Overweighting of Small Probabilities

Zach Burns*  
Andrew Chiu†  
George Wu‡

March 15, 2010

Entry for *Wiley Encyclopedia of*  

*Operations Research and Management Science*

* University of Chicago Booth School of Business, Center for Decision Research, 5807 S. Woodlawn Avenue, Chicago, IL 60637, zburns@chicagobooth.edu  
† University of Chicago Booth School of Business, Center for Decision Research, 5807 S. Woodlawn Avenue, Chicago, IL 60637, achiu@chicagobooth.edu  
‡ University of Chicago Booth School of Business, Center for Decision Research, 5807 S. Woodlawn Avenue, Chicago, IL 60637, wu@chicagobooth.edu
ABSTRACT

The psychological impact of a low probability or rare event is typically large relative to that event’s actuarial likelihood. This *overweighting* follows a two-stage process. First, rare events tend to be overestimated because of the availability heuristic, anchoring on the “ignorance prior,” and coarse chance categories. Second, when making decisions, low probability events are overweighted because of the “possibility effect”—decision makers are more sensitive to probability changes close to 0 than to probability changes away from 0.
Consider the decisions of whether or not to undergo heart surgery, evacuate a city when a tropical storm is approaching, or increase counterterrorist measures. These decisions involve consideration of both the likelihood and the impact of surgical death, a Category 3 hurricane, or a September 11-magnitude terrorist attack, respectively. Assessing the likelihood of these events is difficult, with experts often disagreeing on the chance that the relevant event will occur [1]. Rare events are also important when they involve extreme consequences, whether good or bad. Indeed, much of risk and policy analysis is devoted to the study of low probability, high consequence events, such as a nuclear accident or terrorist attack [2].

Research in judgment and decision making suggests that decision makers do not weight rare events according to their actuarial chances of occurring. Instead, small probability events tend to be overweighted for two reasons: (1) decision makers tend to overestimate the chance that rare events will occur; and (2) small probabilities are overweighted in terms of their impact on decisions. Combined, these two pieces suggest that rare events are given greater psychological weight in our minds than is normatively appropriate.

Although both parts contribute to the overweighting of small probability events, it is nevertheless important to distinguish these two separate processes. Events with a low probability of occurring may be overestimated, or perceived as more likely than they actually are. For example, someone considering a move to San Francisco may overestimate the probability that the Bay Area will suffer from an earthquake in the next 5 years. However, small probability events can loom large in decisions, even when these events are not overestimated. For example, consumers find lottery tickets attractive, even though they may be very much aware of the low chances of winning [3,4]. In this article, we review empirical findings from judgment and
decision making research that illustrate how these two stages contribute to the overweighting of small probability events.

1. A TWO-STAGE FRAMEWORK

In the classical theory of decision under uncertainty, subjective expected utility theory, the utility of each outcome is weighted by the decision maker’s subjective probability that this outcome will obtain (see eorms0312, Expected utility and subjective expected utility) [5,6]. Consider a prospect, $A = (E_1, x_1; \ldots; E_n, x_n)$, that offers outcome $x_i$ if event $E_i$ occurs, where events $E_i$ are mutually exclusive and collectively exhaustive. Subjective expected utility evaluates the prospect $A$ as follows,

$$U(A) = \sum_{i=1}^{n} P(E_i) u(x_i),$$

where the utility function, $u(x_i)$, measures the value the decision maker attaches to outcome $x_i$, and the subjective probabilities, $P(E_i)$, capture the decision maker’s assessment of the likelihood of event $E_i$ and satisfy the basic axioms of probability.

Although subjective expected utility is a standard building block for economic analysis [7], a large number of experimental studies have found that people are inconsistent with this model: subjective probability judgments violate the basic axioms of probability [8], and preferences are inconsistent with the axioms underlying expected utility theory [e.g. 9,10]. Of the numerous alternative models that have been proposed [11], prospect theory is consistent with the broadest set of empirical observations critiquing expected utility theory, such as the common-consequence effect (e.g., the Allais Paradox), the common-ratio effect, the four-fold pattern of risk preferences, and the simultaneous attraction of lottery tickets and insurance (see eorms0687,
Prospect Theory and eorms0636, Paradoxes and empirical violations of normative decision theory) [3,10-11]. Prospect theory uses a probability weighting function, \( \pi(p) \), to capture the distortion of probabilities (see eorms0681, Probability weighting functions), and a value function, \( v(x) \), to account for the evaluation of outcomes.

We extend prospect theory using a two-stage framework proposed by Tversky and Fox [12,13]. Consider a prospect that offers $100 if a thumbtack is flipped point down three times in a row (and $0 otherwise). The two-stage process suggests that the weight assigned to the outcome $100 reflects: (1) a decision maker’s subjective probability that a thumbtack will be flipped point down three consecutive times; and (2) a distortion of the subjective probability that captures the impact of that probability on the attractiveness of the prospect. This framework allows for bias at each of the two stages. More formally, consider a prospect that pays \( x \) if event \( E \) occurs (and 0 otherwise), denoted \((E,x)\) for simplicity. The weight assigned to outcome \( x \) is \( W(E) = \pi(P(E)) \), where \( P(E) \) captures stage 1 (the judgment of the likelihood of \( E \)) and \( \pi(P(E)) \) captures stage 2 (the distortion of \( P(E) \)) as measured by \( \pi(p) \), the probability weighting function. Overestimation of rare events occurs if \( P(E) \), the judged likelihood of event \( E \), is higher than the actuarial probability of \( E \), whereas overweighting of probability occurs if \( \pi(P(E)) > P(E) \).

In Section 2, we discuss overestimation in likelihood judgments, and identify some psychological factors that contribute to this phenomenon. We then discuss overweighting in decisions involving small probabilities in Section 3.
2. STAGE 1: OVERESTIMATION OF RARE EVENTS

2.1. Overestimation

Mortality risks have been commonly used to examine the perception of low probability events. During the last three decades, a number of studies have shown that participants consistently overestimate risks of death due to low probability causes, while underestimating risks of death from high probability causes [14-16]. Consider a classic study conducted by Lichtenstein and colleagues, where participants judged the frequency of lethal events in the U.S. Lichtenstein et al. compared the judged estimates to the actual death rates and found that low frequency events (such as smallpox, poisoning by vitamins, and botulism) were overestimated by a factor of 10, while high frequency events (such as stomach cancer, stroke, and heart disease) were underestimated [17].

In the next three sub-sections, we sketch three psychological mechanisms that drive overestimation of rare events. Our account of mechanisms is by no means comprehensive; the literature contains a number of other often complementary psychological accounts [18-20]. In addition, the overestimation described in this section is not universal. For example, because rare events are not always experienced, they are actually underestimated in some cases as a result [21]. Of course, when rare events are actually experienced, they tend to be overestimated, independent of the psychological mechanisms described below [22].

2.2. Availability

Tversky and Kahneman proposed that people deal with the cognitively difficult task of assigning probabilities by using simplifying heuristics [23]. If you are asked to estimate the divorce rate in your state, you may mentally work through your list of friends, and tally how many of these friends have been divorced (or may likely to become so soon). This example
illustrates the availability heuristic, whereby judged likelihood is based on the ease with which relevant instances come to mind. Although this heuristic often produces quite sensible probability estimates, it may in some cases lead to systematic biases. For example, two people with the same number of divorced friends may nevertheless judge the divorce rate differently. The availability heuristic suggests that a person whose best friends are divorced will tend to overestimate this rate relative to the person whose best friends are married.

Tversky and Kahneman provided a clear empirical demonstration of this heuristic [24]. One group of participants estimated the proportion of words in a typical English text with the letter “n” in the penultimate position. Another group estimated the proportion of words ending in “ing.” Although the former must be larger than the latter, participants overwhelmingly judged “-ing” words to be more common than “-n-” words. This is because it is much easier to recruit examples of “-ing” words than “-n-” words. Put differently, “ing” words are more psychologically available.

We suggest that the availability heuristic will generally give rise to the overestimation of small probability events. The most important low probability events involve extreme consequences, such as plane crashes or winning the lottery [2]. These events tend to be discussed disproportionately and are psychologically more salient, at least in part because high impact events receive media coverage disproportionate to the actual rate of occurrence. For example, Combs and Slovic found that newspapers tend to over-report deaths by homicides, accidents, and natural disasters, and under-report deaths from disease relative to their frequency [25]. In addition, accidents tend to be more dramatic and emotionally evocative, and thus easier to recall [17,26].

2.3. Anchoring on the “ignorance prior”
The French mathematician Laplace proposed a principle that has become known as the “principle of insufficient reason” [27]. Suppose someone is asked to provide probabilities that each of the 16 baseball teams in the American League will advance to the World Series. A non-American who is completely unfamiliar with baseball may lack a compelling reason to judge one team more likely to win than any other, and thus may conclude that each of the 16 teams has an equal chance (1/16) of winning the American League division. Of course, in many situations, individuals have some knowledge to use for making probabilistic judgments. Nevertheless, recent research has suggested that individuals often start by assigning equal probabilities to each of the possible events (the “ignorance prior”) and then make insufficient adjustment to each of these probabilities [28,29]. We suggest that this process will often result in overestimation of small probabilities.

We first present evidence that decision makers anchor on the ignorance prior. One implication of this process is a partition-dependence, in which dramatically different probability estimates are produced, depending on how the event space is partitioned. Fox and Clemen asked Duke MBA students to estimate the likelihood that a random MBA student would fall in each of several salary brackets [28]. In the low condition, there were four salary brackets below the median income of $85,000, and one above. In contrast, the high condition had one bracket below the median and four above that value. Overall, participants in the low condition estimated that there was a 75% chance that a student would earn less than $85,000, compared to 40% in the high condition. As predicted, these probabilities are close to the ignorance prior of 80% for the low condition and 20% for the high condition.

Anchoring and adjustment is a process in which decision makers start by “anchoring” on a particular quantity, even if this number is arbitrary or unrelated to the problem at hand, and
then adjust insufficiently away from the anchor [23]. Of course, in many situations the anchoring step is well-justified. For example, this year’s sales is usually a good starting point for estimating next year’s sales. However, the anchoring and adjustment process often leads to systematic biases. In one demonstration, Kahneman and Tversky asked participants to estimate the percentage of African countries in the United Nations. In the first stage, a wheel was spun with equal chance of it landing on numbers 0 to 100. Participants were then asked whether the actual percentage was higher or lower than the number on the wheel. In the second stage, participants were asked to estimate the actual probability. The correlation between the number spun and the eventual estimate was very high, even though it was clear to participants that the number was randomly generated, and had no informational value for the estimate at hand.

Anchoring and adjustment from the ignorance prior thus leads to overestimation of rare events when the number of events in the partition is not large. In most cases, the partition of events includes a small number of events, such as 2 (e.g., “bags will be lost in transport” or “bags will not be lost in transport”) or 3 (e.g., “the Pittsburgh Steelers win”, “the Pittsburgh Steelers lose”, or “the Pittsburgh Steelers tie”). Thus, rare events will be overestimated if a decision maker begins by assigning each event a $1/n$ chance, where $n$ is the number of events in the partition, and then adjusts insufficiently toward the “true probability.”

2.4. Coarse Chance Categories

A classic study by Piaget and Inhelder found that the typical 4-year old child divides chance events into three categories, “certainly will happen”, “certainly will not happen”, and “will possibly happen” [30]. Adults often act as if their categorization of the event space is similarly coarse. For example, Fischhoff and Bruine de Bruin found that participants provided a disproportionate number of probability judgments of 50%, even for mortality-related risks from
smoking-induced lung cancer and breast cancer (which are relatively small probabilities in the overall population), indicating equal probabilities of “will occur” and “will not occur” [31]. At the other extreme, professional weather forecasts and expert bridge players are relatively well-calibrated (i.e., events which they judge to have a 20% chance of occurring actually occur 20% of the time), partly because these experts have a finer categorization of the probability interval [32,33].

By definition, low probability events are encountered infrequently, and thus individuals do not have very much practice thinking about such events. Reyna and Brainerd suggest that people understand probabilities in “gists,” meaning that they form a general impression rather a precise probability of whether an event will occur [19]. In this way, it is likely that people form broad notions of probabilities in coarse categories instead of a specific number. Put differently, people seem to differentiate across chance categories, but not within [34]. Indeed, people seem to prefer communicating probabilistic information verbally rather than numerically, even though verbal statements are less precise [35]. As a result, rare events are lumped into a single category of “unlikely” outcomes. To the extent that “unlikely” corresponds to probabilities between .15 and .20 [36], this coarse categorization process will lead decision makers to overestimate rare events.

3. STAGE 2: OVERWEIGHTING OF SMALL PROBABILITIES IN DECISION MAKING

3.1. Prospect Theory

We next turn to evidence that small probabilities are overweighted in decision making. Research leading up to the development of prospect theory showed that individuals do not weight probabilities linearly, as would be predicted by expected utility theory. Like expected utility theory, prospect theory assumes that individuals evaluate each alternative and then choose
the alternative with the highest subjective valuation. Prospect theory also assumes that the decision maker values the possible outcomes, the probabilities of obtaining those outcomes, and then combines those valuations together. The major difference between prospect theory and the classical theory is that the valuation of outcomes is governed by the value function, \( v(x) \), while the valuation of the probabilities is governed by the probability weighting function, \( \pi(p) \). The value function is S-shaped, concave for gains and convex for loss, and steeper for losses than gains (see eorms0687, Prospect Theory). The second component, the probability weighting function, is the piece of prospect theory that captures the psychological phenomenon of overweighting of small probabilities.

We should note that small probabilities are not always overweighted. On the contrary, they may be ignored if the probability is below some threshold [37].

3.2. Psychophysics of the Probability Weighting Function

Prospect theory assumes that individuals do not weigh outcomes by their probability, as in expected utility theory, but by some distortion of probabilities. This distortion of probability is captured by prospect theory’s probability weighting function, \( \pi(p) \). The main psychological principle governing the distortion of probability is diminishing sensitivity away from a reference point. For probability, there are two obvious reference points, impossibility and certainty, or 0% chance and 100% chance. The distortion of probability captured by the probability weighting function captures the diminishing sensitivity away from these two reference points. People are most sensitive to changes in probability when they are near 0% or 100% than when the change applies to intermediate probabilities. Thus, most people regard the change from 0% to 1% as significant, because it changes the chance of winning from impossible to possible, a phenomenon known as the possibility effect. In addition, improving a 99% chance at winning by 1% is also
substantial, because the chance of winning has shifted from almost certain to certain, a phenomenon known as the certainty effect. In contrast, a chance from 32% to 33% is seen as inconsequential. Algebraically, both $\pi(0.01) - \pi(0)$ and $\pi(1) - \pi(0.99)$ are larger than $\pi(0.33) - \pi(0.32)$. More generally, a typical probability weighting function is concave for low probabilities and convex for medium to high probabilities (see Figure 1).

Tversky and Kahneman conducted a study that illustrates this nonlinearity [10]. Participants provided certainty equivalents for gambles that provided 5%, 50%, and 95% chances at winning $100 [10]. They found that the certainty equivalents for these gambles were $14, $36, and $78, respectively. The first 5% contributes $14 of value (or $2.80 for each percentage), while the next 45% is only worth $22 (or $0.48 for each percentage). The final 5% adds $22 in value (or $4.40 for each percentage), compared to the previous 45%, which contributes $42 (or $0.93 for each percentage).

Diminishing sensitivity implies that low probabilities are typically given more weight than they would receive using expected utility. Empirically, there is generally overweighting, $\pi(p) > p$, for $p < .35$ [38,39]. This overweighting, combined with an S-shaped value function, is consistent with risk-seeking for low probability gains (such as lottery tickets) and risk-aversion for low probability losses (such as insurance). (Note, however, that overweighting of probabilities is distinct from risk-seeking. For gains, concavity of the probability weighting function for low probabilities “competes” against concavity of the value function. Thus, risk-seeking is observed for probabilities around .10 or less [10].) Medium to high probabilities are typically given less weight than they would receive using expected value. Such underweighting is consistent with risk-aversion for medium to high probability gains, and risk-seeking for medium to high probability losses.
3.3. Affect and Risk

The shape of the probability weighting function also depends on the affective reactions associated with potential outcomes of a risky choice. Affect is the psychological concept that encompasses non-cognitive reactions to stimuli, such as arousal, emotions or moods. Some outcomes are relatively affect-rich, while others are relatively affect-poor. In a set of studies, the probability weighting function was more curved for affect-rich targets (such as kisses and electric shocks) than affect-poor targets (such as money) [40]. In particular, Rottenstreich and Hsee demonstrated a preference reversal in which participants preferred $50 in cash to “the opportunity to meet and kiss your favorite movie star,” but chose a 1% chance at a kiss over a 1% chance at $50. This study suggests that there is more overweighting of small probabilities for affect-rich outcomes than affect-poor outcomes.

The affective account is compelling because many of the risks that are overweighted in judgment are dramatic, fear-inducing and catastrophic, and hence affect-rich. While rare events that are affect-poor in nature may also be overweighted, events that are affect-rich are likely to be overweighted to a larger extent. In the aftermath of the attacks on 9/11, Vice President Dick Cheney set out what has become known as the One Percent Doctrine: “We have to deal with this new type of threat in a way we haven’t yet defined . . . With a low-probability, high-impact event like this . . . if there’s a one percent chance that Pakistani scientists are helping al Qaeda to build or develop a nuclear weapon, we have to treat it as a certainty in terms of our response.” [41] The One Percent Doctrine is an extreme example of overweighting of small probabilities, but less extreme risks are typically overweighted as well, although to a lesser extent (see eorms0652, Acceptable risk).
4. SUMMARY

The overweighting of small probability events is a result of a two-stage process. In the first stage of overestimation of rare events, decision makers take information they receive from the world and recruit from their memory and convert that information into a psychological representation of probability. As such, this stage is particularly susceptible to many of the basic cognitive limitations and shortcuts identified by psychological research, biasing the resulting judgments in a predictable manner. Though by no means an exhaustive list, we have identified several of the most important, including the availability heuristic, anchoring on the “ignorance prior,” and coarse chance categories. In the second stage of overweighting of small probability events, these (perhaps already inflated) probabilities are distorted upward due to the psychophysics of chance and affect.

Although eliminating these biases altogether is probably unrealistic, it is nevertheless possible to reduce these biases. Many of the most effective approaches will require identifying and then addressing the particular psychological mechanism that produced the biased probability judgment [42]. Consider, for example, the availability heuristic. To the extent that salient and vivid instances tend to come to mind and are thus overweighted in a decision, organizations might force decision makers to rely on historical actuarial evidence rather than what evidence they would naturally recall. Alternatively, organizations may develop “cognitive repairs” [43]. For example, a division at Motorola was overly focused on the problems of large customers, even though small customers constituted a large revenue base for the division. Motorola developed a process in which problems were surveyed more systematically and weighted by customer volume. In contrast, anchoring on the ignorance prior requires a different intervention, such as a thoughtful consideration of how to partition the space of events [28].
In the second stage, small probability events are additional inflated, as captured by the probability weighting function. Fortunately, this bias is easily addressed when decisions are made using the standard analytical methods of decision analysis [44]. In decision analysis, outcomes (or the utility of these outcomes) are multiplied by their probability of occurring, with the alternative with the highest expected value (or expected utility) selected.
BIBLIOGRAPHY


A typical prospect theory probability weighting function, $\pi(p)$, which is concave for small probabilities and convex for medium to large probabilities (and thus consistent with the principle of diminishing sensitivity).

FIGURE 1