THE ROYAL SOCIETY

Published online 26 November 2004

A cognitive neuroscience framework for understanding causal reasoning and the law

Jonathan A. Fugelsang* **and Kevin N. Dunbar**

Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH 03755, USA

Over the past couple of decades, there have been great developments in the fields of psychology and cognitive neuroscience that have allowed the advancement of our understanding of how people make judgements about causality in several domains. We provide a review of some of the contemporary psychological models of causal thinking that are directly relevant to legal reasoning. In addition, we cover some exciting new research using advanced neuroimaging techniques that have helped to uncover the underlying neural signatures of complex causal reasoning. Through the use of functional imaging, we provide a first-hand look at how the brain responds to evidence that is either *consistent* or *inconsistent* with one's beliefs and expectations. Based on the data covered in this review, we propose some ideas for how the effectiveness of causal reasoning, especially as it pertains to legal decision-making, may be facilitated.

Keywords: causation; causality; legal; decision-making; functional magnetic resonance imaging; neuropsychology

1. INTRODUCTION

The human mind has evolved many cognitive tools including abstraction, counterfactual thought, deduction and induction, for a vast variety of circumstances that are applied differently depending on the task at hand. Central to these domain general processes is causal thinking. An individual's ability to determine if a precipitating event was the cause of an outcome is essential for making sense of the complex world in which we live. Indeed, many of the learning and evaluative processes in which individuals engage pertain to the development and testing of causal models portraying the relationship between variables of interest (Dunbar 1995; Fugelsang *et al.* 2004). Such causal reasoning processes are evident in tasks ranging from simple everyday reasoning, such as why one's computer crashes, to complex scientific discovery, such as the formulation of the 'Universal Law of Gravitation' in the *Principia* by The Royal Society's former president Isaac Newton (Newton 1999).

This ability to infer causality is not only crucial for human reasoning in general, but also more specifically to the application of law. The legal system often asks lawyers, jurors and judges to determine if an individual's actions were responsible for a specific outcome. In so doing, the legal system requires that individuals reason about the evidence presented to them in an *unbiased* manner, formulating a judgement of causality if and only if the evidence presented to them overwhelmingly depict the acts of the defendant as causally responsible for the outcome under question.

* Author for correspondence (jonf@dartmouth.edu).

One contribution of 16 to a Theme Issue 'Law and the brain'.

In the present essay, we address two main questions that are directly relevant to such legal reasoning: (i) what sources of information do people use to evaluate causality, and (ii) to what degree do people evaluate evidence about causality in an *unbiased* manner? To answer these questions we will discuss both behavioural and fMRI experiments conducted in our laboratory and others.

2. LEGAL DECISION-MAKING AND CAUSAL REASONING

Whereas much decision-making research in the legal domain has focused on the content-oriented (i.e. substance of the trial itself) aspects of legal decision-making (e.g. Matlon 1986), research on the extra-legal aspects of legal decision-making (i.e. decision-making strategies of individuals) has primarily been conducted by cognitive and social psychologists (see Pennington & Hastie (1990) for a comprehensive review). Perhaps not surprisingly, much of this work has focused on criminal jury decision-making processes, especially as they pertain to judgements involving more serious crimes often involving capital punishment (e.g. Constanzo & Constanzo 1994; Wiener *et al.* 1995). The ability of jurors to make decisions in a non-arbitrary and unbiased manner in these situations is of obvious importance. Courts ask jurors to set aside personal beliefs and biases to make judgements in favour of, or against, a defendant based solely on the facts of the case presented to them. This *unbiased* application of the law is crucial, not only for obvious judicial reasons, but also to maintain consistency among rulings within and across jurisdictions. As alluded to in § 1, one's ability to make such judgements is directly related to one's ability to effectively attribute causality when presented with evidence. This evidence can come from a variety of sources and often involves the construction of causal chains of events, whereby the link

between the actions of the defendant and the outcome under question may be separated by several intermediate variables, each with a specific probability of occurring (Einhorn & Hogarth 1986). Indeed, this is not an easy task.

Researchers in both cognitive and social psychology have developed several models that capture different aspects of this causal reasoning process. The predominant view of causality in the psychological literature over the past two decades has dealt with the extent to which people induce causality based on observed statistical covariation-based evidence (e.g. Jenkins & Ward 1965; Rescorla 1968; Kelley 1973; Allan & Jenkins 1980; Einhorn & Hogarth 1986; Cheng & Novick 1990, 1992; Cheng 1997; White 2002). These models of causality stem from the Humean philosophy of radical empiricism (Hume 1978), which is based on the assumption that events that covary are more likely to be judged as causally related than events that do not covary. Such models typically propose computational algorithms by which reasoners are thought to derive an estimate of causality from such observable correlated events, and delineate conditions under which the covariation between variables of interest can be used to infer causality.

There are several different models of causal reasoning, each proposing different accounts of how people induce causality based on observed statistical covariation. Most derive estimates of the degree of covariation based on computing statistical contrasts between the presence and absence of events and outcomes. Perhaps the most popular contemporary model of causality that is based on this covariation principle in the recent psychological literature is the Power PC theory (Cheng 1997; Novick & Cheng 2004). According to this model, an individual considers both the probability of the effect occurring in the presence of the cause [$P(e/c)$] and the probability of the effect occurring in the absence of the cause [$P(e/\sim c)$]. Specifically, Cheng and her colleagues propose that the perceived causal relationship between variables of interest is a function of both the cause's covariation with the given effect [quantified as $P(e/c)-P(e/\sim c)$] and the inverse of the base rate [$1-P(e/-c)$] of that effect. That is, reasoners are suggested to view an individual or an event as a cause of a specific outcome to the extent that the individual or event raises the likelihood of the outcome above some baseline of the outcome occurring when the individual or event is absent. Specifically, the extent to which smoking may be judged as causally responsible for the incidence of lung cancer in a group of patients would depend, not only on the degree to which smoking and lung cancer co-occur together, but also the degree to which lung cancer occurs in the absence of smoking. Proponents of this and other similar covariation-based models (e.g. Cheng & Novick 1990, 1992; White 2002) claim that people are sensitive to the covariation between the cause and effect, and then use this information to derive a measure of the causal link or liability of the causal candidates in question.

An alternative account for how people may judge an individual or an event as liable for an outcome concerns the extent to which that individual or outcome is judged to be a rare or an abnormal event in the given situation. Hilton & Slugoski (1986) have proposed such a model, the Abnormal Conditions Focus model. They propose the causal inference progresses in two stages. The first stage is proposed to involve a judgement about the degree to which a candidate is perceived to be *necessary* for the occurrence of the effect. For example, both oxygen and flame are necessary, but not sufficient variables required for paper to burn; in the absence of either event, paper would not burn. Consequently, both causes would be selected during the first stage of reasoning. The second stage of reasoning involves selecting the abnormal variable from the set of *necessary* causes identified by the first stage of reasoning. That is, a cause is selected that departs from that which is normal for the given circumstances. Hilton and Slugoski provide a helpful example that clarifies these two components. They note that the speed of a train, the weight of the railway cars, and a faulty rail are all necessary components for a train to derail. However, the faulty rail is the one component that is likely to be selected as causally relevant, because it is the single feature from the set that is abnormal for the everyday operation of trains.

Whereas the former models stress the role of observed covariation-based evidence, other cognitive models have examined the degree to which people judge causality based on their beliefs and expectations about what events have the *capacity* or *power* to produce specific outcomes. These models stem from the philosophical tradition of Immanuel Kant (1965) who proposed that causality is an inherent law of nature, not merely an emergent property of statistical regularity. Harre & Madden (1975; see also Hart & Honore 1959; White 1989, 1995; Ahn *et al.* 1995) have elaborated on this philosophical tradition by positing that certain objects are perceived to possess stable properties whose power to produce a specific outcome is based on the 'chemical, physical, or genetic natures of the entities involved' (p. 5). For example, individuals may judge smoking to be causally related to the development of lung cancer owing to their beliefs about the carcinogenic properties of inhaling tobacco, independently of the degree to which they may be thought to correlate in the actual environment.

Harre and Madden further elucidate the relationship between causal agents and enabling conditions. That is, a specific causal outcome is thought to occur only under the appropriate enabling conditions. For example, an individual may possess the disposition to commit a violent crime; however, this disposition may only result in a violent act if the individual is intoxicated. In this example, *intoxication* would act as the enabling or releasing condition that allows the disposition of the individual (that is, to be aggressive) to be released. Based on these defining features of causality, causal roles are defined conceptually, rather than based solely on observing correlations between variables in the environment. The assessment of causal hypotheses, therefore, is thought to be mainly a matter of seeking some object believed to possess the power to produce the effect in question and then determining if the appropriate releasing conditions are present to enable the power of the object to exert the effect (see Dunbar (2002) for an example of this phenomenon in scientists reasoning 'live' in their laboratories). In many cases, this search for an object that possesses the power to produce a specific outcome may supersede the search for evidence about the covariation between variables of interest (e.g. White 1989; Ahn *et al.* 1995).

3. INTEGRATING EVIDENCE WITH ONE'S BELIEFS AND EXPECTATIONS

Recently, our laboratory and others have been conducting several behavioural experiments examining the degree to which an individual's beliefs and expectations (derived from information about the inherent properties of objects) influence how they make causal decisions about covariation-based evidence. Several studies have found that individuals appear to have great difficulty evaluating evidence that is inconsistent with their beliefs. For example, research in a variety of disciplines including cognitive psychology (e.g. Bruner *et al.* 1956; Wason 1968; Mynatt *et al.* 1977; Koriat *et al.* 1980; Klayman & Ha 1987; Evans 1989), judicial reasoning (e.g. Hendry & Shaffer 1989; Pennington & Hastie 1993; Simon 2004; Simon *et al.* 2004) and medical reasoning (e.g. Elstein & Bordage 1979) have all demonstrated examples of biases in evidence-based decision-making. The typical finding is that people are more likely to attend to, seek out and evaluate evidence that is consistent with their beliefs, and ignore or downplay evidence that is inconsistent with their beliefs. In a series of experiments conducted in our laboratory, we have examined how reasoners appear to use their prior knowledge and expectations to *constrain* how they evaluate covariation-based evidence.

In one series of studies we (Fugelsang *et al.* 2004) created a causal thinking situation where participants were asked to test the effectiveness of novel drugs designed to produce a particular outcome. The plausibility of the causal theories was manipulated by presenting participants with a brief introductory statement, which depicted a causal theory that contained either a plausible mechanism of action or an implausible mechanism of action. This manipulation was intended to induce a specific belief about how the potential cause may produce an expected outcome. Evidence was then provided to participants in a trial-by-trial format where they viewed multiple trials of evidence for each type of drug. Under some conditions the candidate cause covaried strongly with an expected outcome; under other conditions the candidate cause covaried weakly with an expected outcome. Here, for example, evidence of a strong covariation would be *consistent* with their beliefs and expectations based on a plausible theory and *inconsistent* with their beliefs and expectations based on an implausible theory.

The basic finding is that people weight the covariation-based evidence stronger when it follows from a theory that contains a plausible mechanism of action than when the evidence follows from a theory that contains an implausible mechanism of action (see Fugelsang & Thompson 2000, 2003; Fugelsang & Dunbar 2004). We have argued that this could be seen as a useful heuristic given the potentially infinite number of covarying causes for every given effect occurring in the natural environment. Using one's beliefs and expectations to *filter* out evidence for implausible theories serves to make the task of building causal theories from evidence achievable. Of course, this heuristic does have a drawback in that potentially valid evidence may be discounted if it is inconsistent with a theory that an individual has strong beliefs in.

4. BRAIN-BASED CORRELATES OF COMPLEX CAUSAL REASONING

An exciting recent approach to the study of human reasoning and decision-making has accompanied the advent of advanced functional brain imaging techniques such as positron emission topography, event-related potentials and fMRI. Using these new techniques, we are able to get a first-hand look at how the brain responds during complex reasoning. Recent work by several cognitive neuroscientists has examined the neurological underpinnings of a variety of complex reasoning and decision-making processes including problem solving (Goel & Grafman 1995; Colvin *et al.* 2001; Fincham *et al.* 2002), analogical reasoning (Wharton *et al.* 2000; Kroger *et al.* 2002; A. Green, J. Fugelsang, N. Shamosh and K. Dunbar, unpublished data), inductive reasoning (Goel & Dolan 2000; Seger *et al.* 2000) and deductive reasoning (Osherson *et al.* 1998; Parsons & Osherson 2001; Goel & Dolan 2003). The major research approach has been to have participants take part in a task that taps a specific reasoning process of interest (e.g. deductive reasoning) and contrast that with a control task that contains much of the same visual and cognitive stimulation but devoid of the specific reasoning process of interest. By contrasting the task-related brain activations of the specific reasoning task with the control task, researchers are able to measure the unique brain activity associated with the specific reasoning process of interest.

We (Fugelsang & Dunbar 2004) have taken a slightly different approach in our research programme on the neural underpinning of complex causal reasoning. Rather than using fMRI to uncover specialized neural circuitry for causal thinking, we have been examining how reasoning with statistical covariation-based evidence that is either consistent or inconsistent with participants' beliefs recruits brain networks that have been implicated in several more domain general cognitive processes. Specifically, we have been using fMRI to uncover the mechanisms by which statistical evidence is integrated with one's beliefs and expectations about that evidence in the brain.

The main question that motivated this research programme was the extent to which people might be more inclined to attend to and assimilate evidence that is consistent with their beliefs, while treating evidence that is inconsistent with their beliefs as error. If that was the case, there are several key brain networks that might be the neural signature of these processes. For example, research in behavioural and cognitive neuroscience has indicated that there are several different brain networks that are invoked during learning (e.g. McDermott *et al.* 1999; Poldrack *et al.* 2002) and in error detection and conflict monitoring (e.g. Botvinick *et al.* 2001; Holroyd & Cole 2002; Kerns *et al.* 2004; Yeung *et al.* 2004) that may be invoked for these different conditions. Based on our prior behavioural research, we predicted that different networks related to learning and conflict monitoring may show increased activity when participants are receiving evidence that is consistent, or inconsistent, respectively, with their beliefs and expectations.

Using a similar paradigm to that of our behavioural experiments, we measured the task-related blood oxygen-level-dependent response as participants observed evidence on the effectiveness of drugs designed to relieve depressive symptoms. The plausibility of the theory of

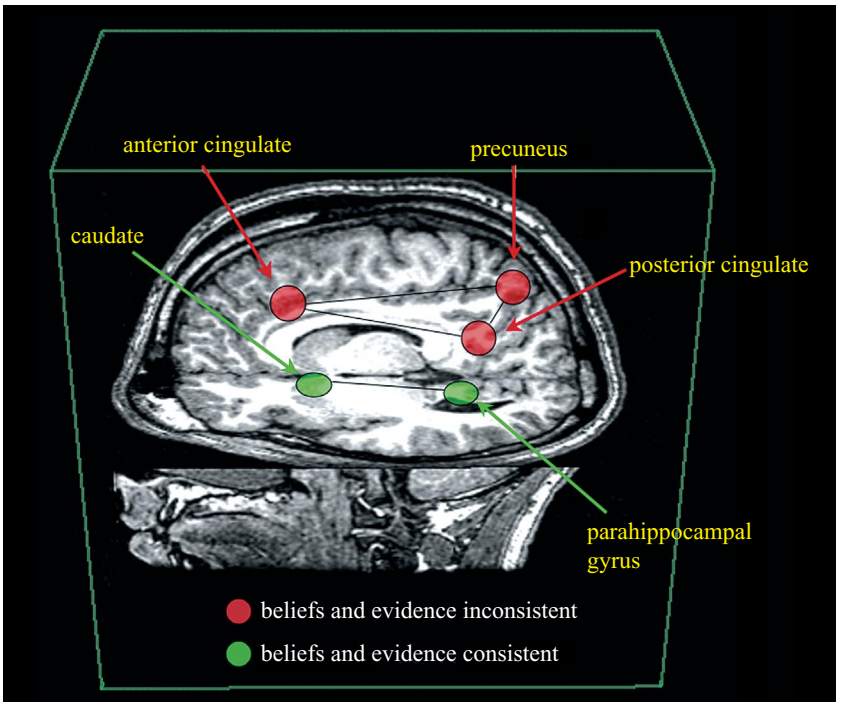

Figure 1. A graphical depiction of a brain-based model displaying the two dissociated networks involved with belief and evidence integration in causal reasoning. Symbols in green depict the observed brain-based network preferentially recruited when people's beliefs and the observed evidence are *consistent*, and symbols in red depict the observed brain-based network preferentially recruited when beliefs and evidence are *inconsistent*.

action of the drug and whether the data were consistent or inconsistent with the theory were varied. We found that when people were reasoning with evidence that was *consistent* with their beliefs, a distinct network of brain regions widely associated with learning and memory were significantly activated, including the caudate and the parahippocampal gyrus. By contrast, when the evidence was *inconsistent* with people's beliefs, a different pattern of activation occurred that is widely associated with error detection and conflict resolution, including the anterior cingulate cortex, posterior cingulate and the precuneus. A graphical depiction of this brain-based model of these findings is depicted in figure 1 where symbols in green depict the observed brain network activated when beliefs and evidence are *consistent*, and symbols in red depict the observed brain network activated when beliefs and evidence are *inconsistent*. These findings provide a neural instantiation for the behavioural interactions between beliefs and evidence that we covered in § 3. Specifically, people's beliefs and expectations may act as a *biological filter* during evidence evaluation by selectively recruiting learning mechanisms for evidence that is consistent with their beliefs and error detection mechanisms for evidence that is inconsistent with their beliefs.

5. FURTHER IMPLICATIONS FOR LEGAL DECISION-MAKING

In this essay, we have covered a variety of experimental studies that demonstrate how people's evaluation of evidence is highly influenced by their beliefs about the objects and events under consideration. That is, people typically *do not* evaluate evidence in an atheoretical manner; rather they use their beliefs and expectations to guide their assessment of the evidence given to them (e.g. Evans *et al.* 1983; Fugelsang *et al.* 2004). This interplay between one's beliefs and evidence has a distinct neural signature, in that evidence that is consistent with one's beliefs is more likely to recruit neural tissue involved in learning and memory, whereas evidence that is inconsistent with one's beliefs is more likely to invoke neural tissue associated with error detection and conflict monitoring.

These findings have important implications for legal reasoning. Specifically, the human brain appears to be specifically sensitive to the degree to which evidence, in the form of statistical information, is consistent with the expectations of the individual. Is this something that the brain does automatically, or do people have conscious cognitive control over this process? Indeed, the answer to this question may have large implications for how one can hope to minimize the influence of such biases when they hinder legal decision-making. The finding that these brain-based dissociations occur when evidence is first being evaluated suggests that people may be unable to set aside their beliefs and expectation when making judgements about causality. These data are corroborated by some recent findings by Fugelsang & Thompson (2003). They provided data that showed that individuals were unable to gauge the degree to which their beliefs and expectations influenced their evaluation of statistical evidence. Taken together, these findings suggest that individuals may not be entirely aware of the extent to which their expectations influence their decisions when judging evidence.

Judges and lawyers are very aware that people can be biased in their reasoning. Our results indicate that there may be brain-based underpinnings for these biases. Knowing that one's beliefs can hinder reasoning, and that this may occur at an automatic or unconscious level, what

additional strides can be taken to minimize the influence of one's beliefs when they may hinder legal reasoning? Several researchers have found that instructions developed to augment the relevance of normative information can reduce the impact that one's beliefs have on their reasoning (Evans 2002; Evans *et al.* 1994). Specifically, Evans and colleagues have found that instructions that emphasize the role of logical form of a problem can result in a decrease in the influence of the content of the scenarios. In addition, Dunbar (1993) found that altering the goal of a reasoner could be very effective in switching a reasoner's strategy from one of confirmation seeking to disconfirmation seeking.

An important avenue for future research would be to examine the role of making individuals cognizant of the potential biasing effects of their beliefs before they engage in legal reasoning. Laying out potential alternative hypotheses before the presentation of evidence may minimize the influence of specific beliefs on the part of the individual asked to weigh the evidence. In addition, the biasing effects of beliefs on evidential evaluation and subsequent judgements are surely influenced by several factors related to the acquisition and maintenance of those beliefs. For example, beliefs may vary in terms of several variables, such as personal relevance, age of initial acquisition and original source of acquisition, that may be orthogonal to the strength of those beliefs. In addition, the extent to which such extra-legal factors (e.g. belief biases in decision-making) influence different content-oriented factors of the judicial process (e.g. judges' instructions, examination of witnesses and the use of exhibits) is still relatively unknown. Further cross-talk between cognitive neuroscientists and legal academics and professionals will do much to inform these future research endeavours.

The authors thank the Honorable Judge Morris Hoffman, Professor Kenneth Kreiling and Professor Oliver Goodenough for providing discussions on various aspects of the research contained in our essay. Research reported in this essay has been funded by grants from Dartmouth College, and the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

Ahn, W., Kalish, C. W., Medin, D. L. & Gelman, S. A. 1995 The role of covariation versus mechanism information in causal attribution. *Cognition* **54**, 299–352.

Allan, L. G. & Jenkins, H. M. 1980 The judgements of contingency and the nature of response alternatives. *Can. J. Psychol.* **34**, 1–11.

Botvinick, M., Braver, T., Barch, D., Carter, C. & Cohen, J. 2001 Conflict monitoring and cognitive control. *Psychol. Rev.* **108**, 625–652.

Bruner, J. S., Goodnow, J. J. & Austin, G. A. 1956 *A study of thinking*. New York: NY Science Editions.

Cheng, P. W. 1997 From covariation to causation: a causal power theory. *Psychol. Rev.* **104**, 367–405.

Cheng, P. W. & Novick, L. R. 1990 A probabilistic contrast model of causal induction. *J. Pers. Soc. Psychol.* **58**, 545–567.

Cheng, P. W. & Novick, L. R. 1992 Covariation in natural causal induction. *Psychol. Rev.* **99**, 365–382.

Colvin, M. K., Dunbar, K. & Grafman, J. 2001 The effects of frontal lobe lesions on goal achievement in the water jug task. *J. Cogn. Neurosci.* **13**, 1129–1147.

Constanzo, S. & Constanzo, M. 1994 Life or death decisions: an analysis of capital jury decision-making under the special issues sentencing framework. *Law Hum. Behav.* **18**, 151–170.

Dunbar, K. 1993 Concept discovery in a scientific domain. *Cogn. Sci.* **17**, 397–434.

Dunbar, K. 1995 How scientists really reason: scientific reasoning in real-world laboratories. In *Mechanisms of insight* (ed. R. J. Sternberg & J. Davidson), pp. 365–395. Cambridge, MA: MIT Press.

Dunbar, K. 2002 Science as category: implications of *in vivo* science for theories of cognitive development, scientific discovery, and the nature of science. In *Cognitive models of science* (ed. P. Caruthers, S. Stich & M. Siegel), pp. 154–170. New York: Cambridge University Press.

Einhorn, H. J. & Hogarth, R. M. 1986 Judging probable cause. *Psychol. Bull.* **99**, 3–19.

Elstein, A. S. & Bordage, G. 1979 Psychology of clinical reasoning. In *Health psychology: a handbook* (ed. G. Stone, F. Cohen & N. Adler), pp. 333–367. San Francisco, CA: Jossey-Bass.

Evans, J. St B. T. 1989 *Bias in human reasoning: causes and consequences*. Hove: Psychological Press.

Evans, J. St B. T. 2002 Logic and human reasoning: an assessment of the deduction paradigm. *Psychol. Bull.* **128**, 978–996.

Evans, J. St B. T., Barston, J. L. & Pollard, P. 1983 On the conflict between logic and belief in syllogistic reasoning. *Mem. Cogn.* **11**, 295–306.

Evans, J. St B. T., Newstead, S. E., Allen, J. L. & Pollard, P. 1994 Debiasing by instruction: the case of belief bias. *Eur. J. Cogn. Psychol.* **6**, 263–285.

Fincham, J. M., Carter, C. S., van Veen, V., Stenger, V. A. & Anderson, J. R. 2002 Neural mechanisms of planning: a computational analysis using event-related fMRI. *Proc. Natl. Acad. Sci. USA* **99**, 3346–3351.

Fugelsang, J. A. & Dunbar, K. 2004 Brain-based mechanisms underlying complex causal thinking. *Neuropsychologia*. (In the press.)

Fugelsang, J. & Thompson, V. 2000 Strategy selection in causal reasoning: when beliefs and covariation collide. *Can. J. Exp. Psychol.* **54**, 13–32.

Fugelsang, J. A. & Thompson, V. A. 2003 A dual-process model of belief and evidence interactions in causal reasoning. *Mem. Cogn.* **31**, 800–815.

Fugelsang, J. A., Stein, C., Green, A. & Dunbar, K. 2004 Theory and data interactions of the scientific mind: evidence from the molecular and the cognitive laboratory. *Can. J. Exp. Psychol.* **58**, 132–141.

Goel, V. & Dolan, R. J. 2000 Anatomical segregation of component processes in an inductive inference task. *J. Cogn. Neurosci.* **12**, 110–119.

Goel, V. & Dolan, R. L. 2003 Explaining modulation of reasoning by belief. *Cognition* **87**, 11–22.

Goel, V. & Grafman, J. 1995 Are the frontal lobes implicated in 'planning' functions? Interpreting data from the Tower of Hanoi. *Neuropsychologia* **33**, 623–642.

Harre, R. & Madden, E. H. 1975 *Causal powers: a theory of natural necessity*. Oxford: Basil Blackwell.

Hart, H. & Honore, A. 1959 *Causation in law*. Oxford: Clarendon.

Hendry, S. H. & Shaffer, D. R. 1989 On testifying in one's own behalf: interactive effects of evidential strength and defendant's testimonial demeanor on jurors' decisions. *J. Appl. Psychol.* **74**, 539–545.

Hilton, D. & Slugoski, B. 1986 Knowledge-based causal attribution: the abnormal conditions focus model. *Psychol. Rev.* **93**, 75–88.

Holroyd, C. B. & Cole, M. G. H. 2002 The neural basis of human error processing: reinforcement learning, dopamine, and error-related negativity. *Psychol. Rev.* **109**, 679–709.

Hume, D. 1978 *A treatise of human nature*. Oxford University Press. [Original work published 1739.]

Jenkins, H. & Ward, W. 1965 Judgments of contingency between response and outcomes. *Psychol. Monogr.* **7**, 1–17.

Kant, I. 1965 *Critique of pure reason*. London: Macmillan & Co. [Original work published 1781.]

Kelley, H. 1973 The process of causal attribution. *Am. Psychol.* **28**, 107–128.

Kerns, J. G., Cohen, J. D., MacDonald, A. W., Cho, R. Y., Stenger, A. & Carter, C. S. 2004 Anterior cingulate conflict monitoring and adjustment in control. *Science* **303**, 1023–1026.

Klayman, J. & Ha, Y. 1987 Confirmation, disconfirmation, and information in hypothesis testing. *Psychol. Rev.* **94**, 211–228.

Koriat, A., Lichtenstein, S. & Fischhoff, B. 1980 Reasons for confidence. *J. Exp. Psychol. Learn.* **6**, 107–118.

Kroger, J. K., Sabb, F. W., Fales, C. L., Bookheimer, S. Y., Cohen, M. S. & Holyoak, K. J. 2002 Recruitment of anterior dorsolateral prefrontal cortex in human reasoning: a parametric study of relational complexity. *Cerebr. Cortex* **12**, 477–485.

McDermott, K. B., Ojemann, J. G., Petersen, S. E., Ollinger, J. M., Snyder, A. Z., Akbudak, E., Conturo, T. E. & Raichle, M. E. 1999 Direct comparison of episodic encoding and retrieval of words: an event-related fMRI study. *Memory* **7**, 661–678.

Matlon, R. J. 1986 Factors affecting jury decision-making. *Soc. Action Law* **12**, No 41- 42.

Mynatt, B. T., Doherty, M. E. & Tweney, R. D. 1977 Confirmation bias in a simulated research environment: an experimental study of scientific inference. *Q. J. Exp. Psychol.* **29**, 85–95.

Newton, I. 1999 *The principia: mathematical principles of natural philosophy*. Berkley, CA: University of California Press. [Original work published 1678.]

Novick, L. & Cheng, P. 2004 Assessing interactive causal inference. *Psychol. Rev.* **111**, 455–485.

Osherson, D., Perani, D., Cappa, S., Schnur, T., Grassi, F. & Fazio, F. 1998 Distinct brain loci in deductive versus probabilistic reasoning. *Neuropsychologia* **36**, 369–376.

Parsons, L. M. & Osherson, D. 2001 New evidence for distinct right and left brain systems for deductive versus probabilistic reasoning. *Cerebr. Cortex* **11**, 954–965.

Pennington, N. & Hastie, R. 1990 Practical implications of psychological research on juror and juror decision-making. *Pers. Soc. Psychol. Bull.* **16**, 90–105.

Pennington, N. & Hastie, R. 1993 The story model of juror decision-making. In *Inside the juror: the psychology of jury decision-making* (ed. R. Hastie), pp. 192–221. New York: Cambridge University Press.

Poldrack, R. A., Clark, J., Pare-Blagoev, E. J., Shohamy, D., Creso Moyano, J., Myers, C. & Gluck, M. A. 2002 Interactive memory systems in the human brain. *Nature* **414**, 546–550.

Rescorla, R. A. 1968 Probability of shock in the presence and absence of CS in fear conditioning. *J. Comp. Physiol. Psychol.* **66**, 1–5.

Seger, C., Poldrack, R., Prabhakaran, V., Zhao, M., Glover, G. & Gabrieli, J. 2000 Hemispheric asymmetries and individual differences in visual concept learning as measured by functional MRI. *Neuropsychologia* **38**, 1316–1324.

Simon, D. 2004 A third view of the black box: coherence based reasoning in law. *Univ. Chicago Law Rev.* **71**, 511–586.

Simon, D., Snow, C. & Read, S. 2004 The redux of cognitive consistency theories: evidence judgements by constraint satisfaction. *J. Pers. Soc. Psychol.* **86**, 814–837.

Wason, P. C. 1968 Reasoning about a rule. *Q. J. Exp. Psychol.* **20**, 273–281.

Wharton, C. M., Grafman, J., Flitman, S. S., Hansen, E. K., Brauner, J., Marks, A. & Honda, M. 2000 Toward neuroanatomical models of analogy: a positron emission tomography study of analogical mapping. *Cogn. Psychol.* **40**, 173–197.

White, P. A. 1989 A theory of causal processing. *Br. J. Psychol.* **80**, 431–454.

White, P. A. 1995 Use of prior beliefs in the assignment of causal roles: causal powers versus covariation-based accounts. *Mem. Cogn.* **23**, 243–254.

White, P. A. 2002 Perceiving a strong causal relation in a weak contingency: further investigation of the evidential evaluation model of causal judgement. *Q. J. Exp. Psychol. A* **55**, 97–114.

Wiener, R. L., Pritchard, C. C. & Weston, M. 1995 Comprehensibility of approved jury instructions in capital murder cases. *J. Appl. Psychol.* **80**, 455–467.

Yeung, N., Botvinick, M. M. & Cohen, J. D. 2004 The neural basis of error-detection: conflict monitoring and the error-related negativity. *Psychol. Rev.* **111**, 931–959.

GLOSSARY

fMRI: functional magnetic resonance imaging