In press. D. J. Koehler and N. Harvey (Eds.) Blackwell Handbook of Judgment and Decision Making.

Inside and outside probability judgment

David Lagnado
University College London

Steven Sloman
Brown University

Please address correspondence to:

David Lagnado
Department of Psychology
University College London
Gower St.
London WC1E 6BT

Email: d.lagnado@ucl.ac.uk
Steven_Sloman@brown.edu

The Inside/Outside distinction

Four brothers have gone for a day at the races. They are preparing to bet on the big race. Harpo is a complete novice. He knows that there are eight horses running, but nothing much else. He figures that the chance of each horse winning is 1/8. Chico is also a novice, but he notices that one of the horses is called ‘Sure Thing’. He figures that this horse has got to win and puts all his money on it. Zeppo is a race track expert (he organized the trip). He has pored over the Racing Post all morning, looking at the previous form of each horse. He manages to identify ten past races very similar to the big race today, all with the same eight horses. ‘Best Shot’ has won eight out of these ten races, so he figures its chances are 8/10. Finally, Groucho is also an expert. Indeed he had dinner with the stable boy last night (he paid). He happens to know that ‘Some Dope’ has been given a new wonder drug that pretty much guarantees it will win. He places his bets accordingly.

Ignoring the relative merits of each strategy for the moment, we can classify these four probability judgments into two broad classes. Harpo and Zeppo, despite their difference in expertise, are both reasoning from the *outside*. They are thinking of the outcome as a member of a set of events or possibilities, and basing their judgments on an appropriate proportion of these. Chico and Groucho, despite their differences, are reasoning from the *inside*. They are basing their judgments on knowledge (or belief) about the properties of a specific horse. The Marx brothers’ two strategies have roots that can be traced back both to efforts to understand the foundations of probability in philosophy and to efforts to understand human probability judgment in psychology.

The epistemic/aleatory distinction in the philosophy of probability

Ever since its inception, the formal notion of probability has been interpreted in two main senses: either in terms of reasonable degrees of belief or in terms of statistical distributions within classes of events. Hacking (1975) terms the former sense *epistemic*, because it concerns states of knowledge or belief, and the latter *aleatory,* because it concerns frequencies or proportions generated by stochastic processes in the world. Philosophical analyses of probability have reinforced this distinction (Carnap, 1950; Ramsey, 1931). Although some theorists have argued for the exclusivity of one approach over the other, the general consensus is that both are important (Gillies, 2001; Hacking, 2001). Certainly from a normative point of view, both provide valid interpretations of the probability calculus. On the epistemic view, the laws of probability furnish laws of *coherence* for our degrees of belief; to violate these laws lays a person open to a Dutch book. That is, if you bet in accordance with an incoherent set of beliefs an opponent can win money from you regardless of the actual outcomes of the events bet upon (Ramsey, 1931). On the statistical view, the laws consist in combination rules for relative frequencies.

The epistemic/aleatory distinction outlined here concerns valid interpretations of the probability calculus, and in that sense speaks to issues of normativity. It is the *ideal* reasoner that has perfectly coherent beliefs or attunes their judgments to the appropriate relative frequencies. However, the philosophical debate has also focused on how people actually employ probability judgments. Do we understand talk of probability in terms of expressions of confidence in what will happen or in terms of relative frequencies? Once again the received opinion is that both interpretations have a certain domain of applicability. Sometimes we use probability statements to express the degree of support our evidence lends to it, on other occasions we use them to refer to a proportion in a class of events. Neither reading alone is sufficient to capture all important aspects of usage (see Hacking, 2001).

A parallel distinction appears in people's judgments of probability. This idea is apparent in early work by Meehl (1954; 1986) on clinical prediction, and also underpins the general framework endorsed by Kahneman and Tversky (1982).

Clinical vs. Statistical prediction

Meehl (1954) was concerned with a basic methodological issue in clinical psychology, to predict how an individual will behave (e.g., Will they re-offend once released from prison? Will they commit suicide due to their depression?). Meehl contrasts 'clinical' prediction, which involves an assessment of the individual case at hand, and attempts to isolate the causally relevant factors that determine subsequent behaviour, with statistical or actuarial prediction, in which a person is assigned to a class of like individuals, and a statistical table of relevant frequencies is used to infer future behaviour. Dawes (1996) summarises a wealth of research showing the superiority of the statistical approach, despite strongly-held beliefs to the contrary within the clinical community: 'People's behaviour and feelings are best predicted by viewing them as members of an aggregate and by determining what variables generally predict for that aggregate and how. That conclusion contradicts expert's claims to be able to analyze an individual's life in great detail and determine what caused what.' (Dawes, 1996, p. 101).

Although concerned with prediction rather than probability judgment per se, and with experts rather than laypeople, the contrast between clinical and statistical methods suggests that a similar distinction may operate in our everyday predictive activities. It also cautions us that despite our apparent 'expertise' in everyday judgment situations (in terms of years of experience and self-confidence), we too may suffer from the neglect of a more statistical approach. Indeed this is exactly the perspective adopted by Kahneman and Tversky (1982) when they introduced the inside/outside distinction.

The Inside vs. Outside view

In their essay 'Variants of Uncertainty' (1982), Kahneman and Tversky maintain that people reach judgments of probability in a variety of ways. In particular, they distinguish a *distributional* mode, where 'the case in question is seen as an instance of a class of similar cases, for which the relative frequencies of outcomes are known or can be estimated', with a *singular* mode, 'in which probabilities are assessed by the propensities of the particular case at hand.' This is illustrated by consideration of the 'planning fallacy' whereby people estimate the time of completion of a project on the basis of factors specific to that particular project, and neglect available information about how long similar projects have taken to complete. In the light of various experimental studies they conjecture that people tend to use the singular mode, even though it is more likely to lead to erroneous judgments. Thus 'people generally prefer the singular mode, in which they take an 'inside view' of the causal system that most immediately produces the outcome, over an 'outside view', which relates the case at hand to a sampling schema' (1982, p. 518).

Note that the common conclusion that a distributional (or statistical) approach leads to more accurate judgments is an *empirical* claim. It is not that adoption of the singular mode is substandard a priori, but that it can result in substandard judgments. The main reason for this is that an inside view will by definition tend to ignore distributional information, and this can be critical to valid probability judgment.

Adopting the basic insight and terminology introduced by Tversky and Kahneman we will characterize an inside perspective on probability judgment as one that focuses on the individual case, and its attendant properties. A judgment of probability is reached by assessing the relation between this case and the to-be-judged category or outcome. In contrast, an outside perspective considers some set or reference class of instances that the individual case belongs to. The distributional properties of this set form the basis for a judgment of probability. To illustrate, consider the case of Hilary, who is applying for a prestigious job vacancy. What are her chances? An inside view will focus just on Hilary's qualities (intelligence, loyalty) and suitability for the post. A judgment of probability will be reached on the basis of how well these features fit the job specification and selection process. In contrast, an outside view would consider one or more sets that Hilary is a member of - the set of other candidates, the set of occasions on which Hilary has applied for previous jobs, etc. Knowledge about distributions over these sets would inform an outside probability judgment.

A survey of the modern literature on probability judgment reveals two types of models that, for the most part, can be classified as taking an inside or an outside view. However, the relation between inside and outside views is complicated in two ways. First, sometimes people try to view a category from the outside -- by examining instances

-- but the sheer number of category instances makes counting ineffective. In such a case, the instances retrieved may be treated as a representative sample and may be subject to property-based comparison (similarity assessment, a typical inside view operation) in order to arrive at a probability judgment. Second, outside views of categories reveal inclusion relations amongst categories that are hidden from the inside view. Therefore, a variety of phenomena of probability judgment reflect the effect of variables that cause people to change from an inside to an outside perspective, and generally also involve an improvement in performance.

Our discussion will divide into two parts. In the first we will concentrate on models and phenomena that reveal the operation of the inside perspective. In the second we will consider models that take a broadly outside perspective, in so far as they involve the explicit consideration of sets of instances. However, while some of the models in this section adopt a purely outside view, others admit of varying degrees of contamination through inside factors.

Inside models

Inside-view models of probability judgment take two forms: those that consider only properties and those that consider both properties and their interrelations. To illustrate, consider a model that states that the judged probability of an event is proportional to the typicality of the event in some relevant category (e.g., the typicality of rain in the category weather). If typicality is defined in terms of the number of properties that an event has in common with other category members (following Rosch's, 1973, definition), then such a model relies only on property overlap to model probability. But typicality can also be defined in terms of relations amongst properties. Such a theory might posit, for example, that typicality reflects the degree to which an event embodies the causal mechanisms that normally govern category members (rain might also be seen as typical because it does not violate the causal principles that normally govern the weather unlike, for example, hail). These causal principles relate properties to one another. In the case of rain, they relate properties like being a liquid and temperature. We consider models that consider only properties and models that also consider property interrelations in turn.

Similarity: probability judgment from property overlap

The representativeness heuristic of probability judgment has been defined in a variety of ways, starting with the earliest discussions by Kahneman and Tversky (1973). Its dominant and most frequent sense concerns similarity (Kahneman & Frederick, 2002). An event is judged representative, and therefore probable, to the degree that it is similar to a model of the event being judged. Sylvain's judged probability that John would make a good husband for Mary reflects Sylvain's judged similarity between John and his model of the kind of man that would make a good husband for Mary.

Evidence for the representativeness heuristic as defined in terms of similarity is vast and broad. Many studies have shown how representative judgment can lead to the neglect of base rate information, violations of basic laws of probability (conjunction and disjunction errors), and a near perfect correlation between people's judgments of similarity and probability (for a recent review see Kahneman & Frederick, 2002). For example, Bar-Hillel and Neter (1993) presented students with the following kind of question:

> Danielle is sensitive and introspective. In high school she wrote poetry secretly…Though beautiful, she has little social life, since she prefers to spend her time reading quietly at home rather than partying. What does she study?

Participants then ranked a list of subject categories according to one of several criteria such as probability, suitability or willingness to bet. The lists included nested subordinate-superordinate pairs (e.g., in the case of Danielle both ‘Literature’ and ‘Humanities’) specifically designed so that the character profile fitted the subordinate category better than the superordinate. There were two main findings. First, people consistently ranked the subordinate category as more probable than the superordinate, in violation of the extension law of probability (whereby a subordinate cannot be more probable than a superordinate category that contains it). Bar-Hillel and Neter termed this a disjunction fallacy. Second, probability rankings were almost perfectly correlated with both suitability and willingness-to-bet rankings (and in a subsequent experiment with actual betting behavior). This suggests that participants in the different judgment conditions used the same underlying process, and one that was not based on an extensional understanding of probability.

Such findings illustrate people’s propensity to take an inside view and neglect the distributional properties of the problem situation. The character profile of Danielle conjures up a stereotype of a sensitive Literature student, and people base their judgments on this picture rather than the structural fact that Literature is a subset of the Humanities. The plausibility of this kind of process, however, depends on a viable model of the similarity judgment itself. Many such models exist. Most models of similarity have been defined in terms of distance or lack of overlap of dimensional values or properties (see Shepard, 1980, for a review). The contrast model of similarity (Tversky, 1977) considers not only the degree to which the properties and values of objects are distinct, but also the degree of commonality amongst objects. Smith and Osherson (1989) applied a version of the contrast model to probability judgments for both conjunction and base rate problems.

One consequence of a feature weighting model like Tversky's (1977) is that a complex feature profile can be highly similar to a target category in some respects, but highly dissimilar in other respects. In accordance with this possibility, Yamagishi (2002) showed that people sometimes make non-complementary binary probability judgments i.e., for two mutually exclusive and exhaustive events A and ~A, their ratings for p(A) and p(~A) sum to over 1, in violation of the laws of probability.

Using enriched descriptions of 'Linda' and 'Bill' from Tversky and Kahneman's (1983) conjunction problems, Yamagishi replicated the finding that judgments of similarity were highly correlated with judged probability, and also showed that judgments of dissimilarity were highly correlated with judged negation probability (the probability that an individual was *not* a member of the target category). For example, an enriched description of Linda (now called 'Rhonda') added features such as 'is pro-life' and 'is very active in her church' that contrasted with the typical feminist features such as 'deeply concerned with issues of discrimination and social justice'. Subsequent judgments of the probability that Rhonda belonged to various categories (e.g. bank teller, feminist bank teller, league of women's voters) were highly correlated with judgments of similarity between Rhonda and these categories. A similar correlation obtained between judgments of the probability that Rhonda was not a member of these categories and judgments of how dissimilar she was from them.

The representativeness heuristic is, in essence, the proposal that judgments of probability are governed by the same mental processes that determine categorization via similarity to a prototype. X is judged likely to be an instantiation of event category Y to the degree that X is a good example of Y. Goodness-of-example is also a key determinant of judgments of the typicality of a category within a superordinate (see Hampton, 1998) and has been viewed as a measure of the similarity of an instance to a category prototype (Rosch, 1973).

However, other categorization models also exist that appear to take an outside view: They attribute typicality and category name judgments, not to similarity to a prototype, but to similarity to a set of exemplars (e.g., Nosofsky, 1984). Correspondingly, one can now find a model of probability judgment that also appeals to exemplar processing (Minerva-DM: Dougherty, Gettys, & Ogden, 1999). This model accounts for some of the phenomena attributed to the representativeness heuristic, and fits a variety of other probability judgment data. However, much of the work of this model is not done by the analysis of exemplars per se, but rather by the similarity relations that determine how exemplars are selected for processing. To this extent, the key to understanding certain judgment phenomena remains the acknowledgment that people tend to view events from the inside, in terms of their properties, and to make judgments by comparing the properties of an event to those of a model of the event being judged.

Associative theories of probability judgment

Associative models of probability judgment (e.g., Gluck & Bower, 1988; Shanks, 1991) are also prototype theories, and typically share the assumption that properties are independent. What distinguishes these models is their reliance on an incremental error-driven learning mechanism. This restricts their applicability to judgment situations where people are exposed to sequentially learned information. The central idea is that during this exposure people learn to associate cues (properties or features) with outcomes (typically another property or a category prediction), and that these learned associations form the basis for subsequent probability judgments. The associative links are supposed to be updated on a trial-by-trial basis according to the Rescorla-Wagner rule (1972).

Within this associative framework several studies have demonstrated biases in probability judgments that are analogues of biases typically found in the heuristics and biases program. Gluck and Bower (1988) demonstrated an analogue of base rate neglect (see also Medin & Edelson, 1988; Shanks, 1991). In an on-line learning task people learn to diagnose two fictitious diseases on the basis of sets of symptoms, and then rate the probability of these diseases given a particular target symptom. The learning environment is arranged so that the conditional probability of each disease is equal, but the overall probability (base rate) of one is high and the other low. Given this structure, the target symptom is a better predictor of the rare disease than the common one, and in line with the associative model people give higher ratings for the conditional probability of the rare disease. The associative account of this probability bias relies on cue competition effects, and cannot be explained by exemplar models such as Minerva-DM (Lewandowsky, 1995; Cobos, Almaraz & Garcia-Madruga, 2003).

Lagnado and Shanks (2002) extended this approach to disjunction errors. People learned to diagnose diseases at two levels of a hierarchy, and were then asked to rate the conditional probabilities of subordinate (e.g., Asian flu) and superordinate categories (e.g., Flu). The learning environment was arranged so that a target symptom was a better predictor of a subordinate disease than it was of that disease's superordinate category. In line with the associative account, people rated the conditional probability of the subordinate higher, even though this violated the extension rule of probability whereby the conditional probability of a subordinate cannot be higher than its superordinate. This suggests that people ignored the subset relation between the diseases, and based their conditional probability judgments on the degree of association between symptom and disease categories. No extensional (or exemplar-based) model could easily account for this finding.

As well as replicating the base-rate effect, Cobos et al. (2003) demonstrated a conjunction effect, where people rated the probability of a conjunction of symptoms higher than one of its conjuncts, and a conversion effect, they confused the conditional probability of symptom given disease with that of disease given symptom (analogous to the inversion confusion, see Villejoubert & Mandel, 2002). Both of these biases can be accommodated by an associative model but not by Minerva-DM's exemplar view.

These learning studies, in common with many one-shot judgment problems, suggest that people base their probability judgments on the degree to which a cue or property is associated with an outcome category, and neglect extensional information provided by the base rates or the set structure of the problem space.

Causality: probability judgment from relational explanation

That people often judge probability by considering the properties of a prototypical event, as opposed to the distributional properties of a category, seems indisputable in light of the evidence. That those properties are treated as independent is far from evident however. Certain examples of the conjunction fallacy already suggest that probability judgment cannot, in general, be reduced to overlap amongst independent properties. Consider the following problem from Tversky and Kahneman (1983):

> Mr. F. was randomly selected from a representative sample of adult males.
>
> Which is more probable?
>
> a) Mr. F. has had one or more heart attacks.
>
> b) Mr. F. has had one or more heart attacks and he is over 55 years old.

58% of their participants chose b) even though the conjunction rule of probability prescribes a). One account of this is that a Mr. F who is over 55 and has had one or more heart attacks is more similar to people's expectations of adult males than a Mr. F who has simply had one or more heart attacks. But an account that is at least as compelling is that the aged Mr. F is more representative because being over 55 is causally relevant to having had a heart attack. Representativeness may draw on the relation between the properties, a relation that is in essence explanatory (age is part of the explanation for heart problems).

Explanatory relations serve as the foundation for various kinds of judgment. To illustrate, Pennington and Hastie (1993) have shown that mock jurors are more likely to attribute guilt to a hypothetical accused if the evidence is presented in chronological rather than random order. They interpret the effect in terms of explanatory coherence: Evidence shown in chronological order facilitates the construction of a story that provides motivation and, more generally, allows the construction of a causal model of events. The key is that jurors are only willing to consider the probability of guilt sufficiently high if the judgment is supported by a causal model. Strong evidence per se does not automatically lead people to conclude guilt; the evidence must sustain a causal model.

Causal explanation in inductive inference

One domain of probability judgment concerns conditional probability: How do people update their beliefs about categories given information about other, related categories? Psychologists have approached this question using arguments with a particular form, such as:

Moose use norepinephrine as a neurotransmitter.

Therefore, deer use norepinephrine as a neurotransmitter.

in which a predicate ("uses norepinephrine as a neurotransmitter") is attributed to one or more premise categories and participants are asked to make a judgment about their belief that the predicate is also true of the conclusion category.

A number of phenomena have been demonstrated with such arguments (for a review, see Sloman & Lagnado, in press-a). One clear fact concerns arguments like the above that use predicates that participants know very little about and cannot use to reason with. Such arguments are judged strong to the extent that the premise and conclusion categories are similar (Rips, 1975). In fact, consider

Every individual bird has sesamoid bones.

Therefore, every individual robin has sesamoid bones.

and

Every individual bird has sesamoid bones.

Therefore, every individual ostrich has sesamoid bones.

Not only are both arguments often assigned a probability of less than 1, even by those people who agree that all robins and ostriches are birds, but people on average assign higher conditional probability to the first, more typical, conclusion than the second (Sloman, 1998). Here the adoption of an inside, similarity-based perspective leads people to neglect a relevant structural constraint (that a property possessed by every member of the superordinate set must be possessed by every member of any of its subsets).

Similarity relations are flexible however, and they change when predicates can be used to reason with (Heit & Rubinstein, 1994). The conditional probability that hawks have an anatomical property (like "has a liver with two chambers") is greater when told that chickens have the property than when told that tigers do. But given a behavioral predicate that concerns hunting (like "prefers to feed at night"), the probability that hawks have it is higher when conditioned on tigers rather than chickens. The interpretation of a predicate picks out a set of relevant properties of the categories under consideration.

The process by which causal knowledge selects relevant features is essentially a process of explanation. Induction is mediated by people's efforts to explain why a predicate would obtain of a category on the basis of relations amongst category features (Sloman, 1994) or by constructing a causal model to explain how a conclusion could be the causal effect of the premise (Medin, Coley, Storms, & Hayes, 2003).

For example, Medin et al. demonstrate that people find the argument

Bananas contain retinum, therefore monkeys do.

more convincing than the argument

Mice contain retinum, therefore monkeys do.

even though mice and monkeys are more similar than bananas and monkeys. Because monkeys are known to eat bananas but not mice, the causal path of ingestion mediates the first argument but not the second.

Here people are basing their probability judgments on the plausibility of causal explanations connecting premise and conclusion, an essentially inside operation, rather than invoking appropriate sets or reference classes.

Mental simulation

Judgments of conditional probability are closely related to probability judgments of conditional if-then statements (Edgington, 1995; Ramsey, 1931). The probability that "If the US had not invaded Iraq, then Iraq would have invaded some other country" cannot be easily distinguished from the probability that "Iraq would have invaded some other country given that Iraq had not been invaded by the US" (Over & Evans, 2003). The two forms seem almost interchangeable, suggesting that analyses of the interpretation of conditional statements of the type exemplified (*epistemic* conditionals) can be applied pretty much in their entirety to the analysis of conditional probabilities. The outline of a method to determine such probabilities was mentioned by Ramsey (1931) and theories of the meaning of conditional statements based on Ramsey's test have been developed by Stalnaker (1968) and Lewis (1976). A natural interpretation of Ramsey's proposal as a psychological hypothesis -- as a strategy for judgment -- is that to judge the probability of q given p, one first imagines a world in which p is true and then examines that world to see what the probability is that q holds in it.

This is basically the idea of the mental simulation heuristic, suggested by Kahneman and Tversky (1982a) as a common means by which people can make probability assessments. Thus people construct an appropriate causal model of the situation and then ‘run’ it using certain parameter settings (e.g., those specified in the antecedent of the conditional). The ease of achieving a target outcome is then taken as a measure of the probability of that outcome, given the initial conditions. Estimates for the probability of an event reached by such a procedure require an inside view; they involve focus on individual scenarios or stories, not the distributional properties of a set of cases.

The simulation heuristic is particularly applicable to situations where people make plans or predictions about the future. A robust empirical finding, termed the *planning fallacy*, is that people tend to underestimate the amount of time that it will take to complete a task or project (for a recent review see Buehler, Griffin & Ross, 2002). An example is the tendency of students to underestimate how long it will take them to finish an academic assignment. Buehler, Griffin and Ross (1994) found that students nearing the end of a one-year honors thesis underestimated their completion time by an average of 22 days.

The planning fallacy likely arises because people estimate time using mental simulations of the project or task, generating a plausible set of steps from initiation to completion. This focus on plausible scenarios can override the consideration of outside factors, such as the past frequencies of delayed completion. The effect obtained even when people made frequency rather than probability judgments, and estimated how many out of 10 tasks would be completed by a relevant date (Griffin & Buehler, 1999). Thus, in spite of a prompt to take an outside perspective, and judge completion in terms of the distribution across a set of similar cases, people persisted in taking an inside view.

Single-path or restricted path reasoning

More generally, there is a wealth of evidence suggesting that people's probability judgments are modulated by the number of alternative scenarios they construct. The judged probability of a target outcome increases when people imagine multiple causes of that outcome, but decreases when they imagine multiple causes of an alternative outcome (Koehler, 1994; Koriat, Lichtehstein & Fischhoff, 1980). Further, in reaching their judgments people tend to restrict themselves to considering a minimal number of possible scenarios (Mynatt, Doherty, & Dragan, 1993; Dougherty, Gettys & Thomas, 1997), often just one.

To illustrate, Dougherty et al. gave people text-based vignettes about a real-life situation (e.g., the circumstances leading up to the death of a firefighter) and then asked them for both a probability judgment and a list of the thoughts they had had in reaching this judgment. They found that participants used a mixture of single-path and several-path strategies, and that the former led to higher probability estimates than the latter. Moreover, although people initially generated several causal scenarios, they tended to eliminate the less plausible of these prior to making their probability judgment. The single-path strategy clearly requires an inside view, and it appears that even when a few alternative paths are entertained, people still adopt an inside strategy to reach their final judgment, but modulate this with knowledge of alternative paths.

Similar simplifying strategies are also apparent when people make probability judgments or predictions based on uncertain premises. Early research on cascaded inference (Steiger & Gettys, 1972) looked at multistage inferences, where the first step involves a probabilistic inference based on a known premise, and the second involves a further probabilistic inference conditional on this uncertain judgment. For example, suppose you want to bet on a horse in the Grand National, and you know that rain will favor 'Water Boy'. You see dark clouds gathering by the race track (this is your known premise). From this you estimate the probability of rain (this is your first step inference). Finally you estimate the probability that 'Water Boy' wins given this inference (this is the second step). In such cascaded inference problems people adopt a 'best guess' or 'as if' strategy: they make their second inference *as if* the most probable outcome at the first step is true rather than probable. In our example this involves assuming that it will rain, and basing your probability estimate that 'Water Boy' wins on this assumption. A normatively more justifiable procedure is to compute a weighted sum over all the alternative pathways using a modified version of Bayes theorem (Jeffrey, 1965). Even with just a few alternatives this would lead to a complex computation; it is not surprising that people use a simplifying heuristic. In effect this heuristic amounts to abandoning an outside perspective across multiple possible chains of events, and focusing on the single most probable, and therefore presumably the most representative, path.

An independent but very similar line of research has been developed in category inference (Murphy & Ross, 1994). The starting point was Anderson's rational model of categorization (1991) and in particular his claim that when making a prediction on the basis of an uncertain categorization people follow a Bayesian rule that computes a weighted average over all potential categories. In contrast to this *multiple category* view, Murphy and Ross provided evidence for a *single category* view where just the most probable category is used to make a prediction. For example, consider the task of predicting whether the insect flying towards you on a dark night is likely to sting you. Let the potential categories in this situation be *Fly*, *Wasp* or *Bee*. According to the multiple category view you compute a weighted average across all three categories in order to determine the probability of being stung. In contrast, on the single category view you base your prediction only on the most probable category (e.g., just one of *Fly*, *Wasp* or *Bee*) and ignore alternative categories.

Ross and Murphy (1996) gave a heuristic explanation of these findings in terms of the availability of categories in memory, and showed that people can incorporate multiple categories when these are made more accessible. However, representativeness also appears to be involved; that is, people focus on the most representative category to guide subsequent inference. Lagnado and Shanks (2003) extended this line of research to on-line learning situations involving hierarchical categories. They too found strong evidence that people focused on the most probable category in order to make a subsequent probability estimate, and showed that these estimates were readily manipulated by priming people to different hierarchical levels (e.g., subordinate or superordinate). These results were explained in terms of the associative links that people build up during learning, and their subsequent activation in the judgment phase. For example, if you think of an individual as a likely Broadsheet reader then you activate the associated property of voting Conservative, whereas if you think of the same individual as a Guardian reader (a subset of Broadsheet readers) you activate the associated property of voting Labour.

These three sets of studies converge on the same conclusion despite quite different paradigms. And they all show a strong influence of the inside perspective; one that avoids consideration of multiple alternative paths, and focuses on a single most probable case.

Finally, various studies by Windschitl et al. (e.g., Windschitl & Young, 2001) have shown that even when people are presented with a full set of alternative outcomes, so that the correct probability for a target outcome is transparent, their intuitive assessments of uncertainty (measured by non-numeric verbal means) are modulated by irrelevant features of the distribution of alternatives. For example, people express greater optimism about winning a 59 ticket raffle in which they hold 17 tickets and the other competitors hold 9, 9, 8, 8, and 8 respectively than when they hold 17 tickets and the other competitors hold 16, 7, 7, 6 and 6. This *alternative outcomes* effect can be explained by a heuristic comparison process that focuses on the target outcome and its strongest competitor. Furthermore, in a learning paradigm where people must base their judgments on memory for the frequencies of past outcomes, the effect is also reflected in their numerical probability estimates (Windschitl, Young & Jenson, 2002).

These studies imply a dissociation between an outside view probability judgment (given by computing the proportion of favorable outcomes) and a more intuitive probability judgment reflected in verbal reports. They also suggest that the latter is based on a comparison process characteristic of the inside perspective. Moreover, the findings further support the idea that people tend to restrict the set of alternatives they consider in order to reach a probability judgment. This can manifest itself in the focus on the most probable inferential path when alternative paths are possible or the focus on just the strongest outcomes when the full distribution of outcomes is transparent.

Outside models

Mental model theory of extensional reasoning

An explicitly outside perspective on probabilistic reasoning is taken by Johnson-Laird, Legrenzi, Girotto, Legrenzi and Carveni (1999) in their mental model theory of extensional reasoning. They argue that naïve reasoners – those who are unfamiliar with the probability calculus – can nevertheless infer the probabilities of events in an extensional fashion. Here extensional reasoning consists in 'inferring the probability of an event from the different possible ways in which it could occur'. The central claim is that people reach probability judgments through the construction of mental models of the problem situation, and subsequent computations over these models.

One key claim of the model theory is that because people construct mental models that focus on what is mentioned in the probability problem, their mental partition of the problem space sometimes fails to correspond to the full set of mutually exclusive and exhaustive possibilities. This can explain a variety of erroneous probabilistic inferences (Johnson-Laird et al., 1999). Another key claim is that in the absence of information to the contrary people assume that each mental model is equiprobable. For example, suppose that you have lost sight of your companions on a country walk. You come to a junction where the road splits into three separate paths. With no extra information, you assign each path a probability of 1/3. In addition, once an equiprobable partition is made, people can compute the probabilities of compound events by proportionality. If two of the paths lead up the hill, and one leads down the hill, the probability that your companions went up the hill is 2/3.

Various studies (e.g., Girotto & Gonzalez, 2003) show that naive reasoners, including children, can compute probabilities based on these principles. Moreover, the typical errors made by participants can be attributed to inappropriate partitions of the problem space. This is also held to underlie common mistakes in probability puzzles such as the Monty Hall problem and Bertrand's three-card problem (Johnson-Laird et al., 1999).

The principle of equiprobability does a lot of work for the model theory, but is inadequate as a means for establishing probabilities *a priori* (Gillies, 2001). Problem spaces often admit of different partitions, and in such cases the equiprobability principle can lead to the same event being assigned inconsistent probabilities. To illustrate, if you divide the earlier problem space into equiprobable paths, the probability that your friends went up the hill is 2/3. However, if you divide the space of possibilities into either 'going up the hill' or 'going down the hill', then the probability of this same event is ½.

Although this is fine if the principle is intended as a psychological heuristic (indeed it helps to explain people's reasoning errors once a particular partitioning is assumed), such examples undermine its status as a logical principle. This severely restricts the model theorists' claims that extensional probabilistic reasoning is deductive. If people are presented with a unique equiprobable partition, then they can deduce other probabilities. But creating an appropriate partition is the main problem in many situations, and this does not reduce to deduction.

A related problem for the model theory, and indeed any extensional theory, concerns the differential weighting of mental models. In many situations we do have some prior weighting of the possibilities. Thus we may see that a coin has an irregular shape, and infer that one side is favoured over the other. The model theory proposes that we encode these weightings by means of numerical tags attached to each model. But the theory does not tell us how to infer such weights when the principle of equiprobability does not apply. Here it appears that we must rely on inside judgments – e.g., about the causal properties of the coin, about its similarity to a normal coin, etc.

In sum, people can sometimes reason from the outside, and possibly do so by constructing mental models, but the weighting of these models is not a logical matter, and will often depend on inside judgments.

Availability: The construction of a set of instances

Tversky and Kahneman (1973) introduced availability as a heuristic method for estimating frequencies or probabilities. People use the availability heuristic whenever they base their estimates on the ease with which instances or occurrences come to mind.

Despite the simplicity of its formulation, the heuristic covers a range of cases. For one, it applies both to the recall of previous occurrences (e.g., how often you remember team X beating team Y) and to the generation of possible occurrences (e.g., how many ways you can imagine a novel plan going wrong). Second, it need not involve *actual* recall or generation, but only an assessment of the ease with which these operations *could* be performed. Our discussion will focus on cases of actual recall or generation because these are most relevant to current models of probability judgment. Issues concerning the subjective feeling of ease of recall rather than the content of recall lie beyond the scope of this chapter (see Schwarz & Vaughn, 2002).

As Tversky and Kahneman point out, availability is an ecologically valid cue to frequency because in general frequent events are easier to recall than infrequent ones. However, the clearest evidence that people use this heuristic comes from studies where it leads to biased estimates. For example, under timed conditions people generate far more words of the form _ _ _ _ *ing* than of the form _ _ _ _ _*n*_ , even though the first class is a subset of the second. This shows that the first form is more available in memory than the second. Further, when one group estimates how many words in a four page novel have the form _ _ _ _ *ing*, and another answers the same question for the form _ _ _ _ _ *n*_ , estimates are much higher for the first. This suggests that in making their frequency estimates people relied on the ease with which they could retrieve instances (Tversky & Kahneman, 1983). Goldstein and Gigerenzer (2002) have suggested a related heuristic for judgment based on recognition rather than recall.

The availability heuristic furnishes one method for constructing a sample of events or instances. A more general account of sampling (and possible biases) is advanced by

Fiedler (2000). This extends the analysis from memory-based search to environmental search. Both kinds of search require an outside perspective, and both can lead to biases in the resulting set of instances. For one, the environment might be sampled in a biased way. Fiedler cites an example concerning the assessment of lie detectors. Many validity studies of such devices incorporate a pernicious sampling bias: of all the people who fail the test, validity assessments only include those who subsequently confess. Those who fail the test but are telling the truth are not counted (cf. positive test strategies, Klayman & Ha, 1987). Another common route to error is when people sample from a biased environment such as the media that over-represent sensational and newsworthy events (Fischhoff, 2002; Slovic, Fischhoff & Lichenstein, 1980).

Systematic biases can also arise when one generates a sample from one's own memory. This can occur due to the intrusion of associative memory processes (Kelley & Jacoby, 2000). Alternatively, it can result from the biased generation of possibilities or scenarios. For example, people tend to recruit reasons to support their own views, and neglect counter-arguments or reasons that support opposing conclusions (Koriat, Lichtenstein & Fischhoff, 1980; Kunda, 1990). In all these cases people take a step towards an outside perspective by gathering a set of instances. Subsequent judgments may be flawed because the sample is atypical of the wider population. Fielder (2000) argues that many judgmental biases arise because -- rather than in spite -- of our ability to process sample information accurately. Samples are often biased, and we lack the meta-cognitive abilities to correct for such biases.

The availability heuristic involves the generation of a set of instances, but it does not specify how people go from this set to a probability judgment. In certain cases this will be relatively transparent, such as when more instances of horse A winning a race rather than horse B are recalled and thus A is predicted to beat B. However, many situations will be more complicated. Suppose A and B have never raced against each other, and A has only raced in easy races, B in hard ones. In this case you may need to weight their number of wins differentially, and for this availability offers little guide.

The main role of availability in probability judgment therefore is to facilitate the selection or construction of a set of instances. Where this process leads to just one instance, the reasoner is likely to rely on an inside judgment (of similarity or association) to reach a probability estimate. Where it leads to a set of instances, the reasoner has a choice. They can resort to an inside strategy by simply averaging across properties of these instances, as they appear to do in frequency versions of the planning fallacy (Buehler et al., 2002; Griffin & Buehler, 1999). Or they can use the distributional properties of the set (e.g., the proportion of Xs that are Ys) to reach a probability judgment from the outside.

Statistical heuristics

The untutored approach to probabilistic reasoning often involves an inside view. What are the factors that promote a shift from an inside to an outside perspective? An influential developmental claim is that people's statistical intuitions stem from their learning about the nature of randomizing devices (e.g., coin tosses, card draws, spinners). Piaget and Inhelder (1951) argued that the child's concept of uncertainty derives from an understanding of physical causality, from causal mechanisms and the predictability of the outcomes they generate. Initially children struggle to comprehend randomizing devices, and invoke ad hoc causal explanations to account for the lack of predictability. Gradually they develop a better understanding; they appreciate the unpredictability of individual outcomes and the relevance of repeated trials.

Drawing on these ideas, Nisbett, Krantz, Jepson and Kunda (1983) argued that people possess *statistical heuristics* – 'intuitive, rule-of-thumb inferential procedures that resemble formal statistical procedures'. Examples of such heuristics include the preference for more rather than less evidence, intuitions about variability, and an appreciation of base rates in certain contexts. People are supposed to apply these heuristics to randomizing devices at a relatively early age, and later on to probabilistic causal systems (e.g., sports events, test performances, weather forecasts). Given that we do possess such heuristics, why do we often persist in reasoning from the inside? Nisbett et al. identified three relevant factors.

Firstly, the use of statistical heuristics depends on the clarity of the sample space and the sampling process. Thus typical chance set-ups promote the use of these heuristics because both the space of alternative outcomes (e.g., the faces of a coin or die) and the repeatability of trials are clear and well-defined. Indeed most games of chance are specifically designed to make these features transparent. For example, the symmetry of the different faces of a die, and the similarity of each separate die throw, make it much easier to take a distributional or outside perspective. This is supported by Girotto and Gonzalez's (2003) findings that in a variety of problems with well-defined sample spaces adults and children reason extensionally. It is also supported by Nisbett et al.'s demonstration that under appropriate conditions people obey the law of large numbers (see also Sedlmier & Gigerenzer, 2000; Fiedler, 2000) and appreciate that homogenous properties require less confirmatory evidence than heterogeneous ones.

A related factor is the transparency of the randomizing element itself. With throws of the dice or spins of the roulette wheel the haphazard nature of the process is readily apparent. Likewise in ball games the movement of the ball reveals elements of randomness. This contrasts with many social domains where the randomizing factor is less tangible. For example, because of the lack of cues to the random elements in job interviews, people are prone to regard the interview as a 'portrait in miniature' rather than a sample from a population.

In support of this claim, Nisbett et al. showed that participants shift from inside to outside reasoning according to the salience of the random process, and Gigerenzer, Hell and Blank (1988) have shown that base rate neglect in the lawyers/engineers problem is reduced if people perform the random sampling themselves. Base rate neglect in the medical diagnosis problem is also reduced when the random sampling in the problem is made explicit (Cosmides & Tooby, 1996).

A third factor cited by Nisbett et al. involves cultural prescriptions to think statistically. They report various studies showing how statistical training promotes the adoption of an outside perspective, and this has been echoed in recent work on training people to reason statistically (Sedlmeier, 1999). Nisbett et al. draw the moral that the driving force in the shift from inside to outside reasoning is cultural evolution. This contrasts with theorists who argue that certain statistical reasoning abilities are hard-wired (Cosmides & Tooby, 1995; Gigerenzer, 2000), and require appropriate representations. A third view is that the ability to reason extensionally develops from other more basic cognitive operations (cf. Heyes, 2003).

Finally, the argument that our conception of random mechanisms develops from our prior understanding of causal systems cuts two ways. As well as suggesting conditions that facilitate an outside view, it suggests reasons that can undermine or distort it; vestiges of our primitive notion of a causal system remain in our ordinary notions of a random process. Both the gambler's fallacy and the hot hand fallacy are testament to this. In the case of the gambler's fallacy people often justify their fallacious inferences (e.g., that after a run of coin tosses resulting in heads, the next toss is more likely to fall tails) by claiming that things must eventually balance out. This is to conceive of chance as a self-correcting process rather than a genuinely random one (cf. Tversky & Kahneman, 1982). Similarly, when basketball fans believe that after a run of successful baskets a player is more likely to succeed with their next shot (Gilovich, Vallone & Tversky, 1985), they defend these erroneous beliefs by appeal to causal explanations (e.g., a 'hot' player is more motivated or confident). In both cases people impose causal structure upon random sequences of events.

Nested sets

One of the virtues of the inside view is that it is able to take advantage of a lot of knowledge very quickly, all the knowledge normally associated with properties of the category being judged. However, its focus on one type of information means that it neglects a different type. Its focus on the internal structure of a category causes it to neglect the distributional structure of category instances and this can lead to certain systematic errors. In particular, an inside view fails to represent class inclusion relations. This is a latent cause of the conjunction fallacy; people make their judgments on the basis of similarity, failing to realize that the conjunction is a subset of a set defined by each constituent (e.g., 55-year-olds who have had heart attacks are a subset of all those who have had heart attacks). Analogously, an inside view fails to represent that subordinate categories are a subset of a superordinate category in inductive inference, explaining why people neglect such relations in the inclusion fallacy (Shafir, Smith, & Osherson, 1990) or inclusion-similarity phenomenon (Sloman, 1998).

An outside view, in contrast, makes class inclusion relations transparent. By virtue of representing a distribution of instances, simple relations amongst sets of instances are represented automatically (Tversky & Kahneman, 1983). Sloman and Over (2003) describe the nested-sets hypothesis as the view that: i. All else being equal, people prefer an inside view; but ii. representing instances requires an outside view that can make set inclusion relations transparent. Closely related hypotheses can be found in Evans, Handley, Perham, Over, and Thompson (2000), and Johnson-Laird et al. (1999).

Evidence in favor of the nested-sets hypothesis includes demonstrations that presenting conjunction problems in the context of diagrams that reveal the nested-set relations among constituent variables reduces the incidence of the conjunction fallacy (Agnoli & Krantz, 1989; Sloman, Over & Slovak, 2003). Diagrams revealing set structure also reduce the incidence of base-rate neglect (Cosmides & Tooby, 1996; Sloman et al., 2003) and reduce error on other probability problems (Yamagishi, 2003). Also, when asked to depict their thought process on paper, people who get problems correct are more likely to draw pictures that represent nested-set relations than people who don't (Agnoli & Krantz, 1989; Gigerenzer & Hoffrage, 1995; Sloman et al., 2003).

The nested-sets hypothesis is also supported by evidence that presenting problems in terms of the frequency of instances facilitates performance over presentation in terms of single-event probabilities (Cosmides & Tooby, 1996; Fiedler, 1988; Tversky & Kahneman, 1983). The frequency frame induces an outside perspective by asking participants to think about multiple instances of the category instead of assuming their more usual inside property-based perspective.

The frequency effect on probability judgment has sparked a lively debate over the status of the fallacies of probability judgment, sometimes referred to as "cognitive illusions." Theorists like Gigerenzer (2000) and Cosmides and Tooby (1996) have argued that probability judgment must be understood in its ecological context. Errors arise on this view when participants are asked to make judgments of single-event probabilities because people did not evolve to make such judgments. Rather, they evolved to make judgments using natural frequencies. In response, Kahneman and Tversky (1996) pointed out that errors have been demonstrated using judgments of frequency since the onset of the program demonstrating cognitive illusions, and that detractors have failed to provide an account of the systematic patterns of behavior that the program has uncovered. Moreover, frequency formulations (even via natural sampling) are neither a necessary nor a sufficient condition for facilitation. Facilitation has been observed with judgments of probability (Ajzen, 1977; Girotto & Gonzalez, 2001; Johnson-Laird et al., 1999; Mellers & McGraw, 1999; Sloman et al., 2003) and errors have been observed with judgments of frequency (Bar-Hillel & Neter, 1993; Gluck &

Bower, 1988a; Lagnado & Shanks, 2002, 2003). The evolutionary rationale is itself suspect as people may well have evolved to use devices like causal models to make single-event probability judgments (Sloman & Over, 2003).

The bulk of the data do indeed show that presenting a problem in terms of frequency can reduce the incidence of error in probabilistic judgment. Gigerenzer and Hoffrage (1995) are surely correct that this occurs because frequency formats can reduce the computational complexity of a problem. The current evidence suggests, however, that this is not the effect of natural frequency via natural sampling per se, but is mediated by the elicitation of an outside perspective that makes nested set relations transparent. In short, people can reason extensionally whether or not they have to process "natural frequencies" (cf. Girotto & Gonzalez, 2001).

Support theory

We have already indicated the tight linkage between probability judgment and categorization: The representativeness heuristic can be viewed as the claim that probability judgments are species of categorization decisions and the study of inductive inference concerns how probability judgments about stable category properties are updated with new evidence. An analysis of how probability judgments are driven by the way categories are described has been offered in the form of Support Theory (Tversky & Koehler, 1994; Rottenstreich & Tversky, 1997).

The evidence in favor of Support Theory is mixed. Some of its implications, such as explicit subadditivity, are well supported (Rottenstreich & Tversky, 1997), but others, such as implicit subadditivity, are not (Sloman, Rottenstreich, Wisniewski, Hadjichristidis, & Fox, 2003). Support Theory is a non-extensional model of probability judgment in the sense that it is defined over descriptions of events and not events themselves. However, the model cannot be described as taking an outside or an inside view as it includes neither a representation of categorical information nor any claims about the process of probability judgment. It is merely a representation of probability judgment. Tversky and Koehler's (1994) claim that the representativeness heuristic and reasons for confidence affect support suggests an inside view. But they also claim that support can be sensitive to relative frequency data, and this would require an outside view. So the model is at a level of abstraction above the inside/outside distinction.

Conclusions

So what are the guidelines that demarcate an outside from an inside probability judgment? To start with, an outside view requires the enumeration of a relevant set of instances or possibilities. But this alone is insufficient. Probability judgments must be based on some distributional property of this set, such as a proportion or relative frequency. Moreover, reasoning from an outside perspective often requires attending to structural relations amongst instances (such as nested set relations) rather than just causal or similarity relations. Finally, although both inside and outside perspectives can lead to fallacious reasoning, the nature of these errors typically differs. Inside judgments tend to violate the laws of probability because they are driven by natural assessments (such as similarity or association) unconstrained by probabilistic coherence. By contrast, outside-view errors tend to arise from an incomplete or inappropriate specification of the problem space.

Our survey suggests an essential tension between the two perspectives. To reach an unbiased probability judgment from the outside requires both an appropriate representation of the problem and a distributional perspective across the relevant units. But the task of achieving a good representation often relies on inside judgments – grouping units by similarity, determining relevance by causality, weighting alternatives by associative strength. And each of these factors can inhibit our ability to see the wider field. The outside view requires that we transcend the specific case, and yet it is the specific case, and all its properties, that make our representations of the world so compelling.

References

Agnoli, F. & Krantz, D. H. (1989). Suppressing natural heuristics by formal instruction: The case of the conjunction fallacy. Cognitive Psychology, 21, 515-550.

Ajzen, I. (1977). Intuitive theories of events and the effects of base-rate information on prediction. Journal of Personality and Social Psychology, 35, 303-314.

Anderson, J. R. (1991). The adaptive nature of human categorization. Psychological Review, 98, 409 - 429.

Bar-Hillel, M. & Neter, E. (1993). How alike is it versus how likely is it: A disjunction fallacy in probability judgments. Journal of Personality and Social Psychology, 65, 1119-1131.

Buehler, R., Griffin, D., & Ross, M. (2002). Inside the planning fallacy: The causes and consequences of optimistic time predictions. In T. D. Gilovich, D. W. Griffin, & D. Kahneman (Eds.), Heuristics and biases: The psychology of intuitive judgment (pp. 250-270). New York: Cambridge University Press.

Buehler, R., Griffin, D. & Ross, M. (1994). Exploring the "planning fallacy": Why people underestimate their task completion times. Journal of Personality and Social Psychology. 67, 366-381.

Carnap, R. (1950). Logical foundations of probability. Chicago: University of Chicago Press.

Cobos, P. L., Almaraz, J., & Garcia-Madruga, J.A. (2003). An associative framework for probability judgment: An application to biases. Journal of Experimental Psychology: Learning, Memory, and Cognition, 29, 80-94

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. Cognition, 58, 1-73.

Dawes, R. M. (1996). House of cards: Psychology and psychotherapy built on myth. New York: The Free Press.

Dougherty, M. R. P., Gettys, C. F., & Ogden, E. E. (1999). MINERVA-DM: A memory processes model for judgments of likelihood. Psychological Review, 106, 180 - 209.

Dougherty, M. R. P., Gettys, C. F., & Thomas, R. P. (1997). The role of mental simulation in judgments of likelihood. Organizational Behavior and Human Decision Processes, 70, 135 - 148.

Evans, J. St. B. T., Handley, S. H., Perham, N., Over, D. E., & Thompson, V. A. (2000). Frequency versus probability formats in statistical word problems. Cognition, 77, 197-213.

Fiedler, K. (2000). Beware of samples! A cognitive-ecological sampling approach to judgment biases. Psychological Review, 107, 659-676

Fiedler, K. (1988). The dependence of the conjunction fallacy on subtle linguistic factors. Psychological Research, 50, 123-129.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1978). Fault trees: Sensitivity of estimated failure probabilities to problem representation. Journal of Experimental Psychology: Human Perception and Performance, 4, 330-334.

Gigerenzer, G. (2000). Adaptive Thinking: Rationality in the real world.: Oxford University Press.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. Psychological Review, 102, 684-704.

Gigerenzer, G., Hell, W., & Blank, H. (1988). Presentation and content: The use of base rates as a continuous variable. Journal of Experimental Psychology: Human Perception and Performance. 14, 513-525

Gillies, D. (2001). Philosophical theories of probability. London: Routledge.

Gilovich, T., Vallone, R. and Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. Cognitive Psychology, 17, 295-314.

Girotto, V., & Gonzalez, M. (2001). Solving probabilistic and statistical problems: A matter of information structure and question form. Cognition, 78, 247-276.

Girotto, V., & Gonzalez, M. (2003). Probabilistic reasoning and combinatorial analysis. To appear in V. Girotto & P. Johnson-Laird (Eds.) The shape of reason: Essays in honour of Paolo Legrenzi. Hove, Psychology Press.

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. Psychological-Review, 109, 75-90

Griffin, D. & Buehler, R. (1999). Frequency, probability, and prediction: Easy solutions to cognitive illusions. Cognitive-Psychology. 1999, 38(1): 48-78

Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. Journal of Experimental Psychology: General, 117, 227-247.

Hacking, I. (2001). An introduction to probability and logic. Cambridge: Cambridge University Press.

Hacking, I. (1975). The emergence of probability. Cambridge: Cambridge University Press.

Hampton, J.A. (1998). Similarity-based categorization and fuzziness of natural categories, Cognition, 65, 137-165.

Heit, E., & Rubinstein, J. (1994). Similarity and property effects in inductive reasoning. Journal of Experimental Psychology: Learning, Memory, and Cognition, 20, 411-422.

Heyes, C. M. (2003). Four routes of cognitive evolution. Psychological Review.

Jeffrey, R. (1965) The logic of decision. New York: McGraw-Hill

Johnson-Laird, P.N., Legrenzi, P., Girotto, V., Legrenzi, M., and Caverni, J-P. (1999). Naive probability: a mental model theory of extensional reasoning. Psychological Review, 106, 62-88.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. D. Gilovich, D. W. Griffin, & D. Kahneman (Eds.), Heuristics and biases: The psychology of intuitive judgment (pp. 49-81). New York: Cambridge University Press.

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. Psychological Review, 103, 3, 582-591.

Kahneman, D. & Tversky, A., (1982). Variants of uncertainty. In D. Kahneman, P. Slovic & A. Tversky (Eds.), Judgment under uncertainty: Heuristics and biases. Cambridge: Cambridge University Press.

Kahneman, D. & Tversky, A., (1982a).The simulation heuristic. In D. Kahneman, P. Slovic & A. Tversky (Eds.), Judgment under uncertainty: Heuristics and biases. Cambridge: Cambridge University Press

Kelley, C.M. & Jacoby, L.L. (2000). Recollection and familiarity: Process-dissociation. In E. Tulving & F.I.M. Craik (Eds.), The Oxford handbook of memory. New York, NY: Oxford University Press.

Klayman, J., & Ha, Y-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. Psychological Review, 94, 211- 228.

Koehler, D. J. (1994). Hypothesis generation and confidence in judgment. Journal of Experimental Psychology: Learning, Memory, and Cognition, 20, 461-469.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. Journal of Experimental Psychology: Human Learning and Memory, 6, 107-118.

Kunda, Z. (1990). The case for motivated reasoning. Psychological bulletin, 108, 480-98.

Lagnado, D. A., & Shanks, D. R. (2002). Probability judgment in hierarchical learning: a conflict between predictiveness and coherence. Cognition, 83, 81-112.

Lagnado, D. A., & Shanks, D. R. (2003). The influence of hierarchy on probability judgment. Cognition, 89, 157- 178.

Lewandowsky, S. (1995). Base-rate neglect in ALCOVE: A critical reevaluation. Psychological Review,102, 185-191

Lewis, D. (1976). Probabilities of conditionals and conditional probabilities. Philosophical Review, 85. 297-315.

Medin, D.L., Coley, J.D., Storms, G. & Hayes, B. (in press). A relevance theory of induction. Psychonomic Bulletin and Review.

Medin, D. L., & Edelson, S. M. (1988). Problem structure and the use of base-rate information from experience. Journal of Experimental Psychology: General, 117, 68-85.

Meehl, P. E. (1954). Clinical vs. statistical prediction: a theoretical analysis and a review of the evidence. Minneapolis, MN, US: University of Minnesota Press.

Meehl, P. E. (1986). Causes and effects of my disturbing little book. Journal of Personality Assessment. 50, 370-375.

Murphy, G. L., & Ross, B. H. (1994). Predictions from uncertain categorizations. Cognitive Psychology, 27, 148-193.

Mynatt, C. R., Doherty, M. E., & Dragan, W. (1993). Information relevance, working memory, and the consideration of alternatives. The Quarterly Journal of Experimental Psychology, 46, 759–778.

Nisbett, R.E., Krantz, D.H., Jepson, D.H. & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. Psychological Review, 90, 339-363.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. Journal of Experimental Psychology: Learning, Memory, and Cognition, 10, 104-114.

Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A., & Shafir, E. (1990). Category-based induction. Psychological Review, 97, 185-200.

Over, D.E., & Evans, J.St.B.T. (2003). The probability of conditionals: The psychological evidence. Mind and Language.

Pennington, N. & Hastie, R. (1993). Reasoning in explanation-based decision making. Cognition, 49, 123-163.

Piaget, J., & Inhelder, B. (1951). The genesis of the idea of chance in the child. Paris: Presses Universitaires de France

Ramsey, F. P. (1931). The foundations of mathematics and other logical essays. London: Routledge and Kegan Paul.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), Classical conditioning II: Current theory and research (pp. 64-99). New York: Appleton-Century-Crofts.

Rips, L. (1975). Inductive judgments about natural categories. Journal of Verbal Learning and Verbal Behavior. 14, 665-681.

Rosch, E.H. (1973) Natural categories. Cognitive Psychology, 4, 328-350.

Rottenstreich, Y., & Tversky, A. (1997). Unpacking, repacking, and anchoring: Advances in support theory. Psychological Review, 104, 406-415.

Schwarz, N., & Vaughn, L. (2002). The availability heuristic revisited: ease of recall and content of recall as distinct sources of information. In T. D. Gilovich, D. W. Griffin, & D. Kahneman (Eds.), Heuristics and biases: The psychology of intuitive judgment (pp. 103-119). New York: Cambridge University Press.

Sedlmeier, P. (1999). Improving statistical reasoning: Theoretical models and practical implications. Mahwah, NJ, US: Lawrence Erlbaum Associates.

Sedlmeier, P. & Gigerenzer, G. (2000). Was Bernoulli wrong? On intuitions about sample size. Journal of Behavioral Decision Making, 13, 133-139

Shafir, E., Smith, E. E. & Osherson, D. (1990). Typicality and reasoning fallacies. Memory & Cognition, 18, 229-239.

Shanks, D. R. (1991). A connectionist account of base-rate biases in categorization. Connection Science, 3, 143-162.

Shepard, R. N. (1980). Multidimensional scaling, tree-fitting, and clustering. Science, 210, 390-398.

Sloman, S. A. (1994). When explanations compete: the role of explanatory coherence on judgments of likelihood. Cognition, 52, 1-21.

Sloman, S. A. (1998). Categorical inference is not a tree: The myth of inheritance hierarchies. Cognitive Psychology, 35, 1-33.

Sloman, S. A. & Lagnado, D. A. (In press a). The Problem of induction. To appear in R. Morrison and K. Holyoak (Eds.) Handbook of Thinking and Reasoning

Sloman, S. A. & Lagnado, D. A. (In press b). Causal invariance in reasoning and learning. To appear in B. Ross Ed. Psychology of learning and motivation.

Sloman, S. A., & Over, D. E. (2003). Probability judgment: from the inside and out. D.E. Over (Ed.) Evolution and the Psychology of Thinking: The Debate. Psychology Press.

Sloman, S. A., Over, D. E., & Slovak, L. (2003). Frequency Illusions and Other Fallacies. Organizational Behavior and Human Decision Processes, 91, 296-309.

Sloman, S. A., Rottenstreich, Y., Wisniewski, E., Hadjichristidis, C., & Fox, C. R. (2003). Unpacking implicit probability judgment. Manuscript submitted for publication.

Stalnaker, R. (1968). A theory of conditionals. American Philosophical Quarterly Monograph Series, 2, 98-112.

Steiger, J. H., & Gettys, C. F., (1972). Best-guess errors in multistage inference. Journal of Experimental Psychology, 92, 1-7.

Smith, E. E., & Osherson, D. N. (1989). Similarity and decision making. In S. Vosniadou & A. Ortony (Eds). Similarity and analogical reasoning. (pp. 60-75). New York, NY, US: Cambridge University Press.

Tversky, A. (1977). Features of similarity. Psychological Review, 84, 327-352

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. Psychological Review, 90, 293-315.

Tversky, A., & Kahneman, D. (1982). Evidential impact of base rates. In Kahneman, D., Slovic, P., & Tversky, A. (Eds.), Judgment under uncertainty: Heuristics and biases. Cambridge: Cambridge University Press.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. Cognitive Psychology, *5*, 207-232.

Tversky, A., & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. Psychological Review, 101, 547-567.

Windschitl, P. D., & Young, M. E. (2001). The influence of alternative outcomes on gut-level perceptions of certainty. Organizational Behavior and Human Decision Processes, 85, 109-134.

Windschitl, P. D., Young, M. E., & Jenson, M. E. (2002). Likelihood judgment based on previously observed outcomes: the alternative-outcomes effect in a paradigm. Memory and Cognition, 30, 469-477.

Villejoubert, G., & Mandel, D. R. (2002). The inverse fallacy: An account of deviations from Bayes's theorem and the additivity principle. Memory & Cognition, 30, 171-178).

Yamagishi, K. (2003). Facilitating Normative Judgments of Conditional Probability: Frequency or Nested Sets? Experimental Psychology, 50, 97-106.

Yamagishi, K. (2002). Proximity, Compatibility, and Noncomplementarity in Subjective Probability. Organizational Behavior and Human Decision Processes, 87, 136-155.