4 Confidence

4.1 Probability

The development of probability theory below follows Kolmogorov (1933), as elaborated upon by authors such as Krantz, Luce, Suppes, and Tversky (1971) and Kreps (1988). We consider the simple case of discrete, finite sets of possibilities.

DEFINITION 4.1.1. Let $S = \{s_1, s_2, \ldots, s_n\}$ be a finite set of possible outcomes in a context C. S is a (finite) *sample space* for C iff exactly one outcome among the elements of S is or will be true in C.

EXAMPLE 4.1.2. Let C be the particular flipping of a coin. Then S = *{Heads, Tails}* is a sample space for C. Another sample space for C is S' = *{Heads is observed, Tails is observed, Cannot observe whether the coin is heads or tails}*. Yet another is S'' = *{Heads is observed and someone coughs, Heads is observed and no one coughs, Tails is observed whether someone coughs or not}*.

In what follows, we will assume the existence of a context without stating it.

DEFINITION 4.1.3. Let S be a sample space, and $\varnothing \neq E \subseteq 2^S$ (E is a nonempty subset of the power set of S, i.e., it is a set of subsets of S). Then E is an *event space* (or *algebra of events*) on S iff for every $A, B \in E$:
(a) $S \setminus A = A^C \in E$ (the S-complement of A is in E)
and
(b) $A \cup B \in E$ (the union of A and B is in E).
We call the elements of E consisting of single elements of S *atomic events*.

COROLLARY 4.1.4. If E is an event space on a sample space S, then $S \in E$.

EXERCISE 4.1.5. Prove 4.1.4.

EXAMPLE 4.1.6. If S = *{Heads, Tails}*, then $E = \{\varnothing$, *{Heads}, {Tails}, {Heads, Tails}*$\}$ is an event space on S. The atomic events are *{Heads}* and *{Tails}*.

DEFINITION 4.1.7. Let S be a sample space and E an event space on S. Then a function $P: E \rightarrow [0,1]$ is a (finitely additive) *probability measure* on E iff for every $A, B \in E$:
(a) $P(S) = 1$
and
(b) If $A \cap B = \varnothing$ (the intersection of A and B is empty, in which case we say that A and B are *disjoint* events), then $P(A \cup B) = P(A) + P(B)$ (*additivity*).
The triple $\langle S,E,P \rangle$ is called a (finitely additive) *probability space*.

COROLLARY 4.1.8. If $\langle S,E,P \rangle$ is a finitely additive probability space, then for all $A, B \in E$:

(a) $P(A^C) = 1 - P(A)$ (*binary complementarity*)
(b) $P(\oslash) = 0$
(c) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
Proof.
(a) $S = A \cup A^C$ by the definition of complementarity. Therefore $P(A \cup A^C) = 1$ by 4.1.7(a). A and A^C are disjoint by the definition of complementarity, so by 4.1.7(b), $P(A \cup A^C) = P(A) + P(A^C)$, so $P(A) + P(A^C) = 1$ and the result follows by subtracting $P(A)$ from both sides of the equation.
(b) $S^C = S \setminus S = \oslash$. Thus $P(\oslash) = P(S^C) = 1 - P(S)$ by 4.1.8a, which by 4.1.7a is 1-1 = 0.
(c) From set theory, we have $A = (A \cap B) \cup (A \cap B^C)$ and $B = (B \cap A) \cup (B \cap A^C)$, $A \cup B = (A \cap B) \cup (A \cap B^C) \cup (B \cap A^C) = (A \cap B) \cup [(A \cap B^C) \cup (B \cap A^C)]$, $(A \cap B) \cap (A \cap B^C) = \oslash$, $(B \cap A) \cap (B \cap A^C) = \oslash$, and $(A \cap B) \cap [(A \cap B^C) \cap (B \cap A^C)] = \oslash \cap \oslash = \oslash$. Therefore, $P(A) = P[(A \cap B) \cup (A \cap B^C)] = P(A \cap B) + P(A \cap B^C)$ and $P(B) = P[(B \cap A) \cup (B \cap A^C)] = P(B \cap A) + P(B \cap A^C)$, and $P(A \cup B) = P\{(A \cap B) \cup \{(A \cap B^C) \cup (B \cap A^C)]\} = P(A \cap B) + P(A \cap B^C) + P(B \cap A^C)$. Substituting, $P(A \cup B) = P(A \cap B) + [P(A) - P(A \cap B)] + [P(B) - P(B \cap A)]$. $B \cap A = A \cap B$, so $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

4.2 Confidence

DEFINITION 4.2.1. Let E be an event space on a sample space S. Then the *confidence* D (for "degree of belief"): $E \rightarrow \Re$ that an agent assigns to events in $F \subseteq E$ is *representable in probability* iff there is a probability measure P on E such that for all events A,B in F, $D(A) > D(B)$ iff $P(A) > P(B)$.

The above defines what it means for confidence judgments to be *qualitatively* consistent with a probability measure. A great deal of theory has been developed around necessary and sufficient conditions for representing confidence judgments in a probability measure. Representation theorems, although possible to prove, exist in multiple forms and are not as easy to interpret psychologically as in the case of preference relations (see Savage, 1954; Kraft, Pratt, and Seidenberg, 1959; Krantz et al., 1971; Suppes, 1976; Kreps, 1988). So, for now we will just provide the above definition and treat people's reported confidence when they are asked to map events onto the unit interval as a candidate for a probability measure, which may be rebutted if it violates probability theory. When confidence agrees with a probability measure, we may refer to the inferred confidence function as *subjective probability*, indicating that it obeys the conditions defining a probability measure. We will leave the question of whether subjective probabilities correspond to observed frequencies to the next section of this course, on Estimation.

DEFINITION 4.2.2. Let E be an event space on a sample space S. Then the *confidence* D (for "degree of belief"): $E \rightarrow [0,1]$ that an agent assigns to events is a *subjective probability measure* iff there is a probability measure P on E such that for all events A in E, $D(A) = P(A)$.

Definition 4.2.2 is more demanding than 4.2.1 in that 4.2.2 requires agreement on the full set of events in a probability space, while 4.2.1 can apply to any subset of an event space. When the event space is large, this has the consequence that 4.2.2 can in practice only be applied to disqualify confidence judgments from fitting a probability measure, rather than positively establishing them as subjective probabilities.

COROLLARY 4.2.2a. If a confidence function D is a subjective probability measure, then it is a

probability measure.
Proof. Follows from the existence of a probability measure P such that $D(A) = P(A)$ for all events in an event space.

The above corollary allows us to apply facts about probabilities directly to confidence judgments in order to ascertain whether a confidence function is a subjective probability measure without defining separate statements about subjective probabilities. When we apply the weaker criterion of representability in probability, however, we will be careful to separate criteria for confidence from criteria for probability measures.

EXPERIMENT 4.2.3. *Subadditivity*. A number of studies performed and cited by Tversky and Koehler (1994) show that people violate the additivity condition of Def. 4.1.7. For confidence ratings, this would imply that $D(A \cup B) = D(A) + D(B)$. The strongest and most consistent violations occur for what are called "implicit disjunctions". Subjects were divided into two groups, one of which evaluated the components of a disjunctive hypothesis such as the chances of "death resulting from heart disease, cancer, or some other natural cause", while the other evaluated an implicit hypothesis that was coextensional with (containing the same cases as) the disjunction, e.g. the chances of "death resulting from natural causes" for people who had died during the previous year. The authors report, "for both probability and frequency judgments, the mean estimate of an implicit disjunction (e.g. death from a natural cause) is smaller than the sum of the mean estimates of its components (heart disease, cancer, or other natural causes)... Specifically [for probabilities], the former equals 58%, whereas the latter equals 22% + 18% + 33% = 73%." For frequencies, the mean estimate for the implicit hypothesis was 56%, while the sum of mean estimates for the disjuncts was 67%. A similar pattern obtained for death from "unnatural causes" (or, in the disjunctive version, "accident, homicide, or some other unnatural cause". The amount of subadditivity increases as the number of disjuncts increased, so that when, for example, cancer was "unpacked" into different types of cancer, the sum of probabilities (or frequencies) assigned to cancers was far greater than for just "cancer". Evidence also exists for "explicit" subadditivity, wherein the sum of probabilities assigned to disjuncts is greater than the probability assigned to an explicit disjunction, but this effect is weaker observed than the implicit version (Rottenstreich and Tversky, 1997).

EXPERIMENT 4.2.4. *Binary complementarity*. A number of experiments reported in Tversky and Koehler (1994) and Wallsten, Budescu, and Zwick (1992) show that subjects' confidence judgments generally obey 4.1.8(a), so that $D(A^C) = 1 - D(A)$. The latter authors, according to Tversky and Koehler, "presented subjects with 300 propositions concerning world history and geography (e.g. 'The Monroe Doctrine was proclaimed before the Republican Party was founded')... True and false (complementary) versions of each proposition were presented on different days." The average sum probability for propositions and their complements was 1.02, which is insignificantly different from 1. Some evidence indicates violations of binary complementarity under limited conditions (ref?).

THEOREM 4.2.4. *Mirroring principle*. If a confidence function D is representable in probability, then for any events A and B in the event space, $D(A) > D(B)$ iff $D(\neg A) < D(\neg B)$.
Proof. (a) *Only if* direction: If $D(A) > D(B)$, then by 4.2.1, there is a probability measure P such that $P(A) > P(B)$, so $-P(A) < -P(B)$, and adding 1 to each side, $1-P(A) < 1-P(B)$, which by 4.1.8(a) implies $P(\neg A) < P(\neg B)$, so by 4.2.1 again, $D(\neg A) < D(\neg B)$. (b) The proof in the *if* direction goes analogously.

EXPERIMENT 4.2.5. *Violation of the mirroring principle – familiarity bias*. Studies reported by Fox and Levav (2000) showed that qualitative likelihood judgments (e.g. "A is more likely than B", "C is less likely than C") can violate the mirroring principle between subjects when one proposition and its complement (negation) are more familiar to subjects than are both the other proposition and that proposition's negation. In one example, 64% of subjects at Duke University said it was more likely that "the winner of the next U.S. presidential election is a member of the Democratic Party" (high familiarity) than that "the winner of the next British Prime Ministerial election is a member of the Labor Party" (low familiarity), but 76% said it was more likely that "the winner of the next U.S. presidential election is *not* a member of the Democratic Party" (high familiarity) than that "the winner of the next British Prime Ministerial election is *not* a member of the Labor Party" (low familiarity). In another example, 59% said it was more likely that "Georgia Tech beats USC in men's basketball tonight" (high familiarity) than that "Washington State beats Washington in men's basketball tomorrow night" (low familiarity), but 74% said it was more likely that "USC beats Georgia Tech in men's basketball tonight" (high familiarity) than that "Washington beats Washington State in men's basketball tomorrow night" (low familiarity). This is presented as an example of a *familiarity bias*, in which people judge familiar events as more likely than unfamiliar events. The result is a violation of the mirroring principle of 4.2.4.

A common assumption in studying judgment and decision making is that inferred inequalities between confidence levels, valuations, and so on, will be invariant under different procedures for eliciting the same inequality. This has come to be known in the literature as *procedural invariance.* We have already seen an example of a violation of procedural invariance in 3-Preference, namely the preference reversal experiments that found people's preferences can switch from xPy to yPx when the elicitation procedure changes, e.g. from choosing to pricing or from accepting to rejecting. Below we define procedural invariance for confidence.

DEFINITION 4.2.6. An agent's confidence D is *procedurally invariant* with respect to two procedures Π and Π' iff the inferred inequality relations $>_{\Pi}$ and $>_{\Pi'}$ are such that for all events A and B, $D(A) >_{\Pi} D(B)$ iff $D(A) >_{\Pi'} D(B)$.

Procedural invariance has strong normative force, and indeed is taken to be so obvious that it is not even stated as an assumption in theories of probability and utility. If an agent reports that they believe one proposition more than another under one elicitation procedure, then it would be hard to justify a switch to the previously less believed proposition merely because the question is asked differently. Yet, the violations of mirroring cited in 4.2.5 appear also to be violations of procedural invariance, because the same two propositions are being compared twice with a different implicit ordering of their likelihoods, one under a procedure that compares two propositions directly, and another under a procedure that compares their negations. This phenomenon has been termed "belief reversal", and it can also occur as a result of a switch from qualitative to quantitative elicitation, as the experiment below shows.

EXPERIMENT 4.2.7. *Belief reversal.* Fox and Levav (2000) conducted studies on the same propositions described in 4.2.5 under two different elicitation procedures. In one (labeled μ), subjects were asked: "Which of the following two events do you think is *more likely* to occur?". In the other procedure (labeled P), subjects were asked: "Please indicate your best estimates of the probabilities (0-100%) of each of the following two events." As noted in 4.2.5, 64% of subjects said it was "more likely" (procedure μ) that "the winner of the next U.S. presidential election is a member of the Democratic Party" (high familiarity) than that "the winner of the next British Prime Ministerial election is a member of the Labor Party" (low familiarity). Under the P procedure, however, only 36% of subjects assigned higher probability to the more familiar event. The μ and P procedures produced similar belief reversals for one of the basketball game comparisons (from 59% under μ to 24% under P) and for another problem involving predicting a legal decision (from 66% under μ to 49% under P). The phenomenon did not occur for all high-low familiarity comparisons, however. Its presence in some cases appears to be an effect of attenuation that the quantitative procedure has on the familiarity bias. The authors propose that "evidence for the alternative hypothesis [the one not being asked about] looms larger in judgments of probability than in judgments of relative likelihood" (p. 287).

EXERCISE 4.2.8. What psychological mechanism might explain why familiar events are judged more likely than unfamiliar ones?

THEOREM 4.2.9. *Conjunction rule*. If a confidence function D is representable in probability, then for any events A and B in the event space, $D(A \cap B) \leq D(A)$.
Proof. From set theory, we have $A = (A \cap B) \cup (A \cap B^C)$, and $(A \cap B) \cap (A \cap B^C) = \emptyset$, so for a probability measure, by 4.1.7(b), $P(A)=P[(A \cap B) \cup (A \cap B^C)]=P(A \cap B)+P(A \cap B^C)$. By the definition of probability, $P(A \cap B^C) \geq 0$, so $P(A) - P(A \cap B) = P(A \cap B^C) \geq 0$. Therefore $P(A) \geq P(A \cap B)$. Since D is representable in probability by hypothesis, then $P(A \cap B) \leq P(A)$ implies that $D(A \cap B) \leq D(A)$.

EXPERIMENT 4.2.10. *Conjunction fallacy.* Tversky and Kahneman (1983) reported a number of studies in which subjects' probability and frequency judgments violated the conjunction rule. In one study, subjects evaluated the relative likelihoods that Bjorn Borg, who was then the most dominant male tennis player in the world, would (a) win the final match at Wimbledon, (b) lose the first set, © lose the first set but win the match, and (d) win the first set but lose the match. The average rankings (1=most probable, 2 = second most probable, etc.) were 1.7 for a, 2.7 for b, 2.2 for c, and 3.5 for d. Thus, subjects on average ranked the conjunction of Borg losing the first set and winning the match as more likely than that he would lose the first set (2.2 versus 2.7). The result has been replicated on a number of other examples, both within- and between-subjects. In this case, the authors' explanation is that people rank likelihoods based on a *representativeness heuristic* which makes the conjunction of Borg's losing the first set but winning the match more representative of Borg than is the proposition that Borg loses the first set. Another study involved natural disasters. Tversky and Kahneman asked subjects to evaluate the probability of occurrence of several events in 1983. Half of subjects evaluated a basic outcome (e.g. "A massive flood somewhere in North America in 1983, in which more than 1,000 people drown") and the other half evaluated a more detailed scenario leading to the same outcome (e.g. "An earthquake in California sometime in 1983, causing a flood in which more than 1,000 people drown.") The estimates of the conjunction were significantly higher than those for the flood. Thus, scenarios that include a cause-effect story appear more plausible than those that lack a cause, even though the latter are extensionally more likely. The causal story makes the conjunction easier to imagine, an aspect of what is known as the *availability heuristic.*

THEOREM 4.2.11. *Disjunction rule*. If a confidence function D is representable in probability, then for any events A and B in the event space, $D(A \cup B) \geq D(A)$.

EXERCISE 4.2.12. Prove 4.2.11.

EXERCISE 4.2.13. Construct an example experiment in which you would expect subjects to violate the disjunction rule.

4.3 Conditional Probability

DEFINITION 4.3.1. The *conditional probability* $P(A|B)$ of an event A given event B is defined as follows: $P(A|B) = P(A \cap B) / P(B)$.

EXERCISE 4.3.2. Formulate and prove the conjunction rule for conditional probability.

EXPERIMENT 4.3.3. *Conjunction fallacy for conditional probability – Linda problem.* Tversky and Kahneman (1983) report a famous example in which "Linda" is described as follows: "Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations." Subjects ranked the likelihood of propositions that included "Linda is a bank teller" and "Linda is a bank teller and active in the feminist movement." In this example, the description of Linda serves as *evidence* on which the conjunction and its component judgments are based, and can therefore be described as a conditional confidence in *hypotheses* about Linda given a description of Linda. The vast majority of subjects evaluated the conjunction as more likely than the hypothesis that Linda is a bank teller, violating the conditional conjunction rule. The authors attribute this to judgments by representativeness, since "feminist bank teller" is more representative of Linda, based on her description, than is "bank teller".

THEOREM 4.3.4. *Bayes's rule.* For events A and B, $P(A|B) = [P(B|A)P(A)] / P(B)$.
Proof. By 4.3.1, $P(A|B) = P(A \cap B) / P(B)$. $A \cap B = B \cap A$, so $P(A \cap B) = P(B \cap A)$. From 4.3.1, then, we can derive $P(B \cap A) = P(B|A)P(A)$, the theorem follows by substitution.

EXAMPLE 4.3.5. *Applying Bayes's rule to medical diagnosis.* Eddy (1982) gives the following example calculation of the probability that a breast lesion is cancerous based on a positive mammogram . From the literature, Eddy estimates the probability that a lesion will be detected through a mammogram as .792. Hence the test will turn up negative when cancer is actually present 20.8% of the time. When no cancer is present, the test produces a positive result 9.6% of the time (and is therefore correctly negative 90.4% of the time). The key fact is the prior probability that a patient who has a mammogram will have cancer, which is taken to be 1%. Thus, Eddy calculates the probability of cancer given as positive test as [(.792)(.01)] / [(.792)(.01) + (.096)(.99)} + .077, applying 4.3.4, so a patient with a positive test has less than an 8% chance of having breast cancer. Does this seem low to you?

EXPERIMENT 4.3.6. *Violations of Bayes's rule – cab problem.* Tversky and Kahneman (1982) gave subjects the following problem:

> A cab was involved in a hit and run accident at night. Two cab companies, the Green and the Blue, operate in the city. You are given the following data:
> (a) 85% of the cabs in the city are Green and 15% are Blue.
> (b) a witness identified the cab as Blue. The court tested the reliability of the witness under the

same circumstances that existed on the night of the accident and concluded that the witness correctly identified each one of the two colors 80% of the time and failed 20% of the time.

What is the probability that the cab involved in the accident was Blue, rather than Green?

We can apply Bayes's rule to obtain the correct probability $P(B|E)$ that the cab was blue given the evidence, i.e. that the witness identified it as such. The answer is 12/29 = .41. Subjects consistently give both modal and median estimates, however, of .80, which is actually $P(E|B)$. Tversky and Kahneman point out that people appear to neglect the prior probability that the cab is blue, which is .15. This prior probability, when factored into Baye's rule, diminishes the likelihood that the cab was actually blue. In a variation of the experiment, item (b) was omitted from the problem, and "almost all subjects [correctly] gave the base rate (.15) as their answer." In another variation, "a witness identified the cab as Blue" was replaced by "a witness identified the color of the cab", and respondents were then asked, "What is the probability that the witness identified the cab as blue?" The authors relate: "The median and modal response to this question was .15. Note that the correct answer is $.2 \times .85 + .8 \times .15 = .29$. In the absence of other data, therefore," the authors conclude, "the base rate was used properly to predict the target outcome and improperly to predict the witness's report." These different versions of the experiment indicate that subjects think simplistically about bayesian updating problems. When information about the likelihood of a hypothesis giving rise to particular evidence is provided and is obviously relevant to the posterior probability, people tend to substitute the likelihood for the posterior probability and ignore prior probabilities. When information about likelihoods is unavailable or is not obviously irrelevant, people use prior probabilities and ignore likelihoods. They don't easily combine data from the two sources to modify each other, as Bayes's rule requires.

EXERCISE 4.3.7. Show using Bayes's rule why the correct answer to the cab problem is .41.

COROLLARY 4.3.8. *Bayes's rule – odds form.* For events A and B,
$P(A|B) / P(\neg A|B) = [P(B|A) / P(B|\neg A)] \times [P(A) / P(\neg A)]$. This may also be written as $O(A|B) = R(B|A)\ Q(A)$, where O is the *posterior odds*, R is the *likelihood ratio*, and Q is the *prior odds.*
Proof. Follows by expanding both the numerator and denominator of $P(A|B) / P(\neg A|B)$ using 4.3.4, and canceling $P(B) / P(B)$ in the result.

EXPERIMENT 4.3.9. *Base rate neglect.* Kahneman and Tversky (1973) gave subjects five personality descriptions that were said to be draw from a collection including either 30 engineers and 70 lawyers (low-engineer) or 70 engineers and 30 lawyers (high-engineer). Subjects were asked to "please indicate your probability that the person described is an engineer, on a scale from 0 to 100." Subjects were offered a bonus for estimates that came close to those of an expert panel who were said to be "highly accurate in assigning probabilities to the various descriptions." An example description was as follows:

Jack is a 45-year old man. He is married and has four children. He is generally conservative, careful, and ambitious. He shows no interest in political and social issues and spends most of his free time on his many hobbies which include home carpentry, sailing, and mathematical puzzles.

Subjects were asked to judge the "probability that Jack is one of the 30 [70] engineers in the sample of 100". Although Bayes's rule does not determine a correct answer for a given description, the odds-form does permit a normative calculation of the ratio between the posterior odds in the high-engineer condition O_H and the posterior odds in the low-engineer condition O_L :

$$O_H/O_L = (Q_H/Q_L)(R/R) = (Q_H/Q_L) = (70/30)/(30/70) = 5.44$$

Thus, as the authors say, “the correct effect of the manipulation of prior odds can be computed without knowledge of the likelihood ratio.” The figure below, reproduced from Kahneman and Tversky (1973), shows the median probabilities given for each description under the two prior-odds conditions. The curved line represents the bayesian prediction. The square symbol on that line represents subjects' estimates when “given no information whatsoever about an individual chosen at random from the sample” (the so-called “null description”). In that case, subjects correctly judge that the posterior probability is determined by the prior odds. In all five of the non-null description cases, however, subjects' responses fall near the identity (diagonal) line, which represents complete determination of posterior odds by the likelihood ratio and neglect of the prior odds. Kahneman and Tversky note: “The effect of of prior probability, although slight, is statistically significant. For each subject the mean probability estimate was computed over all cases except the null. The average of these values was 50% for the low-engineer group and 55% for the high-engineer group (t = 3.23, df = 169, p < .01).”

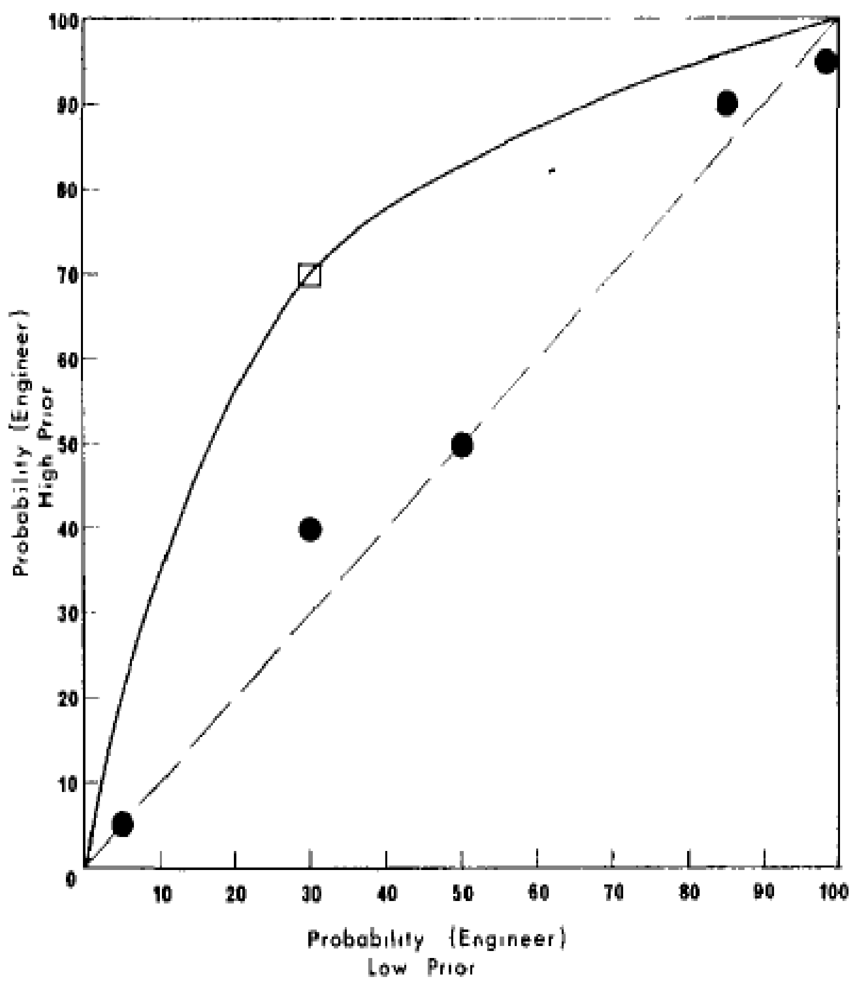

FIG. 1. Median judged probability (engineer) for five descriptions and for the null description (square symbol) under high and low prior probabilities. (The curved line displays the correct relation according to Bayes' rule.)

4.4 Independent Events

DEFINITION 4.4.1. Two events A and B are *independent* iff $P(A \cap B) = P(A)P(B)$.

COROLLARY 4.4.2. Two events A and B satisfying $P(B)>0$ are *independent* iff $P(A|B) = P(A)$.
Proof. (a) *Only if* direction: Since A and B are independent, $P(A \cap B) = P(A)P(B)$ by 4.2a.1. Since $P(B)>0$, $P(A|B) = P(A \cap B)/P(B) = P(A)P(B)/P(B) = P(A)$. (b) *If* direction: $P(A|B) = P(A)$, so multiplying both sides by $P(B)$, $P(A|B)P(B) = P(A)P(B) = P(A \cap B)$.

EXAMPLE 4.4.3. Consider two flips of a coin and let A=*{Heads on the first toss}* and B=*{Tails on the second toss}*. The probability that the coin lands on tails on the second toss is not affected by what happened on the first toss and vice versa, so $P(B|A) = P(B)$ and $P(A|B) = P(A)$. Assuming both sides of the coin have a nonzero probability of landing on top, this argument suffices to establish that the tosses are independent and that, therefore, $P(A \cap B) = P(A)P(B)$. Assuming the coin is unbiased ($P(\{Heads\})=P(\{Tails\})=0.5$), this means that $P(A \cap B) = (.5)(.5) = .25$.

EXERCISE 4.4.4. Consider two six-sided dice (with faces varying from 1 to 6 dots) which are rolled simultaneously, and assume each roll is independent of the other. What is the probability that the sum of the two dice is 7?

The concept of independence can be generalized to an arbitrary number of events with consequences analogous to the two-event case. We will not bother to define the more general version here, but we will apply it to experiments that test whether confidence judgments obey independence.

EXPERIMENT 4.4.5. *Violation of independence – gambler's fallacy.* Tversky and Kahneman (1974) report that subjects on average regard the sequence H-T-H-T-T-H of fair coin tosses to be more likely than the sequence H-H-H-T-T-T, even though both sequences are equally likely, a result that follows from the generalized definition of event independence. An extensive literature indicates that people in general regard, for example, a heads toss as more likely after a long run of tails tosses than after a similar run of tails or a mixed run. This tendency has been called the "gambler's fallacy". People appear to believe in the so-called "law of averages", a folk concept which allows confidence judgments to be at odds with the consequences of event independence. Tversky and Kahneman's explanation is that people expect sequences of tosses to be representative of the process that generates them. H-T-H-T-T-H is rated more likely than H-H-H-T-T-T because the former sequences is more typical of fair coin toss sequences generally, in which H and T are not grouped such that all the Hs precede all the Ts, but rather Hs and Ts are intermingled.

APPLICATION 4.4.6. *Historian commits the gambler's fallacy.* Fischhoff (1982) quotes the historian R.J. Morrison in a passage in which Morrison exhibits the gambler's fallacy. Morrison (1977) showed that the probability of a U.S. Supreme Court vacancy in any given year historically had been about .39, and that the probability each year was independent of whether there had been a vacancy the previous year. Nonetheless, Morrison writes the following about Franklin Roosevelt:

> Roosevelt announced his plan to pack the Court in February, 1937, shortly after the start of his fifth year in the White House. 1937 was also the year in which he made his first appiontment to the Court. That he had this opportunity in 1937 should come as no surprise, because the

> probability that he would go five consequtive years without appointing one or more justices was but .08, or one chance in twelve. In other words, when Roosevelt decided to change the Court by creating additional seats, the odds were already eleven to one in his favor that he would be able to name one or more justices by traditional means that very year (pp. 143-144).

Fischhoff writes: "However, if vacancies do appear at random, then this reasoning is wrong. It assumes that the probabilistic process creating vacancies, like that governing coin flips, has a memory and a sense of justice, as if it knows that it is moving into the fifth year of the Roosevelt presidency and that it "owes" FDR a vacancy. However, on January 1, 1937, the past four years were history, and the probability of at least one vacancy in the coming year was still .39" (p. 344).

EXERCISE 4.4.7. Users of the iPod Shuffle have reported that they think songs are played nonrandomly, i.e. songs too often appear that are related to previous songs, or even the same. Explain how this judgment could come about as a result of the representativeness heuristic.