different numbers to their beliefs when asked for numbers. In the case of aleatory probabilities, people tend to agree on the number, but even this case often allows room for reasonable disagreement. If you want to know the probability of a person getting a certain disease, do you look at the proportion of people who get that disease? or the proportion who match the person in sex and age? and so on.

The theory that probability is based on personal degree of belief allows us to make sense of the idea that unique events have probabilities. If this idea were nonsense, decision making would, to say the least, be very difficult.

4. Assessing probability judgments

Despite the fact that individuals may differ, we can assess individual probability judgments in two different ways, which are called coherence (of judgments with each other) and correspondence (with reality). To assess coherence, we need different probability judgments of related propositions, from the same person. For example, if you believe that the probability of rain tomorrow is .7, and the probability of no rain is .4, you are incoherent, because probabilities (by definition) must add to 1. Several other tests of coherence may be applied,¹⁵ and most of them arise from the relation between probability and decision making.

Measures of correspondence also require several judgments from the same individual, and they apply to that individual in the situations tested. The simplest measure is calibration. If we look at a large number of cases where you say that the probability is .7, then, if you are well calibrated, the event in question will occur 70% of the time. Such measures are useful if we want to use someone else's probability judgments to make our own decisions. Another measure is discrimination, which is whether you give higher probabilities for events that happen more often. You can be well calibrated without discriminating, e.g., give the same probability every day for whether it will rain, the proportion of rainy days in the year. And you can discriminate while being very poorly calibrated, e.g., predict a 10% chance of rain every day that it rains and a 5% chance when it does not.

¹⁵ Baron (2006, 2008).

It may help you to know that probabilistic weather forecasts are almost perfectly calibrated, as well as being pretty good at discriminating in the short term. This is probably because forecasters (human beings, not computers alone) are given extensive feedback about their forecasts and are scored with the use of a scoring rule that rewards useful probability judgments, taking both discrimination and calibration into account. One such rule is the Brier score. When the outcome is known (e.g., rain, snow, or neither), give the outcome that occurred a 1 and outcomes that did not occur a zero. Then look at the probabilities assigned to the outcomes by the forecaster and find the difference between the probability and the revealed truth (1 or 0). Then square these differences and add them up. Perfect forecasts (1.00 to the event that happened, 0 to the others) get a score of 0. Higher Brier scores indicate less useful judgments. If the forecaster says [.7, .1, .2] for [rain, snow, clear] and it rains, the truth is [1, 0, 0] and the Brier score is $.3^2 + .1^2 + .2^2$ or .14, a very good score.

5. Why the Brier score is special

The Brier score has a special property in the context of decision making.¹⁶ Suppose that the probability judgment is made by one person (the judge) and used by another (the decision maker, or DM). The judge does not know what decisions the DM will make on the basis of the judgment. It could be that the DM has a range of options, each with a different probability threshold. In medicine, for example, some decision rules are stated in terms of probabilities, e.g.: “If the probability of cancer is above .30, then we maximize EU by doing a biopsy, but if the probability is between .10 and .30, we maximize EU by ordering another screening test in 6 months.” For a case with a single decision and one threshold, the threshold is determined (in EUT) by comparing the disutility of the two possible errors, misses and false alarms. If the disutility of a miss is very high relative to that of a false alarm, then we want a low threshold for acting, and, conversely, a high threshold if disutility of a false alarm is high.

In this sort of situation, we can reasonably assume that the threshold is equally likely to be anywhere between 0 and 1; i.e., it is uniformly distributed over the interval. In this situation, given a few other reasonable assumptions, the EU of the judgment for the DM who bases a decision on the stated probability is proportional to the (negative of) the Brier score of the judgment.¹⁷ Thus, the Brier score is a generally good estimate of the “goodness” of a probability judgment for the

¹⁶ Baron et al. (2014).

¹⁷ Ibidem, Appendix.

purpose of making decisions, when the action thresholds for these decisions are unknown. And, as a result, the average Brier score of a judge in a given context is a measure of her usefulness in that context.

6. True probability

On the basis of the personal theory of probability, as I have stated it, probability is useful for an a single person making decisions based on her own beliefs and utilities (values). This is sufficient, since decisions are in fact made by individuals, even when they are deciding how to influence a group that must somehow agree on a final decision affecting many people. But groups can benefit from a clear concept of what they are doing and how to talk about it. For this purpose, it can help to clarify the concept of “true probability.”

By the personal theory, the true probability of a proposition could be 0 or 1, depending on whether the proposition is true or false, and otherwise the concept would seem to be nonsense. It is like asking about the true value of the attractiveness of a face. Judges will differ. However, the intuition that there is some sort of right answer to probability questions is a strong one. It is clear that some answers are better than others. Anyone who said, a week before, that Donald Trump had an 80% chance of winning the U.S. election of 2016 needs a psychiatric examination.

Brown (1993) suggested that something like a true probability could be derived within the personal theory by asking what probability a good judge would assign if a standard set of evidence for that judgment were available. Thus, a doctor might reasonably say, “The [true] probability that you have this disease is between 20% and 50%, and I will tell you exactly when I get all the test results back.” Here, the true probability is contingent on the results of the standard tests. We might even make the true probability contingent on *all* the available evidence, although (as Brown notes) the concept of what is “available” is a little slippery. Yet, in most cases we can imagine what “available evidence” means. Once we have all the evidence, the remaining uncertainty is “irreducible.” It is, in a sense, true epistemic uncertainty.¹⁸

This cannot be the whole answer, because judges will still differ. However, we can take one more step, which is to say that the true probability is the judg-

¹⁸ Some interesting cases for this account are those in which public knowledge is available to anyone willing to look for it but private knowledge is available to insiders, who may be prohibited from revealing it. Yet the same insiders may participate in forecasting tournaments or betting markets, to their advantage. We thus need two criteria of “availability,” and the tournaments may perform better as probability judges than what seems possible from the public evidence.

ment made by the best possible judge. Such a judge need not be a single person; it could be a person aided by computers (as in the case of most weather forecasts), or a group of people, possibly aided by computers. To determine the best judge, we apply the result of the last section: the best judge is the one with the lowest Brier score, contingent on all the available evidence. And the ideal judge is one whose Brier score cannot be improved (lowered) any further without collecting evidence that is not available. We can thus create a concept that captures our intuition of what “true probability” would mean if it were something other than 1 or 0. I think we need a better term for it, like “optimal probability.” Note that, for aleatory uncertainty, this optimal probability is the relative frequency, assuming that no other information is available about the path of the flipped coin through the air or the arrangement of cards in a deck. A judge who stated the relative frequency would get the best score.

We can approach the optimal probability empirically, by finding ways to minimize the Brier score [e.g., Tetlock, Mellers, Scoblic (2017)]. When we implement those methods of probability judgment, then we can come close to this optimum.

7. Practical problems

Even if we can define “true probability” in principle, we often have difficulty in practical situations. We are, in a sense, thrown back to the problem of ambiguity. If there is a true probability, we don’t know what it is. And even if the experience of missing information epistemic uncertainty is clear, we still must make decisions. Sometimes, in real situations, an attempt to make a probability judgment can clarify just what information is missing, and we can sometimes get it before the decision must be made.¹⁹ But this is not always possible. What should a decision maker do?

One answer is to consider “second-order probability.” We can ask questions like “What is the probability that, if we had all the available information, our probability judgment would be greater than .5?” We could create a distribution of probabilities of probabilities. This practice may have value if some consequence is defined in terms of a probability threshold, that is, if the probability judgment itself has consequences that are part of a decision. For example, “The regulation says that you cannot build a bridge if the probability of collapse in 100 years is

¹⁹ Savage (1954): ch. 6; summarized by Baron (2008): ch. 7, introduced a general approach to evaluation of the decision to collect information: we compare the EU of the decision without the information to the EU with it. This approach is now applied extensively in medical decision analysis and elsewhere.

greater than 2%." The builder is thus legally responsible if the bridge collapses in 50 years, and someone determines that, if he had all available information, we would have concluded that the probability was 3%. This situation is rare, but Brown gives some real examples²⁰.

In more typical cases, the consequences do not depend on the judgment itself, and the EU of an option is determined by the mean of any second-order distribution, that is, presumably, the original judgment, the best guess. The computation of a second-order distribution should not change the decision, once the best possible judgment is in hand.

An alternative approach, often taken, is to apply some decision rule that does not involve probability, as discussed in the next section. Such rules are often equivalent to choosing the option that would be chosen if the probability of interest were 1, or if it were 0. For example, "If you don't know the risk of default, don't buy that bond." or "The risk of a hurricane here is so low that you certainly don't need hurricane insurance (even if it costs very little). Just ignore it."

These rules all have a general problem. If you throw out your best probability judgment, it is likely that the simple rule will have a lower EU. The expected Brier score of a probability of 0 or 1 is almost never better than that of an intermediate judgment. In particular, if you choose 0 or 1, you could be wrong, and then your Brier score is very high (i.e., very bad).

That said, there are some cases where we can expect the simple rule to do better. These are cases where we have good reason to think that our probability judgments are systematically distorted, e.g., by wishful thinking. One example is a decision about whether to commit adultery.²¹ Those who are faced with such a decision usually would like to believe that they can carry on their tryst in secret, so that the probability of harm to their spouses, and possibly to themselves as a result, is very low, so low as to make adultery seem like the better option. Yet just the fact that they know that they would prefer such a belief means that it is likely to be distorted, and the Biblical prohibition of adultery is the better choice. The same goes for terrorism. Terrorists throughout history have usually thought that their acts would be for the greater good. Although we can imagine cases where this was true, or is true even now, these cases are rare. If we look at the ratio of the number of these true cases to the much larger of false ones, we ought to be convinced that, even if we are in a position to commit an act of terrorism, our judgment that it is "probably" for the greater good is almost certainly incorrect.

²⁰ Brown (1993).

²¹ Hare (1981): sec. 3.2.

J.S. Mill made similar arguments against (e.g.) suppression of free speech: yes, there are cases in which suppression has better consequences, but those who think that they have found such cases are most likely incorrect in their judgment.²²

Such arguments are often identified as favoring “rule utilitarianism,” the view that we should apply utilitarianism to rules and then follow the rules regardless of their perceived consequences. This not what I am suggesting. Rather (as Hare points out), this argument is fully within the scope of act utilitarianism. In a given case, we need to evaluate our probability judgments themselves, thus, in a sense, “correcting” them on the basis of what we know about extraneous influences on them.

Many of the examples below consist of cases where it seems to me that the more basic argument applies. There is no reason in most of these cases to expect such extreme distortions of probability judgments as to justify ignoring them completely. The better justified principle is that we do the best we can and proceed. We can’t do any better than that.

To summarize so far, we should in general go with our best judgment, after gathering whatever information is worth gathering. One exception is when consequences depend on our judgment itself. Another is when we have good reason to think that our judgment is biased in a particular direction.

A final case of interest is one in which we have more than one probability judgment, each from some person or system certified as reasonably good, but the judgments disagree. In this case, a large literature implies that we should simply aggregate the judgments into a single overall judgment.²³ We do not need to choose one source and reject another, and, in fact, will usually do worse with such an approach.

8. Policy examples

I now turn to some examples of the use and misuse of probability in policy decisions to which utilitarian theory may be applied. They are “moral” or “ethical” in the sense that they affect other people, and usually they involve trade-offs between different people, so that some are harmed and others are helped by most changes in policy. Most of these decisions are made by governments, with input from citizens, so citizens, in a sense, also make these decisions as members of a group.

²² Mill (1859).

²³ E.g., Baron et al. (2014) for aggregating large number of judgments; or Lindley, Tversky, Brown (1987) for cases with small numbers.

These cases are characterized by the fact that intuition and law (when relevant) seem to depart from the utilitarian optimum, and the point of the departure seems to be the use of probability. A consequence of the argument I have sketched is that probability is useful for all these decisions. We do not need to assign numbers in order to “use” probability. In most cases, thinking of the decision as one involving probability affects the way we think about it, how we set it up in our minds. Often that set-up alone is sufficient to lead to the answer that would be reached by a more thorough process of assigning numbers and doing calculations, but would not be reached by consulting intuition alone, unaided by the theory.

8.1. Reasonable doubt and the presumption of innocence

Common law requires that juries in most criminal cases presume that the defendant is innocent until proven to be guilty “beyond a reasonable doubt.” The system seems to work most of the time. The point is that false convictions of criminal offenses lead to harsh penalties that do practically nothing but the harm they cause to the convict. (They might have some deterrent effect on others, but they might also cause others to lose faith in the system and be less deterred, thinking that it doesn’t matter so much whether they offend or not.) Jurors seem to be aware of this problem, and prosecutors often lose their cases despite being convinced themselves of the defendant’s guilt. Yet the instructions given to juries are, at least, difficult to understand for any juror who took probability seriously.

What does “presumed innocent” mean? It cannot mean that the probability of guilt is zero. If it meant that, no amount of evidence could change that probability.²⁴ If the presumption of innocence means that the jury should assume “a low probability,” this is inconsistent with the facts: the relative frequency of guilt, given that a defendant is brought to trial, is about 50% or higher, conservatively, by most methods of estimating this proportion.

More realistically, the presumption of innocence is redundant with the instruction that guilt must be proved beyond a reasonable doubt. This means that the threshold for conviction, given the evidence, should be high. How high? Given the argument presented earlier for thresholds, the threshold should be determined by comparison of the relative disutility of the two possible errors, conviction of an innocent person and acquittal of a guilty one. It is reasonable to assume that the former error is many times worse than the latter. Although the probability of ap-

²⁴ In order to conform to the congruence constraints on probability mentioned earlier, the probability of guilt after evidence is available follows Bayes’s rule:

$$p(\text{guilt} | \text{evidence}) = \frac{p(\text{evidence} | \text{guilt})p(\text{evidence})}{p(\text{evidence}) + p(\text{evidence} | \text{not guilty})p(\text{not guilty})}$$

Here $p(\text{guilt} | \text{evidence})$ is the probability of guilt given the evidence, and $p(\text{guilt})$ is the probability of guilt assumed before the evidence is available. Clearly, if the latter is 0, nothing else matters.

prehension and conviction is surely relevant to deterrence of future crimes, a low probability of ultimate conviction can be compensated by a greater punishment²⁵ and often is compensated this way.²⁶

8.2. The precautionary principle

The precautionary principle has been proposed and sometimes enacted (notably, in the European Union) as a basis for regulating new technologies such as the use of genetically modified organisms (GMO) for food. It is similar to the idea of presumption of innocence, but reversed. The idea is that new technologies must be shown to be safe before they are approved. They are, in a sense, presumed guilty. Once again, those who think in terms of probability would have difficulty understanding what this means, and, once again, it could be understood as requiring a high threshold for the probability of “safe” given the available evidence. Sunstein summarizes a number of different versions of the principle, finding all of them to lead to incoherent conclusions.²⁷

A simple way of stating the problem is that new technologies have benefits as well as risks. When we make decisions consistent with EUT, we should take into account the potential disutility of the loss of benefits resulting from delay in adoption of a new technology. If we do not do this, we hurt some people by failing to help them. In the case of GMOs, the benefits include resistance to drought and pests, reduced use of pesticide, lower cost for farmers, and in some cases products that last longer or contain healthier ingredients (“golden rice”).

Although the U.S. never formally adopted any version of the precautionary principle, some of its regulatory agencies act as if they follow it.²⁸ In particular, the Food and Drug Administration has been reluctant to approve new drugs quickly, even when they seem to be effective in curing or preventing otherwise-fatal conditions. This policy has exceptions. And there are some long-term benefits of delay in approval, particularly the possibility of doing well-controlled studies before approval. However, it seems clear that some part of the delay is the fear of having to disapprove some drug already approved, as later evidence indicates that approval was unwarranted.

²⁵ Bentham (1843/1948).

²⁶ Baron, Ritov (2009). An alternative way of thinking about the threshold is Blackstone’s rule: “It is better that ten guilty persons escape, than that one innocent suffer.” Dekay (1996) shows that this is not equivalent to a threshold based on comparison of disutilities.

²⁷ Sunstein (2007).

²⁸ Baron (2006); Bazerman, Baron, Shonk (2001).

More generally, these biases illustrate “omission bias,”²⁹ a preference for harms caused by omission (inaction) over lesser harms caused directly by commission (action). Utilitarianism adopts a more inclusive definition of causality, “but for” causality (“But for the choice I made, the outcome would not have occurred”), which concerns options rather than actions. Some parts of the law also adopt this definition, particularly tort law, where negligence is cause for a lawsuit.

Note that omission bias is found both for probabilities within individuals and for numbers when the choice affects groups, to approximately the same degree.³⁰

8.3. Insurance against catastrophic risks

Insurance companies, with few exceptions, typically refuse to provide insurance against “unknown risks,” such as the risk of a military attack on a cargo ship, or a meltdown of a nuclear power plant, in part because they perceive the probability to be ambiguous.³¹

8.4. Charity and diversification

When we invest money, diversification is a good idea. Each investment has some risk, and the risks are not completely correlated. The whole idea of “hedging” is to find investments with negatively correlated risks, so that the combination has much less risk than either investment alone. Risk is bad because of the declining marginal utility of money, the concave form of the utility function for money. The disutility of a large loss is greater than the utility of a gain of the same size, and, therefore, the EU of a risky investment is less than the utility of its expected value (the average outcome in dollars when weighted by probability of each possible amount). Diversification goes in the direction of hedging.

When individuals contribute to charity, they seem to carry over the same intuition. When my mother died, I had all her mail forwarded to me. I must have gotten renewal notices for 20 different charities, all good causes I’m sure. But she seemed to respond to any solicitation by writing a check. Putting aside the cost of all this mailing for what amounted to relatively small contributions (if only because there were so many), this is an ineffective strategy, out of line with the utilitarian theory of charity advanced now by organizations that promote “effective altruism.”

²⁹ Ritov, Baron (1990).

³⁰ Ibidem.

³¹ E.g., Kunreuther, Hogarth, Meszaros (1993).

The argument for diversification is irrelevant in this case, although people do intuitively think that diversification is reasonable.³² Unless you are Bill Gates or George Soros, the amount you can contribute to to each charity is very small compared to its total. The utility of the total budget is probably marginally declining for most charities – with too much money they have to look for new ways to spend it – but any small part of that function is indistinguishable from a straight line. Thus, the harm (disutility) caused by taking \$100 from the charity is very close to the benefit (utility) of contributing \$100.

Given that there is no benefit of diversification, the best strategy is to decide which of the charities you want to support does the most good for what you are willing to contribute to it, and then give all your available money to that one.³³ This is non-intuitive, but it follows from the position I have advocated here. Of course you do not know for sure which charity is most efficient but you could think in terms of probabilities. Each charity could have a probability for each possible level of efficiency, and you could estimate (in principle) the expected efficiency for each one.

You may think, “If I am correct, then this makes sense, but what if I misestimate the probabilities of different levels of efficiency?” You may think that this is a case where probabilities are truly ambiguous. Yet the arguments I have made imply that this doesn’t matter. Your best guess is still your best guess. From the outset, you knew that risk was involved, since you cannot know exactly the efficiency of different programs, especially when they have different goals, which you need to put on the same scale of utility for comparison. A charitable donation is like any other expenditure: what you expect is not necessarily what you get. The fact that it is risky does not make it any less valuable. It could be better than what you expect, as well as worse. The same applies to other sorts of personal commitments. When a scientist pursues a line of research, or a politician tries to get an idea passed into law, it is possible that the effort is wasted, or that it succeeds. Risk is everywhere.

8.5. Climate change and diversification

What applies to individuals does not apply directly to nations. Consider the problem of what a big nation like the U.S. should do if it had decided, through its

³² Baron, Szymanska (2010).

³³ The restriction here to “what you are willing to contribute to it” is, I think, necessary. Utilitarianism must be applied to real options, and the option of giving all or most of our money away is, for most of us, not on the table when we make these decisions. The limits of altruistic motivation, while somewhat modifiable (even by ourselves over the long run) are limits on our options, just as are many other facts about the world.

government, to do something about climate change. There are many different paths to take: reduce the use of carbon fuels (with many sub-methods of doing that), do research on alternative sources of energy (again, many possibilities here), do research on “negative emissions” (ways of removing carbon from the atmosphere, either at the time it would be emitted or directly from the air), plant trees or counter deforestation, build defenses against rising oceans, reduce excessive population growth in low-lying areas in Africa, find ways to accept more refugees from lands destroyed by rising oceans, and so on.

The argument I made for charity could be applied here: pick the best one and put everything into it. But, for a large nation, or a group of nations acting together, the utility function for each of these approaches is indeed concave, and sometimes sharply so. For example, research on nuclear fusion for energy generation requires a fairly substantial sum to get off the ground, but, given that sum, the limiting factor is more likely to be the slow progress of science. Speeding up that process is possible with more money, but the expected payoff is probably lower than that of funding the initial technology and research. The same can be said about research on negative emissions (although here I suspect that we are nowhere near the point of saturation by too much money). And once solar energy reaches a certain level of efficiency, subsidies from government will not do much further good beyond what the market will do by itself.

In sum, in this case, and similar cases, it pays to diversify.

8.6. Voting and citizen participation

A final example of the utilitarian use of probability, similar to the case of charity, is voting, or, more generally citizen participation in politics (including protesting, letter writing, and political work). Political scientists have understood for some time that voting is not justified as a way of advancing self-interest.³⁴ The problem is the low probability of being the decisive (pivotal) voter. If you are not the decisive voter, you cannot affect the outcome (to a first approximation – sometimes the size of the vote affects the power of a “mandate”). Even if you stand to gain a fortune if one side of an election wins, the EU of a vote in most elections is still no more than that of a couple of small coins, if your utility comes only from the money you get yourself. A rational voter who understands probability must have some other reason to vote, aside from self-interest.

Altruism is one source of utility. Altruists gain utility from increases in other people’s utility. From the point of view of us all, we would want people to vote

³⁴ Downs (1957); many voters do not know this and think that voting is worthwhile on grounds of self-interest alone, as found by Baron (2012).

out of altruism rather than other sources of motivation. Other sources could lead people to vote in ways that cause harm to others rather than doing good. People do vote for other reasons, such as a sense of duty, which may or may not conflict with the utilitarian rationale for voting.

A simple analysis of that rationale, for the individual voter, depends on two factors: the probability of being pivotal (or, more generally, having an effect) and the magnitude of the effect, in terms of utility. The magnitude depends on both the average effect for each person and the number of people affected. The probability is roughly proportional to, but less than, the reciprocal of the number of voters N . However, as N increases, the magnitude of the effect usually increases too, by a factor directly proportional to N . Thus, the reduced probability with a large number of voters is roughly canceled out by the increased magnitude of the effect.³⁵ If the average benefit per person, your degree of altruism toward those who benefit, and the number of them are sufficiently great, voting for what is best for these people is rational for you.

Importantly, the number of people affected is often orders of magnitude greater than N (the number of voters). Many issues at stake in elections have effects on foreigners, children and other citizens who do not vote, and people not yet born, around the world. A prime example is climate change, although many such issues concerning worldwide resources exist.

For a utilitarian voter with some altruism toward people in general, it is thus rationally worth the effort to vote, and to be sufficiently informed to vote for the better side, if the ballot contains proposals (or candidates) that affect very large numbers of people. It is not rational to vote at all if the voter thinks in terms of narrow self-interest alone. When altruism is limited to those in the voter's in-group, such as citizens of the same nation, it may or may not be rational. When one side of a choice on the ballot is better for the in-group and the other side is better for the world, then a voter who votes for the in-group side is harming the rest of the world. We should see this for what it is, a form of immoral behavior, something that results from a way of thinking that we, at least those of us who do care about people in general, should want to discourage.³⁶

This sort of utilitarian view, which takes probability into account, thus contrasts with two other views that ignore probability. One, already mentioned, is that voting is justified by narrow self-interest. This view ignores the low probab-

³⁵ Edlin, Gelman, Kaplan (2007, 2008).

³⁶ The arguments for such wide concern, often called cosmopolitanism, are outside of my topic, but clearly it fits well with utilitarian theory.

ity of having any effect. The other view is that voting is not worthwhile at all because it has no effect. This is the equivalent of treating very low probabilities as if they were zero.

9. Conclusion

I have argued here that a complete utilitarian theory of choices requires us to deal with uncertainty, and that probability provides a conceptual foundation for dealing with uncertainty. This means that we can understand what we are doing when we think about probability and utility of outcomes. Sometimes probability theory can be applied directly as a practical tool, as in in the sort of “utilitarian decision analysis” that I have promoted,³⁷ but that has not been my topic here. Rather, I have tried to show that we can understand probability as a personal judgment of degree of belief. People can differ, but each person can usually do no better than to use her own judgments of beliefs and values to make decisions, including decisions that affect others. This is true even when each person is a member of a group that will make the decision through some method of aggregation, such as voting. In this case, it also makes sense to ask whether our probability judgments are as useful as they could be.

I have also argued that misconceptions about probability have led to decisions and policies with potential or real harmful consequences. Many of these are equivalent to treating decisions as if some probabilities were 1 or 0. Other cases arise from thinking about probability as if it were objective.

References

- Baron J. (1993), *Morality and Rational Choice*, Kluwer, Dordrecht.
- Baron J. (1996), “Norm-Endorsement Utilitarianism and the Nature of Utility,” *Economics and Philosophy* 12 (2): 165–182.
- Baron J. (2004), “Normative Models of Judgment and Decision Making,” [in:] *Blackwell Handbook of Judgment and Decision Making*, D.J. Koehler, N. Harvey (eds), Blackwell, London: 19–36.
- Baron J. (2006), *Against Bioethics*, MIT Press, Cambridge (MA).
- Baron J. (2008), *Thinking and Deciding* (4th edition), Cambridge University Press, New York.
- Baron J. (2012), “The ‘Culture of Honor’ in Citizens’ Concepts of their Duty as Voters,” *Rationality and Society* 24 (1): 37–72.

³⁷ Baron (2006, 2008).

- Baron J., Frisch D. (1994), "Ambiguous Probabilities and the Paradoxes of Expected Utility," [in:] *Subjective Probability*, G. Wright, P. Ayton (eds), Wiley, Chichester (Sussex): 273–294.
- Baron J., Mellers B.A., Tetlock P.E., Stone E., Ungar L.H. (2014), "Two Reasons to Make Aggregated Probability Forecasts More Extreme," *Decision Analysis* 11 (2): 133–145.
- Baron J., Ritov I. (2009), "The Role of Probability of Detection in Judgments of Punishment," *Journal of Legal Analysis* 1 (2): 553–590.
- Baron J., Szymanska E. (2010), "Heuristics and Biases in Charity," [in:] *The Science of Giving: Experimental Approaches to the Study of Charity*, D. Oppenheimer, C. Olivola (eds), Taylor and Francis, New York: 215–236.
- Bazerman M.H., Baron J., Shonk K. (2001), *You Can't Enlarge the Pie: The Psychology of Ineffective Government*, Basic Books, New York.
- Bentham J. (1843/1948), *An Introduction to the Principles of Morals and Legislation*, Blackwell Publisher, Oxford.
- Brown R.V. (1993), "Impersonal Probability as an Ideal Assessment Based on Accessible Evidence: A Viable and Practical Construct?" *Journal of Risk and Uncertainty* 7 (2): 215–235.
- DeKay M.L. (1996), "The Difference Between Blackstone-Like Error Ratios and Probabilistic Standards of Proof," *Law and Social Inquiry* 21 (1): 95–132.
- Downs A. (1957), *An Economic Theory of Democracy*, Harper and Row, New York.
- Edlin A., Gelman A., Kaplan N. (2007), "Voting as a Rational Choice: Why and how People Vote to Improve the Well-being of Others," *Rationality and Society* 19 (3): 293–314.
- Edlin A., Gelman A., Kaplan N. (2008), "Vote for Charity's Sake," *The Economists' Voice* 5 (6).
- Ellsberg D. (1961), "Risk, Ambiguity, and the Savage Axioms," *Quarterly Journal of Economics* 75 (4): 643–699.
- Fox C.R. Ülkümen G. (2011), "Distinguishing Two Dimensions of Uncertainty," [in:] *Perspectives on Thinking, Judging, and Decision Making*, W. Brun, G. Keren, G. Kirkebøen, H. Montgomery (eds), Universitetsforlaget, Oslo.
- Hare R.M. (1963), *Freedom and Reason*, Oxford University Press (Clarendon Press), Oxford.
- Hare R.M. (1981), *Moral Thinking: Its Levels, Method and Point*, Oxford University Press (Clarendon Press), Oxford.
- Kaplow L., Shavell S. (2002), *Fairness versus Welfare*, Harvard University Press, Cambridge (MA).
- Kaplow L., Shavell S. (2004), "Any Non-Welfarist Method of Policy Assessment Violates the Pareto Principle: Reply," *Journal of Political Economy* 112 (1): 249–251.
- Kunreuther H., Hogarth R., Meszaros J. (1993), "Insurer Ambiguity and Market Failure," *Journal of Risk and Uncertainty* 7 (1): 71–87.
- Lindley D.V., Tversky A., Brown R.V. (1979), "On the Reconciliation of Probability Assessments" (with commentary), *Journal of the Royal Statistical Society: Series A* 142 (2): 146–180.
- Mill J.S. (1859), *On Liberty*, J.W. Parker & Son, London.

- Ritov I., Baron J. (1990), "Reluctance to Vaccinate: Omission Bias and Ambiguity," *Journal of Behavioral Decision Making* 3 (4): 263-277.
- Savage L.J. (1954), *The Foundations of Statistics*, Wiley, New York.
- Sunstein C.R. (2007), *Worst-Case Scenarios*, Harvard University Press, Cambridge (MA).
- Tetlock P.E., Mellers B.A., Scoblic J.P. (2017), "Bringing Probability Judgments into Policy Debates via Forecasting Tournaments," *Science* 355 (6324): 481-483.