Reasoning and Decision Making

Mike Oaksford
School of Psychology
Birkbeck College, University of London

Nick Chater
Department of Cognitive, Perceptual and Brain Sciences &
Centre for Economic Learning and Social Evolution (ELSE)
UCL

&

Neil Stewart
Department of Psychology
University of Warwick

Acknowledgements: Nick Chater is supported by a Senior Research Fellowship from the Leverhulme Trust.

In this chapter, we introduce some recent developments in the areas of human reasoning and decision making. We focus on the how people use given information to make inferences concerning new information (i.e., reasoning) or to decide what to do (i.e., decision making). The fields of reasoning and decision making are both large, and we will be selective. In particular, in our discussion of reasoning, we shall focus on theories of how people reason with conditionals, i.e., theories of the nature of the linkage between given and inferred information. Regarding decision making, we focus on decision-under-risk, using problems which are explicitly described in linguistic or symbolic terms.

Reasoning

Perhaps the fundamental question in the Psychology of Reasoning is: do people reason correctly (Wason, 1960; Wason & Johnson-Laird, 1972)? To answer this question requires relating data on how people *do* reason to a normative theory of how they *should* reason. The normative theory typically adopted in the field is deductive logic. To be rational is, on this view, to be logical.

We focus in this chapter on experimental work on human *deductive* reasoning as opposed to *inductive* reasoning. There are various ways of marking this distinction. Perhaps the most fundamental is that in a deductive argument, if the premises are the true, the conclusion must be true. In an inductive argument, the premises merely make the conclusion plausible or probable. Thus, an argument from observing specific white swans to the conclusion *all swans are white* is inductive---here, because it is entirely possible that later counterexamples (e.g., black swans) exist. That is, the conclusion can be defeated by further information and so the argument is *defeasible*. Second, inductive reasoning relies on content (Goodman, 1953). For example, suppose a long-term Königsberg resident notes that all swans observed so far have lived in Königsberg. The observer is unlikely to confidently conclude that *all swans live in Königsberg*. The difference depends on content: colour, but not geographical location, are properties likely to be shared by all members of a species. By contrast, defeasibility and the effect of content do not affect deductive validity. However, they do affect how people reason deductively.

The standard logic of the conditional, *if…then*, has been assumed to provide the normative standard in much of the experimental work on human deductive reasoning. In standard logic, the meaning of logical terms (*if p then q, p or q, p and q*)

is given by a truth function, mapping all possible truth value assignments to the constituent propositions (p, q) to a truth value. The conditional is true, if and only if p, the antecedent, is false or q, the consequent, is true; otherwise it is false. That is, it is only false just where p is true and q is false. This is the *material impliciation* semantics of the conditional. This semantics licenses a variety of formal (i.e., content-independent) rules of inference. Despite the existence of a variety of formal rules that logically can be derived involving the conditional, the psychology of reasoning has typically concentrated its research effort on only two, the conditional syllogisms, *modus ponens* (MP) and *modus tollens* (MT). For these rules of inference, if the premises (above the line) are true ("¬" = not), then the conclusion (below the line) must be true, i.e., they are logically *valid*:

$$\text{MP} \quad \frac{p \rightarrow q,\ p}{\therefore q} \qquad\qquad \text{MT} \quad \frac{p \rightarrow q, \neg q}{\therefore \neg p}$$

In psychological reasoning experiments these valid rules of inference are usually paired with two logical fallacies, *denying the antecedent* (DA) and *affirming the consequent* (AC):

$$\text{DA} \quad \frac{p \rightarrow q, \neg p}{\therefore \neg q} \qquad\qquad \text{AC} \quad \frac{p \rightarrow q, q}{\therefore p}$$

Over the last 50 years the question of the quality of people's deductive reasoning has been pursued using a number of experimental paradigms. The three paradigms which have been most studied are the Wason selection task (Wason, 1968), quantified syllogistic reasoning (Johnson-Laird & Steedman, 1978), and conditional inference (Taplin, 1971). In each case, standard deductive logic makes precise predictions about people's performance.

Wason's selection task is like being asked to consider four birds, two of which you only know their species, one is a swan and one is a crow, and two of which you only know their colour, one is white and one is black. The question a participant must address is which birds must be examined to confirm or disconfirm that *all swans are white*. The logical form of this claim is $All(x)(Swan(x) \rightarrow White(x)$, i.e., a universally quantified conditional, which is only false if one finds a black swan. The question of which birds to look at has a determinate logical answer if it is assumed that the domain of x is restricted to just the four birds under consideration. Only the swan and the black bird could falsify the claim in this restricted domain and so only these birds must be examined. However, in experimental versions of this task using letters and

numbers, e.g., *if there is an A on a one side of a card, then there is a 2 on the other*, people mainly ask to see the reverse of the *A* and the *2* cards (Wason, 1968). That is, rather than attempting to falsify the hypothesis, people appear to choose evidence that might confirm it, by revealing a card with an *A* on one side and a *2* on the other. This behaviour was labelled *confirmation bias*.

Quantified syllogistic reasoning involves the logical quantifiers, *All P are Q*, *Some P are Q*, *No P are Q*, and *Some P are not Q* (capital *P* and *Q*s are used to distinguish these *predicates* from the *propositional* variables used in describing the conditional syllogisms). A quantified syllogism involves two of these statements as premises connected by a *middle term* (*Q*), e.g.,

$$\begin{array}{l} \textit{Some P are Q} \\ \underline{\textit{All Q are R}} \\ \therefore \textit{Some P are R} \end{array}$$

This is a logically *valid* syllogism. The *All* statement has the same conditional logical form as in the swan example. As the end terms (*P, Q*) and the middle term can assume four configurations (called *figures*) in the premises and there are 16 possible combinations of quantifiers, there are 64 possible syllogisms (512 if you include the conclusion and the order of end terms in the conclusion). If people responded logically in tasks where they are asked whether the conclusion follows logically from the premises, they should endorse the valid syllogisms and not endorse the invalid syllogisms. However, people show systematically graded behaviour, i.e., they reliably endorse certain valid syllogisms more than others. Moreover, they also endorse invalid syllogisms over which they also show systematically graded behaviour.

In conditional inference tasks, participants are given the two valid inference rules (MP and MT) and the two fallacies (DA and AC) and are asked which they wish to endorse. If they are reasoning logically, they will endorse the valid inferences but not the fallacies. However, people typically select MP more than MT. Moreover, they also select DA and AC but select AC more than DA and occasionally select more AC than MT (Schroyens & Schaeken, 2003). *Content* also matters. For example, people endorse the MP inference more for a conditional such as, *if the apple is ripe, then it will fall from the tree*, than for the conditional, *if John studies hard, then he will do well in the test* (Cummins, 1995). Furthermore, the difference can be directly located in the differential *defeasibility* of these two conditionals. It appears, for example, much easier for people to generate scenarios in which John will not do well in the test

(e.g., he is depressed, he has a low IQ…etc) than that the ripe apple remains forever stuck to the tree. As we pointed out above, defeasibility and effects of content are normally considered properties of inductive, rather than deductive inference.

In summary, the three experimental paradigms (conditional inference, the selection task, and syllogistic reasoning) that have been the main focus of empirical research in the psychology of reasoning reveal considerable deviations from logical expectations. The currently most active area of research is on conditional inference. This is because over the last 10 years or so there have been considerable theoretical and methodological advances in this area. Some are shared with other inferential modes but they can be best exemplified in conditional inference. Moreover, it is unequivocally agreed both in philosophical logic (e.g., Bennett, 2003) and in experimental psychology (e.g., Evans & Over, 2004) that the conditional is the core of human inference.

The response to apparently illogical responses in these tasks is to appeal to cognitive limitations and/or the nature of people's mental representations of these arguments. So, on a *mental logic* view (e.g., Rips, 1994), people tend to draw the MT inference less than MP, because they only have the MP rule of inference (see above) in their mental logic. Consequently, they must draw the MT inference using a *suppositional* strategy (they suppose the denial of the conclusion and show that this leads to a contradiction). This strategy is more cognitively demanding and so fewer participants complete it. On the *mental models* account, people have no mental inference rules but rather construct a mental representation of the possibilities allowed by a conditional over which they draw inferences. These possibilities relate directly to the truth conditions of the conditional: they are representations of the states of affairs in the world that are not ruled out assuming the conditional is true. Moreover, because of working memory limitations they do not mentally represent all of these possibilities:

$$p \qquad q \qquad\qquad (1)$$
$$\ldots$$

(1) shows the *initial* mental model for the conditional. It shows just the possibility that *p* is true and *q* is true but misses out the false antecedent possibilities (¬*p* *q* and ¬*p* ¬*q)* which are also true instances of the conditional. (1) allows MP, because the categorical premise, *p*, matches an item in the model which suggests it "goes with" *q*, the conclusion of MP. However, this model does not match the categorical

premise, $\neg q$, of the MT inference. (1) needs to be *fleshed out* with the false antecedent truth table cases for a match to be found for $\neg q$, which suggests it "goes with" $\neg p$. This extra mental operation makes the MT inference harder and so fewer people endorse it.

Yet how could such an error prone reasoning system have evolved? How could it lead to successful behaviour in the real world? Over the last ten years or so alternative accounts of human reasoning based on probability theory have been proposed which may address these questions. Moreover, they directly address the fact that inductive properties like content dependence and defeasibility arise, even when people are presumed to be solving deductive reasoning tasks. For example, interpreting *birds fly* to mean that the conditional probability that something flies given it is a bird, $P(fly(x)|bird(x))$, is high, e.g., .95, is consistent with the probability that something flies given that it is a bird and an Ostrich being zero or close to zero, i.e., $P(fly(x)|bird(x),Ostrich(x)) \approx 0$.

The source of these probability judgements is world knowledge (Oaksford & Chater, 2007). In truly deductive inference, people should ignore their prior knowledge of the specific content of the premises. Stanovich and West (2000) note that people find this difficult, calling this "the fundamental computational bias." Different theories take different approaches to addressing this bias. They vary from making adjustments to a theory designed to account for standard logical inference, as in mental models (Johnson-Laird and Byrne, 2002), to rejecting standard logic as the appropriate normative standard for these psychological tasks, as in the probabilistic approach (Oaksford & Chater, 2007).

Recently researchers have begun quantitatively to compare models of reasoning using a "model-fitting" approach---building mathematical accounts of different models, and testing how closely the predictions of these models fit the empirical data (in the selection task [Oaksford & Chater, 2003a; Klauer, Stahl, & Erdfelder, 2007], syllogistic reasoning [Klauer, Musch, & Naumer, 2000], and conditional inference [Oaksford & Chater 2003b; Oaksford, Chater, & Larkin, 2000; Oberauer, 2006; Schroyens & Schaeken, 2003]).

How have accounts of conditional reasoning responded to the apparent influence of content and defeasibility? The mental logic approach does not address the issue directly, and it has been suggested that such influences arise from non-deductive reasoning mechanisms (e.g., Rips, 1994, 2001). Mental model theory

addresses these issues directly. Johnson-Laird and Byrne (2002) argue that mental models of conditionals can be modulated by their prior knowledge that rules out or rules in various truth-functional possibilities. They call this process *semantic and pragmatic modulation*. The process of semantic and pragmatic modulation may even lead to the representation of possibilities that falsify the conditional, i.e., the $p \quad \neg q$. For example, they argue that a conditional such as, *if there is gravity* (*which there is*), *then your apples may fall* induces the following mental models:

$$p \qquad q \qquad (2)$$
$$p \qquad \neg q$$

The false antecedent possibilities are not considered because gravity is always present but on any given occasion the apples may or may not fall. Notice that the modal "may" here is represented simply by listing both consequents as possible, which appears radically oversimplified, from the standpoint of conventional logic. Johnson-Laird and Byrne (2002) discuss no less than ten possible interpretations of the conditional by showing how the ten different models they specify may capture the intended meanings of various examples that differ in content (see, Johnson-Laird & Byrne, 2002, p 667, Table 4). Each of these ten different models license different patterns of inference.

Johnson-Laird and Byrne (2002) relate their ten interpretations to specific examples that motivate the interpretations. Frequently, this involves the inclusion of *modal* terms like *possibly* or *may* in the consequent (*q*) clause that linguistically marks the fact that the consequent *may* not occur given the antecedent. This suggests that the surface form of the conditional can trigger the appropriate interpretation. This may directly involve accessing information from pragmatic world knowledge rather than indicating that such a search for a counterexample, e.g., a case where the apple does not fall, would be successful (see Schroyens and Schaeken, 2003 for an alternative viewpoint). One problem for this account is that it appeals directly to semantic and pragmatic intuitions in order to generate predictions; and indeed, the underlying mental models representations serve as a notation for describing these intuitions, rather than constrainting them.

New probabilistic approaches to conditional inference directly address defeasibility and effects of problem content by starting from a different normative theory of conditionals. The key idea is that the probability of a conditional, *if p then q*, is the conditional probability, $P(q|p)$. In probability logic (Adams, 1998), $P(q|p)$ is

given by the subjective interpretation provided by the *Ramsey Test*. As Bennett (2003, p. 53) says:

> "The best definition we have [of conditional probability] is the one provided by the Ramsey test: your conditional probability for *q* given *p* is the probability for *q* that results from adding $P(p) = 1$ to your belief system and conservatively adjusting to make room for it."

Recent evidence shows that people do regard the probability of a conditional to be the conditional probability (Evans, et al, 2003; Oberauer & Wilhlem, 2003; Over, et al, 2005). For example, Evans et al (2003), assessed people's probabilistic interpretations of conditional rules and their contrapositives (*if* $\neg q$ *then* $\neg p$). They tested three possibilities. First, material implication predicts that the probability of a conditional should be $1 - P(p,\neg q)$, i.e., 1 minus the probability of finding a falsifying case. Second, the conditional probability account predicts that the probability of a conditional should be $P(q|p)$. Finally, they test the possibility that the probability of the conditional is the joint probability, $P(p,q)$. According to material implication, conditionals and their contrapositives should be endorsed equally because they are logically equivalent. Consequently, there should be a strong correlation between ratings of how likely the conditional and its contrapositive are to be true. However, according to the conditional probability account, $P(q|p)$ and $P(\neg p|\neg q)$ can differ considerably and would not be expected to reveal a perfect correlation.

Evans et al (2003) varied $P(q|p)$, $P(\neg q|p)$ and $P(\neg p)$ by describing the distribution of cards in packs of varying sizes. For example, given a conditional *if the card is yellow then it has a circle printed on it*, participants were told that there are four yellow circles, one yellow diamond, sixteen red circles and sixteen red diamonds (Oaksford et al's [2000] used similar manipulations). So $P(q|p) = .8$, $P(\neg q|p) = .2$ and $P(\neg p) = 32/37$. On material implication, increases in $P(\neg p)$ should increase ratings of $P($*if the card is yellow then it has a circle printed on it*$)$, because if the conditional is true, if the antecedent is false; according to conditional probability, they should be independent of $P(\neg p)$; and according to conjunction interpretation, they should decrease with increases in $P(\neg p)$, because $\neg p$ implies that the conjunction is false. The evidence supported the conditional probability interpretation, with some evidence for a joint probability interpretation.

Recently, Over et al (2005) replicated these findings for everyday conditionals which were pre-tested for $P(p)$ and $P(q)$ as in Oaksford, Chater and Grainger (1999) and Oaksford et al (2000, Experiment 3). While replicating the effect of conditional probability, in contrast to Evans et al (2003), they found that the conjunctive interpretation was rarely adopted by participants. Consequently, the conjunctive interpretation is probably an artefact of unrealistic stimuli (Oberauer & Wilhelm, 2003).

The results seem to confirm that the probability of the conditional equals the conditional probability, i.e., $P(p \rightarrow q) = P(q|p)$, as Adams (1998) account of the probability conditional requires. However, it is the implications for inference of this change in normative focus that is of fundamental importance to the psychology of reasoning. Via the Ramsey test, the probability conditional reveals the total dependence of conditional inference on prior world knowledge. Most recent work has attempted to integrate these insights in to a psychological theory of conditional reasoning. We have already looked at the mental model approach (Johnson-Laird & Byrne, 2002). We now look at the other approaches that have been suggested.

Perhaps the most direct approach has been taken by Oaksford, Chater, and Larkin (2000, Oaksford & Chater, 2007). They have proposed a computational level account of conditional inference as dynamic belief update (Oaksford & Chater, 2007). So if a high probability is assigned to *if x is a bird, x flies*, then on acquiring the new information that *Tweety is a bird*, one's degree of belief in *Tweety flies* should be revised to one's degree of belief in *Tweety flies given Tweety is a bird*, i.e., one's degree of belief in the conditional. So using P_0 to indicate *prior* degree of belief and P_1 to indicate *posterior* degree of belief, then:

$$P_1(q) = P_0(q \mid p), \text{ when } P_1(p) = 1. \qquad (3)$$

Thus according to this account, the probability with which someone should endorse the MP inference is the conditional probability. This is the approach taken in Oaksford et al (2000).

However, as Oaksford and Chater (2007) point out there is a problem with extending this account to MT, DA and AC (Sober, 2002). The appropriate conditional probabilities for the categorical premise of these inferences to conditionalize on are $P(\neg p|\neg q)$, $P(\neg q|\neg p)$, and $P(p|q)$ respectively. However, the premises of MT and the fallacies do not entail values for these conditional probabilities (Sober, 2002, Sobel,

2004, Wagner, 2004). Oaksford et al (2000) suggested that people had prior knowledge of the marginals, $P(p)$ and $P(q)$, which together with $P(q|p)$ do entail appropriate values (see, Wagner [2004] for a similar approach) ($P_0(q|p) = a$, $P_0(p) = b$, $P_0(q) = c$):

$$\text{MP} \quad P_1(q) = P_0(q \mid p) = a \tag{4}$$

$$\text{DA} \quad P_1(\neg q) = P_0(\neg q \mid \neg p) = \frac{1 - c - (1 - a)b}{1 - b} \tag{5}$$

$$\text{AC} \quad P_1(p) = P_0(p \mid q) = \frac{ab}{c} \tag{6}$$

$$\text{MT} \quad P_1(\neg p) = P_0(\neg p \mid \neg q) = \frac{1 - c - (1 - a)b}{1 - c} \tag{7}$$

Equations (4) to (7) show the posterior probabilities of the conclusion of each inference assuming the posterior probability of the categorical premise is 1. As can be seen in Figure 1, this account provides a close fit with the empirical data.

In summary, recently the psychology of reasoning, and the psychology of conditional reasoning in particular, has shifted its focus away from the old debates about whether rule based mental logic approaches or mental models provided the best account of human inference. The emergence of computational level models framed in terms of probability theory rather than standard logic has fundamentally changed the questions being asked. The questions are now whether or how to incorporate these new insights. Should we view the probability conditional as a wholesale replacement for the standard logic based mental models approach? How does world knowledge modulate reasoning and/or provide probability information?

Decision Making

Whereas reasoning concerns how people use given information to derive new information, the study of decision making concerns how people's beliefs and values determine their choices. As we have seen, in the context of reasoning, there is fundamental debate concerning even the most basic elements of a normative framework against which human performance should be compared (e.g., whether the framework should be logical [e.g., Johnson-Laird & Byrne, 1991; Rips, 1994] or probabilistic [Oaksford & Chater, 2007]). By contrast, expected utility theory is fairly widely assumed to be the appropriate normative theory to determine how, in principle, people ought to make decisions.

Expected utility theory works by assuming that each outcome, *i*, of a choice can be assigned a probability, Pr(*i*) and a utility, *U*(*i*) and that the utility of an uncertain choice (e.g., a lottery ticket; or more generally, any action whose consequences are uncertain), is:

$$\sum \Pr(i)U(i) \qquad (8)$$

That is, expected utility of a choice is the sum of over outcome, *i*, of the utility *U*(*i*) of each outcome, weighted by its probability Pr(*i*) given that choice. Expected utility theory recommends the choice with the maximum expected utility.

This normative account is breathtakingly simple, but hides what may be enormous practical complexities—both in estimating probabilities and establishing what people's utilities are. Thus, when faced with a practical personal decision (e.g., whether to take a new job, which house to buy, whether or whom to marry), decision theory is not easy to apply—because the possible consequences of each choice are extremely complex, their probabilities ill-defined, and moreover, we often have little idea what preferences we have, even if the outcomes were definite. Thus, one difficulty with expected utility theory is practicability in relation to many real-world decisions. Nonetheless, where probabilities are utilities can be estimated with reasonable accuracy, expected utility is a powerful normative framework.

Can expected utility theory be used as an explanation not merely for how agents *should* behave, but of how agents actually *do* behave? Rational choice theory, which provides a foundation for explanation in microeconomics and sociology (e.g., Elster, 1986) as well as perception and motor control (Körding & Wolpert, 2006), animal learning (Courville, Daw & Touretzky, 2006) and behavioral ecology (Krebs & Davies, 1996), assumes that it does. This style of explanation involves inferring the probabilities and utilities that agents possess; and using expected utility theory to infer their choices according to those probabilities and utilities. Typically, there is no specific commitment concerning whether or how the relevant probabilities and utilities are represented—instead, the assumption is that preferences and subjective probabilities are "revealed" by patterns of observed choices. Indeed, given fairly natural consistency assumptions concerning how people choose, it can be shown that the observed pattern of choices can be represented in terms of expected utility—i.e.,

appropriate utilities and subjective probabilities can be inferred (Savage, 1954), with no commitment to their underlying psychological implementation. Indeed, this type of result can sometimes be used as reassurance that the expected utility framework is appropriate, even in complex real-world decisions, where people are unable explicitly to estimate probabilities or utilities.

The descriptive study of how people make decisions has, as with the study of reasoning, taken the normative perspective as its starting point; and aimed to test experimentally how far normative assumptions hold good. In a typical experiment, outcomes are made as clear as possible: for example, people may choose between monetary gambles, with known probabilities; or between gambles and fixed amounts of money.

A wide range of systematic departures from the norms of expected utility are observed in such experiments, as demonstrated by the remarkable research programme initiated by Kahneman, Tversky and their colleagues (e.g., Kahneman, Slovic & Tversky, 1982; Kahneman & Tversky, 2000). Thus, for example, people can be induced to make different decisions, depending on how the problem is "framed." Thus, if a person is given £10 at the outset, and told that they must choose either a gamble, with a 50% chance of keeping the £10, and a 50% chance of losing it all; or they must give back £5 for certain, they tend to prefer to take the risk. But if they are given no initial stake, but asked whether they prefer a 50-50 chance of £10, or a certain £5, they tend to play safe. Yet, from a formal point of view these choices are identical—the only difference is that in one case the choice is framed in terms of losses (where people tend to be risk-seeking); rather than gains (where they tend to be risk-averse).

Expected utility theory cannot account for framing effects of this type—only the formal structure of the problem should matter, from a normative point of view; the way in which it is described should be irrelevant. Indeed, expected utility theory can't well account for the more basic fact that people are not risk neutral (i.e., neutral between gambles with the same expected monetrary value) for small stakes (Rabin, 2000). This is because, from the standpoint of expected utility theory, people ought to evaluate the possible outcomes of a gamble in "global" terms—i.e., in relation to the impact on their life overall. Hence, if a person has an initial wealth of £10,000, then both the gambles above amount of choosing between a 50-50 chance of ending up with a wealth of £10,010 or £10,000, or a certain wealth of £10,005 (see, Figure 2).

One reaction to this type of clash between human behaviour and rational norms is the observation that the human behaviour is error-prone—and hence, where this is true, expected utility will be inadequate as a *descriptive* theory of choice. A natural follow-up to this, though, is to attempt to modify the normative theory so that it provides a better fit with the empirical data. A wide range of proposals of this sort have been put forward, including from prospect theory (Kahneman & Tversky, 1979). Indeed, prospect theory, by far the most influential framework, was deliberately conceived as an attempt to find the minimal modifications of expected utility theory that would describe human choice behavior (Kahneman, 2000).

In essence, prospect theory modifies expected utility theory in three main ways. First, monetary outcomes are considered in isolation, rather than aggregated as part of total wealth. This fits with the wider observation that people view different amounts of money, or indeed goals, quantities or events of any kind, one-by-one, rather than forming a view of an integrated whole. This observation is the core of Thaler's (1985) "mental accounting" theory of how people make real-world financial decisions.

Second, prospect theory assumes that while the value function (i.e., relating money to subjective value) for positive gains is concave (i.e., negatively accelerating, indicating risk aversion in an expected utility framework), the value function for losses is convex (i.e., positively accelerating, see Figure 3a). This implies that the marginal extra pain for an additional unit of loss (e.g., each extra pound or dollar lost) decreases with the size of the loss. Thus, people are risk-seeking when a gamble is framed in terms of losses, but risk-averse when it is framed in terms of gains, as we noted above. Moreover, the value function is steeper for losses than for gains, which captures the fact that most people are averse to gambles with a ½ chance of winning £10, and a ½ chance of losing -£10 (Kahneman & Tversky, 1979). This phenomenon, *loss aversion*, has been used to explain a wide range of real world phenomena, including the status quo bias (losing one thing and gaining another tends to seem unappealing, because the loss is particularly salient, Samuelson and Zeckhauser, 1988) and the equity premium puzzle (share returns may be "unreasonably" high relative to fixed interest bonds, because people dislike falls in stock prices more than they like the equivalent gains, Benartzi and Thaler, 1995).

The final key modification of expected utility theory is that prospect theory assumes that people operate with a distorted representation of probability (Figure 3b).

They overestimate probabilities near zero; and underestimate probabilities near 1, such that the relation between probability, $p(i)$ and the "decision weights," $w(i)$, which are assumed to determine people's choices, as related by an inverse-S shape. According to prospect theory, this distortion can explain the so-called "four-fold pattern" of risky decision making---that, for small probabilities, risk-preferences reverse both for gains and losses (Table 1). That is, while people are normally risk-averse for gains, they still play lotteries---according to prospect theory, this is because they drastically overestimate the small probability of winning. Similarly, while people are normally risk-seeking for losses, they still buy insurance---according to prospect theory, this is because the drastically overestimate the small probability of needing to claim on that insurance.

The machinery of prospect theory integrates values and decision weights to assign a value to each gamble (where this is any choice with an uncertain outcome), just as in expected utility theory, so that the value of a risky option is:

$$\sum w(i)v(i) \qquad (9)$$

where $w(i)$ is the decision weight (i.e., distorted probability) for outcome i; and $v(i)$ is the value of that outcome. Thus, the value of a risky option is the sum of the products of the subjective value of each possible outcome and the subjective "weight" (distorted probability) assigned to each outcome. Prospect theory, and other variants of expected utility, hold with the assumption that people represent value and probability on some kind of absolute internal scale; and that they integrate these values by summing the product of weight and value over possible outcomes, to obtain the value of each gamble. Two recent psychological theories, however, set aside the structure of expected utility theory; they are inspired not by the attempt to modify normative considerations, but instead to trace the consequences of assumptions about the cognitive system.

One recent approach (Brändstatter, Gigerenzer & Hertwig, 2006) focuses on processing limitations, and on the consequences of assuming that the cognitive system is not able to integrate different pieces of information, and that, instead, people can only focus on one piece of information at a time. This assumption is controversial. In perceptual judgements (e.g., concerning the identity of a phoneme, or the depth of a surface), many theories explicitly assume (linear) integration between difference

sources of information (Schrater & Kersten, 2000)---in a probabilistic framework, this corresponds, roughly, to adding logs of the strength of evidence provided by each cue. Many models of higher-level judgement have assumed that information is also integrated, typically linearly (e.g., Hammond, 1996). However, Gigerenzer and colleagues (e.g., Gigerenzer & Goldstein, 1996) have influentially argued that high-level judgements---most famously, concerning the larger of pairs of German cities---do not involve integration. Instead judgement is assumed involve considering cues, one at a time---if a cue determines which city is likely to be larger, that city is selected; if not, a further cue is chosen, and the process is repeated. There has been considerable, and on-going, controversy concerning the circumstances under which integration does or does not occur, in the context of judgement (Hogarth & Karelaia, 2005).

Brändstatter, Gigerenzer and Hertwig's (2006) innovation is to show that a non-integrative model can make in-roads into understanding how people make risky decisions---a situation which has been viewed as involving a trade-off between "risk" and "return" almost by definition. Their model, the priority heuristic, has the following basic form. For gambles which contain only gains (or £0), the heuristic recommends considering features of the gambles in the order: minimum gain, probability of minimum gain, maximum gain. If gains differ by at least 1/10 of the maximum gain (or, for comparison of probabilities, if probabilities differ by at least 1/10), choose the gamble which is "best" on that feature (defined in the obvious way). Otherwise move to the next feature in the list, and repeat.

To see how this works, consider the gambles illustration the "four-fold" pattern of risky choice, described by Kahneman and Tversky (1979), in Table 1. For the high probability gamble over gains, the minimum gain for the certain outcome is £500; but the minimum gain for the risky gamble is £0; this difference is far more than 1/10 of the maximum gain, £1000. Hence, the safe option is preferred. By contrast, for the low probability gamble, the difference between the minimum gains for the options is just 50p, which is much less than 1/10 of the maximum gain of £1000. Hence, this feature is abandoned, and we switch to probability of minimum gain—this is clearly higher for a certain gamble---as there is only one outcome, which is by definition the minimum. The risky gamble, with the smaller probability of minimum gain, is therefore preferred. Thus, we have risk seeking behavior with small

probabilitities of large gains (and hence an explanation of why people buy lottery tickets).

Brändstatter, Gigerenzer and Hertwig propose a modification of the heuristic for gambles containing just losses, where "gain" is replaced by "loss" throughout, so that the feature order is: minimum loss, probability of minimum loss, maximum loss. If gains differ by at least 1/10 of the maximum loss (or probabilities differ by at least 1/10), choose the gamble which is "best" on that feature (defined in the obvious way). Otherwise move the next feature in the list, and repeat. Tracing through the argument described above for the "loss" gambles in Table 1, yields the conclusion that people should appear risk-seeking for losses, except where there is a small probability of a large loss; here people will again be risk-averse (e.g., they will buy insurance).

The priority heuristic model does, however, make some extremely strong and counterintuitive predictions—e.g., that if the minimum gains differ sufficiently, then all other features of the gambles (including the probability of obtaining those gains) will have no impact on choice. In extreme cases, this seems implausible. For example, a certain 11p should be preferred to a .999999 probability of £1 (and otherwise £0). Brändstatter, Gigerenzer and Hertwig (2006) restrict their account, however, to cases for which the expected values of the gambles are roughly comparable—where they are not, the gamble with the obviously higher expected value is chosen, and the priority heuristic is not invoked.

Another recent approach to risk decision making, starting from cognitive principles rather than a normative economic account, is Decision by Sampling (DbS, Stewart, Chater & Brown, 2006). This viewpoint assumes that people have no underlying internal "scales" for utility or probability—but nonetheless, it turns out to be possible to reconstruct something analogous to the value and decision weight functions from prospect theory. If people assess the gut feel of a magnitude in relation to prior examples, the statistical distribution of such magnitudes is likely to be important. Other things being equal, this distribution will provide an estimate of the probabilities of different comparison items being considered in particular judgements. Thus, if small sums of money are much more commonly encountered than large sums of money, then it is much more likely that people will consider small sums of money as comparison items, other things being equal. Therefore, the difference in "gut" feel between £5 and £50 will be much greater than that between £1005 and £1050, because sampling an item in the first interval (so that the lower and upper items will

be assigned different ranks), is much more likely than sampling in the second. More generally, the attractiveness of an option, according to DbS, is determined by its rank in the set of comparison items; and hence, its typical attractiveness (across many sampling contexts) can be estimated by its rank position in a statistical sample of occurrences of the relevant magnitude. Figure 4a shows a sample of "positive" sums of money---credits into accounts from a high street bank; plotting monetary value against rank (Figure 4b) then produces a concave function, reminiscent of those in utility theory and prospect theory. Thus, the "gut" attractiveness of a sum of money is, on average, a diminishing function of amount. The similar analysis for losses (using bank account debits as a proxy) yields a convex function of value against losses, as in prospect theory. Moreover, for losses, the statistical distribution is more skewed towards small items, which has the consequence that ranks change more rapidly for small values for losses than for gain. This corresponds to a steeper value curve for losses and gains, and hence captures loss aversion. Indeed, putting the curves of rank against value together (Figure 4c) yields a curve strikingly reminiscent of that postulated in prospect theory.

In both reasoning and decision making, indeed, there is a certain air of paradox in human performance (Oaksford & Chater, 1998). Human common-sense reasoning is far more sophisticated than any current artificial intelligence models can capture; yet people's performance on, e.g., simple conditional inference, while perhaps explicable in probabilistic terms, is by no means effortless and noise-free; and similarly, in decision making, it appears that "low-level" repeated decision making may be carried out effectively. But perhaps this situation is not entirely paradoxical. It may be that both human reasoning and decision making function best in the context of highly adapted cognitive processes such as basic learning, deploying world knowledge, or perceptuo-motor control. Indeed, what is, striking about human cognition is the ability to handle, even to a limited extent, reasoning and decision making in novel, hypothetical, verbally stated scenarios, for which our past experience and evolutionary history may have provided us with only minimal preparation.

References

Adams, E. W. (1998). *A primer of probability logic.* Stanford: CLSI Publications.

Benartzi, S., & Thaler, R.H (1995). Myopic loss aversion and the equity premium puzzle. *Quarterly Journal of Economics*, *110*, 73-92.

Bennett, J. (2003). *A philosophical guide to conditionals.* Oxford England: Oxford University Press.

Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: Making choices without trade-offs. *Psychological Review*, *113*, 409-432.

Courville, A. C., Daw, N. D., & Touretzky, D. S. (2006). Bayesian theories of conditioning in a changing world. *Trends in Cognitive Sciences*, *10*, 294-300.

Cummins, D. D. (1995). Naïve theories and causal deduction. *Memory & Cognition, 23*, 646-658.

Elster, J. (Eds.), (1986). *Rational choice*. Oxford: Basil Blackwell.

Evans, J.St.B. T., Handley, S. H., & Over, D. E. (2003). Conditionals and conditional probability. *Journal of Experimental Psychology: Learning, Memory and Cognition, 29,* 321-355.

Evans, J.St.B.T., & Over, D.E. (2004). *If.* Oxford, England: Oxford University Press.

Gigerenzer, G. & Goldstein, D (1996) Reasoning the fast and frugal Way: Models of bounded rationality. *Psychological Review*, *103*, 650-669.

Goodman. (1954). *Fact, fiction, and foreca*st. London: The Athlone Press.

Hammond, K. R. (1996). *Human judgment and social policy: Irreducible uncertainty, inevitable error, unavoidable injustice*. Oxford: Oxford University Press.

Hogarth, R. M., & Karelaia, N. (2005). Simple models for multi-attribute choice with many alternatives: When it does and does not pay to face trade-offs with binary attributes. *Management Science, 51*, 1860-1872.

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

Johnson-Laird, P.N., & R.M.J. Byrne. (1991). *Deduction.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review, 109*, 646-678.

Johnson-Laird, P. N., & Steedman, M. (1978). The psychology of syllogisms. *Cognitive Psychology, 10,* 64–99.

Kahneman, D. (2000). Preface. In D. Kahneman & A. Tversky, (Eds.), *Choices, values and frames* (pp. ix-xvii). New York: Cambridge University Press and the Russell Sage Foundation.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.), (1982). *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decisions under risk. *Econometrica, 47,* 313-327.

Kahneman, D., & Tversky, A. (Eds.), (2000). *Choices, values and frames.* New York: Cambridge University Press and the Russell Sage Foundation.

Klauer, K. C., Musch, J., & Naumer, B. (2000). On belief bias in syllogistic reasoning. *Psychological Review, 107,* 852-884.

Klauer, K. C., Stahl, C., Erdfelder, E. (2007). The abstract selection task: New data and an almost comprehensive model. *Journal of Experimental Psychology: Learning, Memory and Cognition, 33,* 680-703.

Körding, K. P. & Wolpert, D. M. (2006). Bayesian decision theory in sensorimotor control. *Trends in Cognitive Sciences, 10,* 319-326.

Krebs, J. R., & Davies, N. (Eds.) (1996). *Behavioural ecology: An evolutionary approach* (4th edition). Oxford: Blackwell.

Oaksford, M., & Chater, N. (1998). *Rationality in an uncertain world*. Psychology Press: Hove, England.

Oaksford, M., & Chater, N. (2003a). Optimal data selection: Revision, review and re-evaluation. *Psychonomic Bulletin & Review, 10,* 289-318.

Oaksford, M., & Chater, N. (2003b). Conditional probability and the cognitive science of conditional reasoning. *Mind & Language, 18,* 359-379.

Oaksford, M. & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford: Oxford University Press.

Oaksford, M., Chater, N., & Grainger, B. (1999). Probabilistic effects in data selection. *Thinking and Reasoning, 5,* 193-244.

Oaksford, M., Chater, N., & Larkin, J. (2000). Probabilities and polarity biases in conditional inference. *Journal of Experimental Psychology: Learning, Memory and Cognition, 26,* 883-889.

Oberauer, K. (2006). Reasoning with conditionals: A test of formal models of four theories. *Cognitive Psychology, 53,* 238-283.

Oberauer, K. & Wilhelm, O. (2003). The meaning(s) of conditionals: Conditional probabilities, mental models and personal utilities. *Journal of Experimental Psychology: Learning, Memory and Cognition,* 680-739.

Rabin, M. (2000). Diminishing Marginal Utility of Wealth Cannot Explain Risk Aversion. In D. Kahneman & A. Tversky (Eds.) *Choices, Values, and Frames* (pp. 202-208). New York: Cambridge University Press.

Rips, L. J. (1994*). The Psychology of proof.* Cambridge, MA: MIT Press.

Rips, L. J. (2001). Two kinds of reasoning. *Psychological Science, 12,* 129-134.

Samuelson, W. F. & Zeckhauser, R. J. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty, 1,* 7-59.

Savage, L.J. (1954). *The Foundations of Statistics*. New York, NY: Wiley.

Schrater, P. R. & Kersten, D. (2000). How optimal depth cue integration depends on the task. *International Journal of Computer Vision, 40,* 71-89.

Schroyens, W., & Schaeken, W. (2003). A critique of Oaksford, Chater and Larkin's (2000) conditional probability model of conditional reasoning. *Journal of Experimental Psychology: Learning, Memory and Cognition, 29,* 140-149.

Sobel, J. H. (2004). *Probable modus ponens and modus tollens and updating on uncertain evidence.* Unpublished manuscript, Department of Philosophy, University of Toronto, Scarborough.

Sober, E. (2002). Intelligent design and probability reasoning. *International Journal for Philosophy of Religion, 52,* 65-80.

Stanovich, KE., & West, R.F (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences, 23,* 645-665.

Stewart, N., Chater, N., & Brown, G. D. A. (2006). Decision by sampling. *Cognitive Psychology*. *53,* 1-26.

Taplin, J.E.(1971). Reasoning with conditional sentences. *Journal of Verbal Learning and Verbal Behavior, 10,* 219-225.

Thaler, R. (1985). Mental accounting and consumer choice. *Marketing Science, 4,* 199-214.

Wagner, C. G. (2004). Modus tollens probabilized. *British Journal for Philosophy of Science,* 55, 747-753.

Wason, P.C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology, 12*, 129-140.

Wason, P.C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology, 20,* 273-281.

Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of Reasoning: Structure and Content.* London: Batsford.

Figure and table captions

Figure 1 *The behaviour of Oaksford et al's (2000) conditional probability model*

How the posterior probability of the conclusion (P_1(Conclusion)) varies as a function of the prior probability of the conclusion (P_0(Conclusion)) and the prior probability of the categorical premise (P_0(Premise)) for DA, AC and MT with P_1 (Premise) = 1.

Figure 2. *The utility function in conventional expected utility theory*.

The utility function is usually assumed to have a convex shape, as shown, although this is not an essential part of the theory. A concave utility function implies that the utility of, say, £50, is greater than the average utility of £0 and £100. This implies risk aversion---because risky options involve such averaging of good and poor outcomes. Note that expected utility theory applies to overall wealth, rather than directly to the outcomes of the gambles. If the gambles are small in relation to overall wealth, this implies that the utility curve is fairly flat, and hence that risk average should be small. The high levels of risk aversion shown in laboratory experiments are difficult reconcile with expected utility theory (Rabin, 2000).

Figure 3. *The value and probability weighting functions in prospect theory*.

The value function, (a), in prospect theory is concave in gains, and convex in losses, implying risk-aversion for gambles with positive outcomes; and risk-seeking choices for gambles with negative outcomes. The slope of the value function is steeper in losses than gains, implying loss-aversion. "Decision weights" (b) are presumed to be systematically distorted with respect to "true" probabilities, in an inverse-S shape, such that probabilities near zero are overweighted and probabilities near one are underweighted. This weighting aims to explain why people buy lottery tickets, despite basic risk aversion for gains; and why the buy insurance, despite basic risk-seeking preference for losses. In both cases, low probabilities (of winning, or of needing to file an insurance claim) are presumed to be overestimated.

Figure 4. *Decision by Sampling, and Money*

Decision by sampling assumes that people evaluate dimensions such as amount of money, probability, time or quality, in terms of their *ranking* against other items of the same type. Here, we consider the subjective value of money from this viewpoint. Panel a. shows the distribution of credits in a UK bank account, which we treat as a proxy for the distribution of positive sums of money that people encounter. Panel b. shows the sum data in cumulative form---plotting sum of money against the relative rank of that sum of money, in this distribution. Note that this curve mirrors the concave utility curve (Figure 2), typically used to explain risk aversion in the expected utility framework. Panel c. shows the result of extending the analysis to losses, using bank debit data. The resulting function is strikingly similar to that postulated in prospective theory (Figure 3a), but derived purely from environmental structure. (Reprinted with permission from Stewart, N., Chater, N. & Brown, G. D. A. (2006). Decision by sampling. *Cognitive Psychology*. *53*, 1-26.).

Table 1. *The four-fold pattern of risky choice*.

For moderate probabilities, people are presumed to be risk-seeking for gains, and risk-averse for losses (Kahneman & Tverksy, 1979). These patterns may be reversed when probabilities are small, according to prospect theory, because these small probabilities are substantially overweighted.

Figure 1

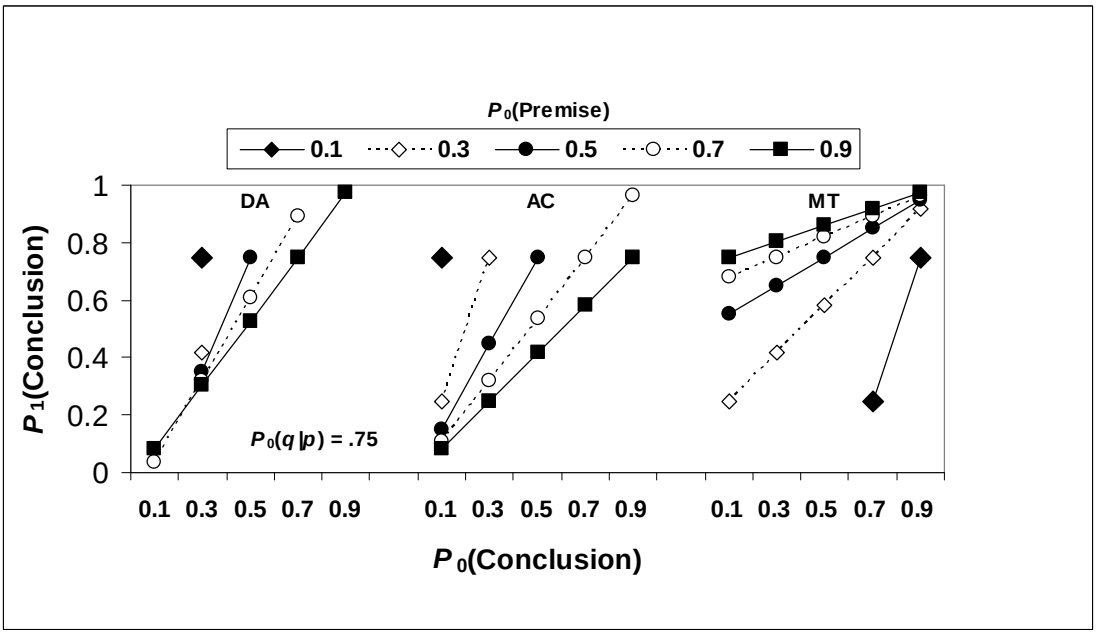

P0(Premise)
0.1
0.3
0.5
0.7
0.9
DA
AC
MT
P1(Conclusion)
P0(q|p) = .75
1
0.8
0.6
0.4
0.2
0
0.1 0.3 0.5 0.7 0.9
0.1 0.3 0.5 0.7 0.9
0.1 0.3 0.5 0.7 0.9
P0(Conclusion)

Figure 2

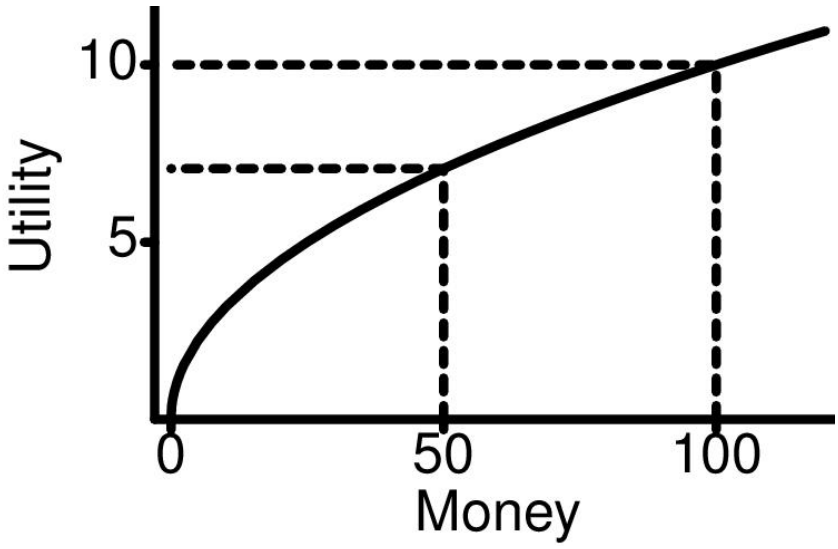
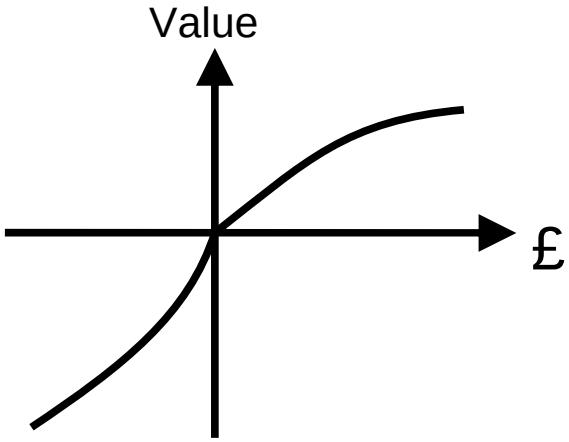


Figure 3

a.



b.

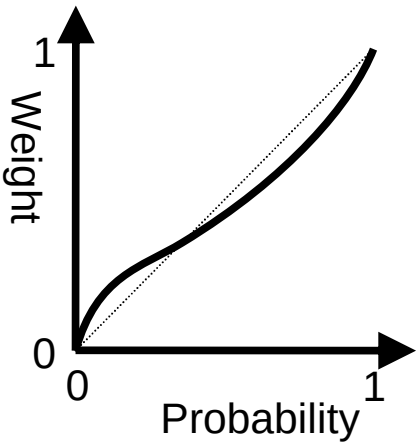
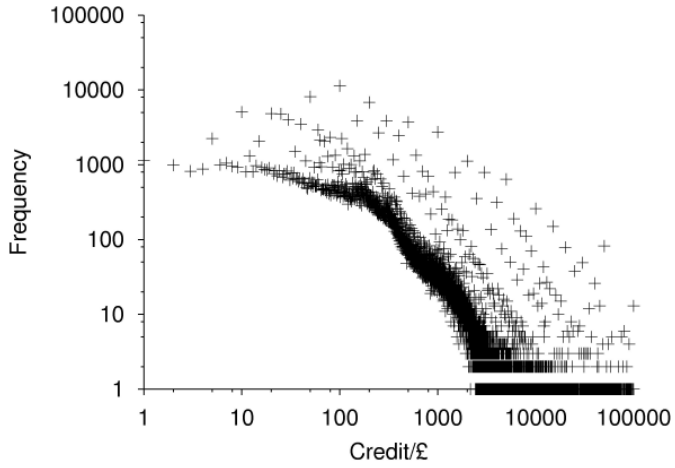
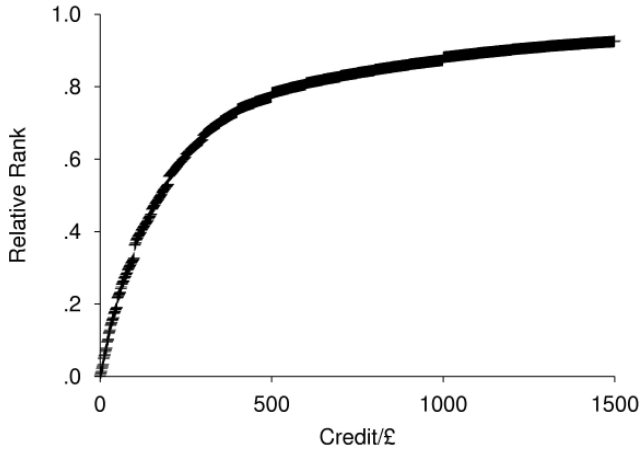


Figure 4

a.


100000
10000
1000
100
10
1
Frequency
1
10
100
1000
10000
100000
Credit/£

b.


1.0
.8
.6
.4
.2
.0
Relative Rank
0
500
1000
1500
Credit/£

c.

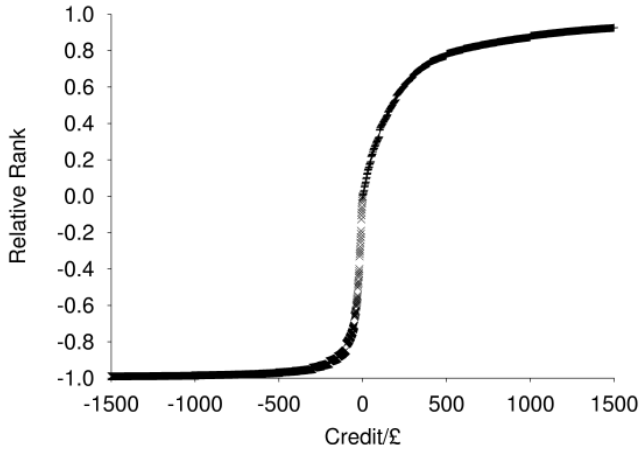

1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
Relative Rank
-1500
-1000
-500
0
500
1000
1500
Credit/£

Table 1

	Small probability	High probability
Gain	certain 50p vs 1/2000 probability of £1,000 choose gamble (risk seeking)	certain £500 vs 1/2 probability of £1,000 choose certainty (risk aversion)
Loss	certain -50p vs 1/2000 probability of -£1,000 choose certainty (risk aversion)	certain -£500 vs 1/2 probability of -£1,000 choose gamble (risk seeking)

Further Reading

Adler, J. E., & Rips, L. J. (2008). *Reasoning*. Cambridge: Cambridge University Press. (Collection of summaries and classic papers in the field)

Baron, J. (2008). *Thinking and deciding* (4th edition). Cambridge: Cambridge University Press. (A classic textbook, but thoroughly updated)

Braine, M. D. S., & O'Brien, D. P. (1998). *Mental logic*. London: Taylor & Francis. (Edited collection of papers on the mental logic theory)

Evans, J. St.B., T., & Over, D. E. (2004). *If*. Oxford: Oxford University Press. (An extended account of how different logical analyses of the conditional bear on explaining the experimental data on conditional reasoning)

Evans, J. St.B., T. (2007). *Hypothetical thinking*. Brighton: Psychology Press. (An exhaustive review of the psychology of reasoning and how the "suppositional" theory can explain it)

Gigerenzer, G. & Selten, R. (2001). Bounded rationality: The adaptive toolbox.. Cambridge, MA: MIT Press (An interdisciplinary collection on boundedly rational models of decision making).

Johnson-Laird, P. N. (2006). *How we reason*. Oxford: Oxford University Press. (The most recent instantiation of how the mental models theory accounts for the experimental data on human reasoning)

Kahneman, D. & Tversky, A. (Eds.) (2000). *Choices, frames and values*. Cambridge: Cambridge University Press. (A classic collection on the heuristics and biases approach to decision making).

Oaksford, M., & Chater, N. (2007). *Bayesian rationality*. Oxford: Oxford University Press. (An account of the conceptual underpinnings of the probabilistic approach and of how it explains the experimental data on human reasoning)

Glossary

Inductive reasoning: Reasoning in which the truth of the premises confers only a higher plausibility on the conclusions

Deductive reasoning: Reasoning in which the truth of the premises guarantees with certainty that the conclusion is true

Defeasible reasoning: Reasoning in which the a conclusion may be overturned by subsequent information.

Mental logic: The psychological theory of reasoning that assumes that various syntactic inference rules form part of the architecture of the mind.

Mental models: The psychological theory of reasoning that assumes that humans are logical in principle but because they represent logical terms by the possible states of affairs they allow reasoning can be systematically biased.

Probabilistic approach: The psychological theory of reasoning that assumes that human mind is adapted to dealing with an uncertain world and so human reasoning mechanisms are probabilistic.