ELSEVIER

Neuropsychologia 43 (2005) 1204–1213

NEUROPSYCHOLOGIA

www.elsevier.com/locate/neuropsychologia

Brain-based mechanisms underlying complex causal thinking

Jonathan A. Fugelsang, Kevin N. Dunbar*

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

Received 12 March 2004; received in revised form 15 October 2004; accepted 21 October 2004

Abstract

We use functional magnetic resonance imaging (fMRI) and behavioral analyses to study the neural roots of biases in causal reasoning. Fourteen participants were given a task requiring them to interpret data relative to plausible and implausible causal theories. Encountering covariation-based data during the evaluation of a plausible theory as opposed to an implausible theory selectively recruited neural tissue in the prefrontal and occipital cortices. In addition, the plausibility of a causal theory modulated the recruitment of distinct neural tissue depending on the extent to which the data were *consistent* versus *inconsistent* with the theory provided. Specifically, evaluation of data *consistent* with a plausible causal theory recruited neural tissue in the parahippocampal gyrus, whereas evaluating data *inconsistent* with a plausible theory recruited neural tissue in the anterior cingulate, left dorsolateral prefrontal cortex, and precuneus. We suggest that these findings provide a neural instantiation of the mechanisms by which working hypotheses and evidence are integrated in the brain.

Keywords: Scientific reasoning; Learning; Error detection; Conflict monitoring; fMRI

1. Introduction

For over 400 years scientists (Crick, 1990), philosophers (Bacon, 1620/1854), cognitive psychologists (Dunbar, 2002), and even politicians (Healy, 1996) have debated the preferred way for people to think and reason about data. In the cognitive laboratory, decades of research have clearly established that one's knowledge influences how people interpret data in their environment. These findings have come from a variety of theoretical traditions including the investigation of heuristics and biases in decision-making (e.g., Gigerenzer & Goldstein, 1996; Kahneman & Tversky, 1996; Todd & Gigerenzer, 2000; Tversky & Kahneman, 1974), belief-bias effects in deductive reasoning (e.g., Evans, 1989; Evans, Barston, & Pollard, 1983; Goel & Dolan, 2003; Klauer, Musch, & Naumer, 2000), and knowledge mediation in causal and scientific reasoning (Fugelsang & Thompson, 2000, 2001, 2003; Fugelsang, Stein, Green, & Dunbar, 2004; Klahr, Fay, & Dunbar, 1993; Koehler, 1993). A common thread through these approaches is that the knowledge people possess changes how they evaluate information provided to them. Specifically, the knowledge individuals bring to bear on a task has been shown to greatly influence their tendency to carry out that task in a way traditionally deemed as normatively appropriate.

A prevalent form of human inference where knowledge modulates the analyses of data is *causal reasoning*. Here, the reasoner must ascertain the extent to which variables are causally related based on one or more causal cues (e.g., covariation, mechanism, temporal and spatial contiguity). Recent work conducted in our laboratory has shown that the degree to which data are evaluated is modulated by the plausibility of the causal theory being tested (Fugelsang & Thompson, 2000, 2003; Fugelsang et al., 2004). Specifically, we have shown that the plausibility of a causal theory guides the analyses of data such that reasoners may be more inclined to assess data that are encountered during the evaluation of a plausible theory as opposed to data encountered during the evaluation of an implausible theory.

By what mechanism does this knowledge mediation occur? Recent cognitive models have converged on the notion that attentional processes mediate much of theory and data interactions in a number of reasoning domains (e.g.,

* Corresponding author. Tel.: +1 603 646 2998; fax: +1 603 646 1419.
E-mail address: kevin.n.dunbar@dartmouth.edu (K.N. Dunbar).

0028-3932/$ – see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neuropsychologia.2004.10.012

Evans, 2003; Fugelsang & Thompson, 2003; Gigerenzer & Goldstein, 1996; Kahneman & Tversky, 1996; Klauer, Musch, & Naumer, 2000). However, the locus of such effects has remained relatively elusive. There are at least two possible ways in which attentional processes can mediate the interplay between theory and data in the domain of causal reasoning. These hypotheses concern the extent to which one's attention, and subsequent working memory processes are drawn to data encountered during the evaluation of plausible versus implausible theories. One possibility is that reasoners quickly accept with little deliberation data encountered while evaluating a plausible theory and closely scrutinize data encountered while evaluating an implausible theory. Conversely, reasoners may preferentially attend to data encountered while evaluating a plausible theory and ignore data encountered while evaluating an implausible theory. These hypotheses can be dissociated by examining the extent to which brain networks typically associated with attention, working memory, and executive processes, such as the prefrontal cortex (e.g., Curtis & D'Esposito, 2003; Smith & Jonides, 1999), are selectively recruited when encountering data during the evaluation of plausible versus implausible theories.

A second component of these hypotheses concern the mechanisms by which data *consistency* interacts with the plausibility of the causal theory being tested. Are people more inclined to attend to, associate, and integrate data *consistent* with a theory while treating data *inconsistent* with a theory as erroneous? Research in behavioral and cognitive neuroscience indicates that there are a number of key brain networks that are invoked during learning versus error detection and conflict monitoring that may provide a neural basis for operationalizing such biases in causal reasoning.

Concerning the former, both patient studies (e.g., Bernasconi et al., 2003; Damasio, Eslinger, Damasio, Van Hoesen, & Cornell, 1985; Hay, Moscovitch, & Levine, 2002; Milner, Corkin, & Teuber, 1968) and functional imaging studies (e.g., Kapur et al., 1996; Kelley et al., 1998; McDermott et al., 1999; Poldrack et al., 2002; Ranganath et al., 2003) have highlighted the primary role of the parahippocampal gyrus and related mesial structures in declarative learning and memory. Specifically, the parahippocampal gyrus and adjacent structures in the temporal lobes are thought to be crucial for binding stimulus features into an episodic memory trace (Moscovitch, 1992; Wagner, Maril, & Schacter, 2000) thus allowing successful subsequent retrieval of information retrospectively. Concerning the latter, numerous ERP and fMRI studies using a variety of tasks including variants of the Stroop task (e.g., Bush et al., 1998; Kerns et al., 2004), the Eriksen Flanker task (e.g., Fan, Flombaum, McCandliss, Thomas, & Posner, 2003; van Veen, Cohen, Botvinick, Stenger, & Carter, 2001), and probabilistic learning paradigms (e.g., Holroyd et al., 2004) have highlighted the predominant role of the anterior cingulate cortex in error detection and conflict monitoring. Indeed, a number of key theoretical papers have recently been devoted to understanding the central role of the anterior cingulate cortex in error detection and conflict monitoring (e.g., Botvinick, Braver, Barch, Carter, & Cohen, 2001; Bush, Luu, & Posner, 2000; Holroyd & Coles, 2002; van Veen & Carter, 2002; Yeung, Botvinick, & Cohen, 2004).

Based on these prior findings, and the hypothesized mechanisms of reasoning described above, we predict that the disparate networks associated with learning versus conflict monitoring will show increased activity when participants evaluate data that are *consistent* versus *inconsistent*, respectively, with the theory provided to them. To address these issues, we developed a causal reasoning task where the strength of statistical data is manipulated orthogonally to the plausibility of the theory being tested. To do this, we adapted a methodology commonly used in the cognitive laboratory to measure causal reasoning processes based on the strength of covariation-based statistical data. This methodology takes into account the combined role of the *sufficiency* and *necessity* of observed statistical relationships. The sufficiency of a cause is determined by the probability that the effect occurs in the presence of a cause [i.e., $P(e/c)$], whereas the necessity of a cause is determined by the probability that the effect occurs in the absence of a cause [i.e., $P(e/\sim c)$]. Using these two components, the covariation between a potential cause and outcome can be determined by subtracting the latter equation from the former [i.e., $P(e/c) - P(e/\sim c)$]. This metric of covariation, commonly referred to as the ΔP_c *coefficient*, is featured prominently in contemporary theories of causal thinking (e.g., Cheng, 1997; Cheng & Novick, 1990; Novick & Cheng, 2004; White, 2002) and numerous experiments conducted in the cognitive laboratory support the assumption that people do indeed make causal inferences to a large degree based on the observed covariation between variables (e.g., Allan & Jenkins, 1980; Fugelsang & Thompson, 2000, 2001, 2003; Fugelsang et al., 2004; Spellman, 1996; White, 2002). We were predominantly interested in examining (1) the degree to which theory plausibility biases the evaluation of statistical covariation-based data, and (2) the neural foundations that subserve these biases.

2. Method

2.1. Participants

Fourteen participants (6 males, 8 females; age range 18–31 years) took part in the study and were paid $10. All participants were right-handed, reported no significant abnormal neurological history and had normal or corrected-to-normal visual acuity. Informed written consent for all participants was obtained prior to the experiment in accordance with the guidelines established by the Committee for the Protection of Human Subjects at Dartmouth College.

Table 1
Mean (standard deviations in parentheses) pre-tested plausibility and familiarity ratings of the causal theories

Causal theory	Plausibility rating (0–10), $N = 23$	Familiarity rating (1–7), $N = 13$
Plausible theories		
Serotonin reuptake inhibitor	8.17 (1.07)	2.69 (1.75)
Monoamine oxidase inhibitor	8.22 (0.95)	2.31 (2.18)
Implausible theories		
Protein binder	0.78 (0.80)	2.31 (2.32)
Topoisomerase inhibitor	1.17 (1.54)	2.08 (1.65)

2.2. Design and apparatus

A standard block design was used with 50 s of task followed by 30 s of fixation only rest trials. Visual stimuli were presented using a G4 PowerBook computer running PsyScope 2.5.1 software (Cohen, MacWhinney, Flatt, & Provost, 1993). Stimuli were projected to participants using an Epson (model ELP-7000) LCD projector onto a screen positioned at the head end of the fMRI scanner bore. Participants viewed the screen through a mirror. Cushions were used to minimize head movement.

2.3. Stimuli and task

Using fMRI, we measured the task related blood oxygen level dependent (BOLD) response as participants observed covariation-based data on the effectiveness of drugs designed to relieve depressive symptoms. The plausibility of a theory was manipulated by presenting participants with a brief introductory statement that contained either (1) a direct causal mechanism of action linking a red pill to a mood outcome, or (2) no direct causal mechanism of action linking a red pill to a mood outcome (see Appendix A). Here, we define theory plausibility in terms of the degree to which a mechanism of action exists that links the candidate cause to the effect under consideration (see Ahn, Kalish, Medin, & Gelman, 1995; Harre & Madden, 1975; White, 1989). The causal mechanisms consisted of biological agents that were equated for complexity. Table 1 presents the mean pre-rated plausibility and familiarity ratings for these stimuli obtained from an independent sample of Dartmouth College undergraduate students who did not receive the covariation-based data manipulation. Participants were given no explicit causal mechanism information for the blue pill and were instructed to treat it as a placebo condition.

Data were then provided to participants in a trial-by-trial format where they viewed 20 trials of data each lasting 2.5 s for each of the four causal theories provided. These data were presented in combinations of the cause (a *red pill* or a *blue pill*) and the effect (*happiness* or *neutral* outcome) co-occurring. Fig. 1 presents a graphical depiction of these four event types. Under some conditions the *red pill* and *happiness* covaried strongly, under other conditions the *red pill* and *happiness* covaried weakly. This was accomplished by varying the frequency with which each of the four event types (*red pill*/*happiness*, *red pill*/*neutral*, *blue pill*/*happiness*, *blue*

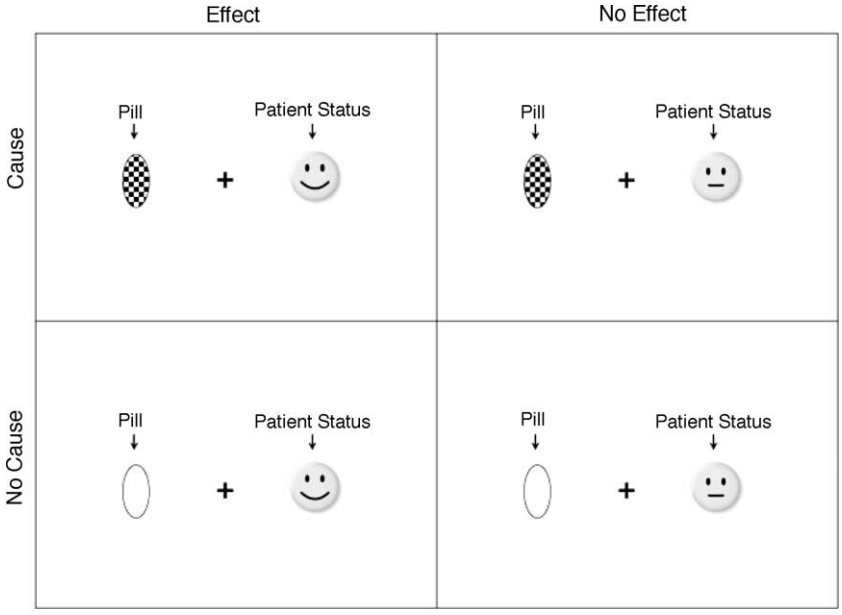

Fig. 1. Example stimuli representing the four possible combinations of the candidate cause (checked pill vs. white pill) and effect (happiness vs. neutral emotion). Note that the stimuli in the actual experiments utilized a red pill and a blue pill in the place of the checked pill and the white pill.

Table 2
Event frequencies used for the computation of covariation-based data strength

Event frequencies				Degree of covariation		
ce	c∼e	∼ce	∼c∼e	*P*(e/c)	*P*(e∼c)	Covariation (ΔP_c)
18	2	4	16	18/20	4/20	0.7 (strong)
10	10	4	16	10/20	4/20	0.3 (weak)

Note: (ce) represents the number of times the cause and effect co-occurred; (c∼e) represents the number of times the cause occurred in the absence of the effect; (∼ce) represents the number of times the effect occurred in the absence of the cause; (∼c∼e) represents the number of times the effect was absent when the cause was absent.

pill/*neutral*) occurred. Table 2 presents the event frequencies used to manipulate covariation information on a trial-by-trial basis. Note that the strong covariation- and weak covariation-based data conditions represented an actual covariation of 0.7 and 0.3, respectively, as measured by the ΔP_c *coefficient*. Note also that high covariation-based data encountered during the evaluation of a plausible causal theory and low covariation-based data encountered during the evaluation of an implausible causal theory would both constitute *consistent* data, whereas low covariation-based data encountered during the evaluation of a plausible theory and high covariation-based data encountered during the evaluation of an implausible theory would both constitute *inconsistent* data. After participants received 20 trials of data, they were asked to judge the effectiveness of the red pill in causing the happiness using a scale that ranged from 1 (low) to 3 (high). This procedure was repeated four times: once for each level of the theory plausibility and covariation manipulations. Therefore, each participant took part in all conditions using a completely within subjects design.

2.4. Image acquisition

Imaging was performed on a 1.5 T whole body scanner (General Electric Medical Systems Signa, Milwaukee, Wisconsin) with a standard head coil. Anatomical images were acquired using a high-resolution 3D spoiled gradient recovery sequence (SPGR; 124 sagittal slices, TE = 6 ms, TR = 25 ms, flip angle = 25°, voxel size = 1 mm × 1 mm × 1.2 mm). Functional images were collected in runs using a gradient spin-echo echo-planar sequence sensitive to blood oxygen level-dependent (BOLD) contrast ($T2^*$) (TR = 2500 ms, $T2^*$ evolution time = 35 ms, flip angle = 90°, 3.75 mm × 3.75 mm in-plane resolution). During each functional run, 40 sets of axial images (25 slices; 5.5-mm slice thickness, 1 mm skip between slices) were acquired allowing complete brain coverage.

2.5. Statistical analysis

All data were analyzed using SPM99 software (Wellcome Department of Cognitive Neurology, London, UK; Friston et al., 1995). For each functional run, data were preprocessed to remove sources of noise and artifact. Functional data were realigned within and across runs to correct for head movement using a six parameter, rigid body alignment technique (Kiebel, Ashburner, Poline, & Friston, 1997; Woods, Grafton, Holmes, Cherry, & Mazziotta, 1998) and coregistered with each participant's anatomical data. Functional data were then transformed into a standard anatomical space (3 mm isotopic voxels) based on the ICBM 152 brain template (Montreal Neurological Institute), which approximates Talairach and Tournoux (1988) atlas space using higher order polynomial then non-linear basis functions (Ashburner & Friston, 1999). Normalized data were then spatially smoothed (10 mm full-width-at-half-maximum) using a Gaussian kernel in order to optimize signal-to-noise (Skudlarski, Constable, & Gore, 1999) and abide by the assumptions of Gaussian random field theory (Worsley & Friston, 1995). The normalized and smoothed images were then used for the subsequent statistical analysis. For each subject, a general linear model (Friston et al., 1998) incorporating task effects (modelled as a box-car function convolved with the canonical hemodynamic response function), a mean, and a linear trend were used to compute parameter estimates (β) and *t*-contrast images (containing weighted-parameter estimates) for each comparison at each voxel. A random-effects analysis (Friston, Holmes, Price, Buchel, & Worsley, 1999; Lazar, Luna, Sweeney, & Eddy, 2002) consisting of one-sample *t*-tests with a hypothesized mean of 0 was then applied to the individual subject t-contrast images to create mean *t*-images (thresholded at $P = .001$, uncorrected).

3. Results

The results are presented in two sections. The first section presents the omnibus analyses of theory plausibility (implausible versus plausible), and the strength of the covariation-based data (strong versus weak) for the behavioral judgments. The second section presents the fMRI random-effects group analyses. Effect size estimates in the behavioral results section were computed using *partial* η^2.

3.1. Behavioral results

Fig. 2 presents the mean effectiveness ratings for the two theory types for both strong and weak covariation-based data. These data reveal that the participants' causal judgments were influenced by both the plausibility of the theory, $F(1,13) = 5.2$, M.S.E. = 0.495, $\eta^2 = 0.29$, $P < .05$, and the covariation between the occurrence of the red pill and the outcome, $F(1,13) = 81.37$, M.S.E. = 0.148, $\eta^2 = 0.86$, $P < .01$. Importantly, there is also a significant interaction between theory plausibility and covariation, $F(1,13) = 10.48$, M.S.E. = 0.170, $\eta^2 = 0.45$, $P < .01$ revealing that the covariation manipulation has a greater effect for plausible theories (*mean difference* = 1.29) than implausible theories (*mean difference* = 0.57). Consistent with prior behavioral work (i.e.,

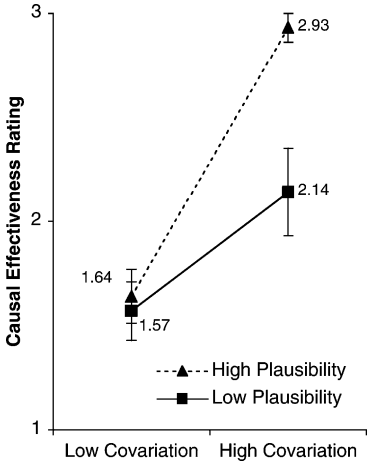

Fig. 2. Mean causal effectiveness ratings for the two theory types (low plausibility vs. high plausibility) for both weak data (low covariation) and strong data (high covariation) after the 20 presentation trials.

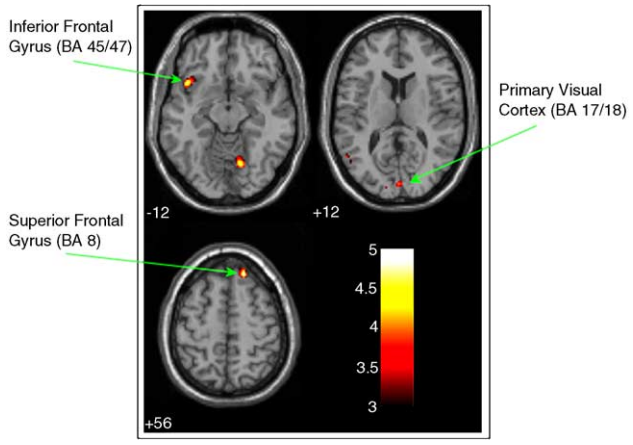

Fig. 3. Unique task associated BOLD activations occurring when participants encountered data while evaluating a plausible vs. an implausible theory.

Fugelsang & Thompson, 2000, 2003; Fugelsang et al., 2004), these data reveal a *belief-bias* in causal reasoning whereby the effects of covariation are larger when evaluating a plausible as opposed to an implausible causal theory.

3.2. fMRI results

We analyzed the task related BOLD response for conditions in which subjects encountered data while evaluating a plausible versus an implausible theory. Fig. 3 shows that regions typically associated with working memory and executive processing (Curtis & D'Esposito, 2003), including bilateral prefrontal regions (right superior frontal gyrus [BA 9] and the left inferior frontal gyrus [BA 45/47]) are significantly ($P < .001_{\text{uncorrected}}$) more activated when subjects encountered data during the evaluation of a plausible theory as opposed to an implausible theory. In addition, encountering data during the evaluation of a plausible theory preferentially recruits neural tissue in the primary visual cortex (BA 17/18). These latter findings are consistent with recent work establishing the relationship between visual attention and working memory and the subsequent recruitment of neural tissue in primary and secondary regions of the visual cortex (Rees & Lavie, 2001; Rees, Frith, & Lavie, 1997).

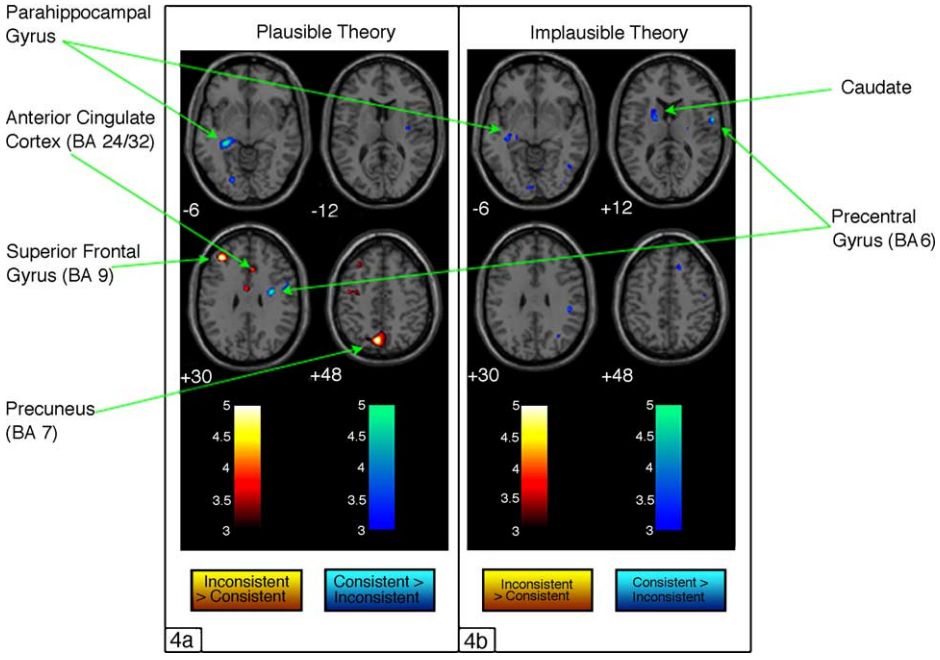

Fig. 4. Unique task associated BOLD activations occurring when viewing data *inconsistent* vs. *consistent* with a plausible theory (a) and an implausible theory (b). Note that the activations denoted by red to yellow are for the conditions in which the provided theory and data are *inconsistent* and the activations denoted by blue to green are for the conditions in which the theory and data are *consistent*.

To assess the degree to which plausibility modulates the integration of data as a function of theory and data *consistency*, we examined the task related BOLD function for conditions in which theory and data are *consistent* (i.e., plausible theory and strong data; implausible theory and weak data) versus conditions in which theory and data are *inconsistent* (i.e., plausible theory and weak data; implausible theory and strong data). Fig. 4 shows that when theory and data are *consistent*, a distinct network of brain regions widely associated with learning and memory (Kelley et al., 1998; McDermott et al., 1999) are preferentially recruited, including the caudate, and the parahippocampal gyrus. In contrast, when theory and data are *inconsistent*, a different pattern of activation occurs that is widely associated with error detection and conflict monitoring (Botvinick et al., 2001; Yeung et al., 2004; Holroyd & Coles, 2002), including the left dorsolateral prefrontal cortex (BA 9), dorsal regions of the anterior cingulate cortex (BA 24/32), and the precuneus (BA 7). Importantly, this latter brain network is only significantly activated when participants encounter data that conflicts with a plausible causal theory.

4. Discussion

In the present experiment, we show that people display specific behavioral and neural response patterns as a function of the relationship between theory and data. Theory and data have an interactive effect on participants' causal judgments whereby data are weighted more heavily when they are encountered during the evaluation of plausible as opposed to implausible causal theories. These data are consistent with recent models of scientific causal thinking and hypothesis testing that incorporate theory and data interactions (Dunbar, 1993; Fugelsang & Thompson, 2000, 2003; Klahr et al., 1993; Klayman & Ha, 1987; Koehler, 1993). Here, it is proposed that using one's knowledge to constrain the use of statistical data is an adaptive strategy. Specifically, given the potentially infinite number of covarying causes for every naturally occurring effect, it is preferable to focus one's attention on data encountered during the evaluation of plausible as opposed to implausible hypotheses. In this way, using one's knowledge to *filter* out data for implausible theories serves to make the task of evaluating causal hypotheses from statistical data feasible.

By contrasting the selective activations associated with encountering data while evaluating plausible versus implausible theories, we were able to dissociate the extent to which one's attention, and subsequent working memory processes, are drawn to data as a function of theory plausibility. Concurrent with the behavioral response patterns, we show that the brain responds differently to incoming data as a function of the plausibility of the theory being tested. Specifically, the preferential recruitment of prefrontal and occipital cortices for conditions in which data are encountered during the evaluation of plausible theories directly correspond to those conditions in which plausibility modulated the greater use of the covariation-based data in participants' behavioral judgments. These patterns of activation suggest that the individuals in the current study may have preferentially devoted more attentional/working memory resources when encountering data during the evaluation of plausible as opposed to implausible theories. These findings are consistent with extensive research demonstrating that the prefrontal cortex is involved in a vast array of tasks that require the active encoding and maintenance of patterns of stimuli (Smith & Jonides, 1999). In addition, and perhaps most relevant for the current experiment, the prefrontal cortex has also been linked to the initiation of *bias signals* to other structures in the brain (Miller & Cohen, 2001). These bias signals from the prefrontal cortex are proposed to guide the flow of activity along specific neural pathways in order to establish the proper mappings between inputs and outputs needed to perform a specific task. Considering the observed *belief-bias* in the behavioral judgments, one can envision such a role for the prefrontal cortex in the current experiment.

Furthermore, the plausibility of a theory influenced the degree to which data *consistent* versus *inconsistent* with that theory invoked disparate neural tissue associated with learning or conflict monitoring. Specifically, the selective activations of the caudate and parahippocampal gyrus under conditions in which theory and data were *consistent* imply that participants may be more apt to efficiently encode data under such conditions. Unexpectedly, in both cases in which theory and data were *consistent*, the precentral gyrus was preferentially recruited in concert with the parahippocampal gyrus. One possible explanation for this finding reflects the extent to which preparatory motor functions might be occurring during the data accumulation phase of the task. Specifically, participants in the current task were required to withhold their causal response until after the 20 trials of data had been presented. The fact that regions typically associated with motor functions were recruited during this data accumulation period is consistent with a continuous flow model of information processing (e.g., Cohen, Dunbar, & McClelland, 1990; Eriksen & Shultz, 1979). Here, the accumulation, deliberation, and response associated with a particular task are proposed to occur continuously during all portions of the process, rather than in a serial manner whereby response information would not contribute to the cognitive process until the task demanded.

The preferential recruitment of the anterior cingulate, left dorsolateral prefrontal cortex, and precuneus imply that participants likely perceived data as error when it was *inconsistent* with a plausible causal theory. The fact that the precuneus and dorsolateral prefrontal cortex are recruited in concert with the anterior cingulate cortex provides important additional information for understanding the types of cognitive mechanisms that may be applied when participants encounter data under such conditions. Considering first the precuneus, there has now been considerable evidence suggesting that the precuneus may have a predominant role in the reallocation of attentional resources (e.g., Handy, Hopfinger, & Mangun,

Table 3
BOLD signal increases at $P \leq .001$ (minimum 5 voxels) for all contrasts of interest as a function of the theory plausibility and covariation manipulations

Brain region	*X*	*Y*	*Z*	*T*-value
Plausible > implausible				
Superior frontal gyrus	18	38	51	5.64
Occipital lobe	3	−84	15	4.57
Inferior frontal gyrus	−39	14	−13	4.98
Implausible > plausible				
No significant activations				
Strong covariation > weak covariation				
Parahippocampal gyrus	−24	−38	−11	4.62
Weak covariation > strong covariation				
Middle temporal gyrus	56	−6	−11	4.28
Lingual gyrus	−6	−64	3	4.42
Precuneus	−3	−68	42	4.49
Plausible strong covariation > plausible weak covariation				
Precentral gyrus	33	−6	30	5.81
Parahippocampal gyrus	−27	−30	−3	5.26
Plausible weak covariation > plausible strong covariation				
Precuneus	−6	−69	45	6.47
Superior frontal gyrus	−36	36	30	5.44
Anterior cingulate cortex	9	21	27	4.53
Implausible strong covariation > implausible weak covariation				
No significant activations				
Implausible weak covariation > implausible strong covariation				
Precentral gyrus	60	0	9	5.43
Caudate	−18	3	15	4.19
Parahippocampal gyrus	−24	−24	−9	4.18

2001; Mazoyer, Wicker, & Fonlupt, 2002; Raichle, 2000; Raichle et al., 2001). That is, when participants reallocate attention away from a task (commonly found in resting states) the precuneus often exhibits increased activity as measured by the fMRI BOLD signal (Table 3). In addition, the selective dorsolateral prefrontal recruitment in this condition may be the result of the active *inhibition* of the attentional processes associated with the task. Recently, Goel and Dolan (2003) found preferential recruitment of the dorsolateral prefrontal cortex in a deductive reasoning task when beliefs and logic were in conflict and required the *inhibition* of a response. Taken together, these findings suggest that the network involving the precuneus, anterior cingulate, and dorsolateral prefrontal cortex may represent the active reallocation of attentional resources when presented with statistical data that are *inconsistent* with one's a priori theory. Indeed, a similar synergistic relationship between the anterior cingulate and the dorsolateral prefrontal cortex has also recently been observed in the Stroop task (Kerns et al., 2004).

These data also speak to the growing work examining the role of the anterior cingulate in error detection and conflict monitoring. Here we show that the brain responds to *conceptual* inconsistencies in a similar manner to what others have found using *perceptual* inconsistencies (e.g., Bush et al., 1998; Kerns et al., 2004). Importantly, the extent to which conflict monitoring networks are recruited as a function of data inconsistency depends on the nature of the inconsistency. That is, regions in the dorsal anterior cingulate cortex, dorsolateral prefrontal cortex, and precuneus were only preferentially recruited when participants encounter data that were inconsistent with a plausible causal theory, and not an implausible causal theory. One possible explanation for this finding is that one's knowledge of the presence of causal mechanisms is more sensitive to inconsistency than one's knowledge of the absence of causal mechanisms. Here, individuals may be more inclined to preserve their beliefs in the face of conflicting data when the representation of that belief reflects a known mechanism of action (Ahn et al., 1995). In addition, the fact that these disparate neural patterns selectively occurred when participants encountered data while evaluating plausible rather than implausible theories provides additional support for the preceding arguments regarding the priority that data for plausible theories may receive. These findings, taken together with the main effects analyses of theory plausibility (both in terms of behavior and the brain), suggest that attentional/working memory priority is given to the analyses of data during the evaluation of plausible as opposed to implausible causal theories.

The fMRI data may also provide a neural instantiation for the growing body of research on *confirmation bias* that has been examined over the past several decades (see Nickerson, 1998 for review). For example, research in cognitive psychology (e.g., Bruner, Goodnow & Austin, 1956; Evans, 1989; Klayman & Ha, 1987; Mynatt, Doherty, & Tweney, 1977; Wason, 1968), scientific thinking (e.g., Cohen, 1985; Gorman, 1989; Mitroff, 1974; Tweney, 1989; Tweney & Doherty, 1983), judicial reasoning (e.g., Fugelsang & Dunbar, 2004; Hendry & Shaffer, 1989; Pennington & Hastie, 1993), medical reasoning (e.g., Elstein & Bordage, 1979), and politics (e.g., Healy, 1996) have all noted the preponderance of confirmatory-based reasoning strategies across many disparate domains. Providing a neural mechanism by which these biases operate may assist in the development of techniques to minimize such biases when they may hinder effective reasoning (see Dunbar, 1993; Evans, 2002; Evans, Newstead, Allen, & Pollard, 1994 for examples of the reduction of reasoning biases through instructional manipulations). A fruitful avenue for future research would be to directly compare belief-bias effects in causal and deductive reasoning within the same individuals. In contrast to our findings within the domain of causal reasoning, several theoretical accounts of deductive reasoning propose that it is the unbelievable information that demands the most attention and working memory processes (e.g., Evans, 1989; Newstead, Pollard, Evans, & Allen, 1992; Oakhill, Johnson-Laird, & Garnham, 1989). Determining the degree to which causal and deductive reasoning recruit common or distinct neural circuitry will aid in the development of more comprehensive general models of reasoning and provide mechanisms describing when processing may differ when the task demands it. In addition, the extent to which these biases, and concurrent recruitment of disparate neural tissue, are the result of automatic or controlled reasoning processes (see Evans, 2003; Fugelsang

& Thompson, 2003) is an important avenue for future research.

These data may also contribute to the development of more comprehensive models of human causal reasoning. Recent models of causality in cognitive science, computer science, developmental psychology, and philosophy have begun to adopt a Bayes net approach to understanding the acquisition and representation of causal knowledge (Glymour, 2001; Gopnik et al., 2004; Pearl, 2000). The data obtained in our experiment are consistent with a Bayesian formulation if one takes into account the role of prior knowledge when judging probabilistic data. When judging probabilistic data, prior knowledge is typically operationalized in terms of one's knowledge of, or use of base-rate information (Bar-Hillel, 1980; Kahneman & Tversky, 1973; Peterson & Beach, 1967). Here, the *posterior* odds of a given hypothesis [$p(H/D)$] using a Bayesian formulation are a product of the *prior* odds of the given outcome occurring [$p(H)$] and the current data to be evaluated [$p(D/H)$]. One can express our manipulations of theory plausibility and covariation-based data probabilistically and thus onto the Bayesian formulations of $p(H)$ and $p(D/H)$, respectively. Here, the degree to which covariation-based data [$p(D/H)$] influences the strength of one's causal judgment when evaluating a theory [$p(H/D)$] is determined in part by the plausibility of the theory being tested [$p(H)$]. Therefore, if the theory being tested is implausible [e.g., small $p(H)$], the covariation-based data is unlikely to significantly impact one's judgment when evaluating a specific causal hypothesis [$p(H/D)$]. By incorporating the role of *priors*, a Bayesian account of causal reasoning may serve as a useful tool to further our understanding of complex reasoning behavior.

Remarkably, the human brain appears to have evolved a particular mechanism for treating the myriads of potentially conflicting information to which an organism is exposed. One of the main riddles of understanding the scientific mind is that there are an infinite variety of models and theories that can be invoked to explain a set of data. By having a mechanism that limits possible interpretations, the brain makes the sheer number of models to be considered tractable.

Acknowledgements

We would like to thank Jonathan Evans, Clay Holroyd, and an anonymous reviewer for comments and helpful suggestions on a previous version of this manuscript. Research reported in this paper has been funded by a grant from Dartmouth College awarded to Kevin Dunbar.

Appendix A. Plausible and implausible causal theories

A.1. Plausible causal theory

(1) Past research has demonstrated that peoples' feelings of happiness *are* directly related to the level of *serotonin* in the brain. The *red pill* is a "selective serotonin reuptake inhibitor". This pill blocks the recycling process for the *serotonin* which then keeps more of this neurotransmitter in the brain available to communicate with other nerve cells.

(2) Past research has demonstrated that peoples' feelings of happiness *are* directly related to the level of *norepinephrine* in the brain. The *red pill* is a "monoamine oxidase inhibitor". Monoamine oxidase is an enzyme that breaks down *norepinephrine* in the brain. Monoamine oxidase inhibitors inhibit this enzyme, thus allowing a greater supply of this neurotransmitter to remain available in the brain.

A.2. Implausible causal theory

(1) Past research has demonstrated that the growth of small amounts of the bacteria *staphylococcus* in the body has *no* direct link to peoples' feelings of happiness. The *red pill* is a "topoisomerase inhibitor". Topoisomerase is an enzyme that is necessary for the reproduction of *staphylococcus* in the body. "Topoisomerase inhibitors" inhibit this enzyme, thus restricting the ability of *staphylococcus* to replicate.

(2) Past research has demonstrated that the growth of small amounts of the bacteria *clostridium* in the body has no direct link to peoples' feelings of happiness. The red pill is a "protein binder". The cell walls of bacteria are continuously expanding through the synthesis of proteins and amino acids. In order for a bacteria cell to flourish and reproduce, the cell wall must be able to expand with the growing interior. "Protein binders" bind to specific amino acids and proteins thus inhibiting the cell wall of *clostridium* to synthesize.

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 judgments of contingency and the nature of response alternatives. *Canadian Journal of Psychology*, *34*, 1–11.

Ashburner, J., & Friston, K. J. (1999). Nonlinear spatial normalization using basis functions. *Human Brain Mapping*, *7*, 254–266.

Bacon, F. (1620/1854). *Novum Organum (Translated by B. Monatgue)*. Philadelphia, PA: Parry & McMillan.

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica*, *44*, 211–233.

Bernasconi, N., Bernasconi, A., Carmanos, Z., Antel, A. B., Andermann, F., & Arnold, D. L. (2003). Mesial temporal damage in temporal lobe epilepsy: A volumetric MRI study of hippocampus, amygdala, and parahippocampal region. *Brain*, *126*, 462–469.

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

Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking*. New York: Wiley.

Bush, G., Luu, P., & Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Science, 4*, 215–222.

Bush, G., Whalen, P. J., Rosen, B. R., Jenike, M. A., McInerney, S. C., & Rauch, S. L. (1998). The counting stroop: An interference task specialized for functional neuroimaging: Validation study with functional MRI. *Human Brain Mapping, 6*, 270–282.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review, 104*, 367–405.

Cheng, P. W., & Novick, L. R. (1990). A probabilistic contrast model of causal induction. *Journal of Personality and Social Psychology, 58*, 545–567.

Cohen, I. B. (1985). *Revolution in science*. Cambridge, MA: Harvard University Press.

Cohen, J. D., Dunbar, K., & McClelland, J. (1990). On the control of automatic processes: A parallel distributed processing account of the stroop effect. *Psychological Review, 97*, 332–361.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers, 25*, 257–271.

Crick, F. (1990). *What mad pursuit: A personal view of scientific discovery*. New York: Basic Books.

Curtis, C. E., & D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends in Cognitive Science, 7*, 415–423.

Damasio, A. R., Eslinger, P. J., Damasio, H., Van Hoesen, G. W., & Cornell, S. (1985). Multimodal amnesic syndrome following bilateral temporal and basal forebrain damage. *Archives of Neurology, 42*, 252–259.

Dunbar, K. (1993). Concept discovery in a scientific domain. *Cognitive Science, 17*, 397–434.

Dunbar, K. (2002). Understanding the role of cognition in science: The science as category framework. In P. Carruthers, S. Stich, & M. Siegal (Eds.), *The cognitive basis of science* (pp. 154–170). New York, NY: Cambridge University Press.

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

Eriksen, B. A., & Shultz, D. W. (1979). Information processing in visual search: A continuous flow conception and experimental results. *Perception & Psychophysics, 25*, 249–263.

Evans, J. St. B. T. (1989). *Bias in human reasoning*. Hillsdale, NJ: Erlbaum.

Evans, J. St. B. T. (2002). Logic and human reasoning: An assessment of the deduction paradigm. *Psychological Bulletin, 128*, 978–996.

Evans, J. St. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Science, 7*, 454–459.

Evans, J. St. B. T., Barston, J. L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition, 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. *European Journal of Cognitive Psychology, 6*, 263–285.

Fan, J., Flombaum, J. I., McCandliss, B. D., Thomas, K. M., & Posner, M. I. (2003). Cognitive and brain consequences of conflict. *Neuroimage, 18*, 42–57.

Friston, K. J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M. D., & Turner, R. (1998). Event-related fMRI: Characterizing differential responses. *Neuroimage, 7*, 30–40.

Friston, K. J., Holmes, A. P., Price, C. J., Buchel, C., & Worsley, K. J. (1999). Multisubject fMRI studies and conjunction analyses. *Neuroimage, 10*, 385–396.

Friston, K., Holmes, A., Worsley, K., Poline, J. B., Frith, C., & Frackowiak, R. (1995). Statistical parametric maps in functional imaging: A general approach. *Human Brain Mapping, 2*, 189–210.

Fugelsang, J. A., & Dunbar, K. N. (2004). A cognitive neuroscience framework for understanding causal reasoning and the law. Philosophical Transactions of The Royal Society of London: Series B, *359*, 1749–1754.

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. *Canadian Journal of Experimental Psychology, 58*, 86–95.

Fugelsang, J. A., & Thompson, V. A. (2000). Strategy selection in causal reasoning: When beliefs and covariation collide. *Canadian Journal of Experimental Psychology, 54*, 13–32.

Fugelsang, J. A., & Thompson, V. A. (2001). Belief-based and covariation-based cues affect causal discounting. *Canadian Journal of Experimental Psychology, 55*, 70–76.

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

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

Glymour, C. (2001). *The mind's arrows: Bayes nets and graphical causal models in psychology*. Cambridge, MA: The MIT Press.

Goel, V., & Dolan, R. J. (2003). Explaining modulation of reasoning by belief. *Cognition, 87*, B11–B22.

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review, 111*, 3–32.

Gorman, M. E. (1989). Error, falsification and scientific inference: An experimental investigation. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology A, 41*, 385–412.

Handy, T. C., Hopfinger, J. B., & Mangun, G. R. (2001). Functional neuroimaging of attention processes. In R. Cabeza & A. Kingstone (Eds.), *Handbook on functional neuroimaging of cognition* (pp. 75–108). Cambridge, MA: MIT Press.

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

Hay, J. F., Moscovitch, M., & Levine, B. (2002). Dissociating habit and recollection: Evidence from Parkinson's disease, amnesia, and focal lesion patients. *Neuropsychologia, 40*, 1324–1334.

Healy, B. (1996). *The dangers of trial by Dingell*. The New York Times.

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. *Journal of Applied Psychology, 74*, 539–545.

Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review, 109*, 679–709.

Holroyd, C. B., Nieuwenhuis, S., Yeung, N., Nystrom, L., Mars, R. B., Coles, M. G., et al. (2004). Dorsal anterior cingulate cortex shows fMRI response to internal and external error signals. *Nature Neuroscience, 7*, 497–498.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*, 237–251.

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

Kapur, S., Tulving, E., Cabeza, R., McIntosh, A. R., Houle, S., & Craik, F. I. (1996). The neural correlates of intentional learning of verbal materials: A PET study in humans. *Cognitive Brain Research, 4*, 243–249.

Kelley, W. M., Miezin, F. M., McDermott, K. B., Buckner, R. L., Raichle, M. E., Cohen, N. J., et al. (1998). Hemispheric specialization in human dorsal frontal cortex and medial temporal lobe for verbal and nonverbal memory encoding. *Neuron, 20*, 927–936.

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.

Kiebel, S. J., Ashburner, J., Poline, J. B., & Friston, K. J. (1997). MRI and PET coregistration: A cross validation of statistical paramet-

ric mapping and automated image registration. *Neuroimage, 5*, 271–279.

Klahr, D., Fay, A., & Dunbar, K. (1993). Heuristics for scientific experimentation: A developmental study. *Cognitive Psychology*, *25*, 111–146.

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

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

Koehler, J. J. (1993). The influence of prior beliefs on scientific judgments of evidence quality. *Organizational Behavior and Human Decision Processes*, *56*, 28–55.

Lazar, N. A., Luna, B., Sweeney, J. A., & Eddy, W. F. (2002). Combining brains: A survey of methods for statistical pooling of information. *Neuroimage*, *16*, 538–550.

Mazoyer, P., Wicker, B., & Fonlupt, P. (2002). A neural network elicited by parametric manipulation of the attention load. *Neuroreport*, *13*, 2331–2334.

McDermott, K. B., Ojemann, J. G., Petersen, S. E., Ollinger, J. M., Snyder, A. Z., Akbudak, E., et al. (1999). Direct comparison of episodic encoding and retrieval of words: An event-related fMRI study. *Memory*, *7*, 661–678.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167–202.

Milner, B., Corkin, S., & Teuber, H. L. (1968). Further analysis of the hippocampal amnesic syndrome: 14-year follow-up study of H.M. *Neuropsychologia*, *6*, 215–234.

Mitroff, I. (1974). *The subjective side of science*. Amsterdam: Elsevier.

Moscovitch, M. (1992). Memory and working with memory: A component process model based on modules and central systems. *Journal of Cognitive Neuroscience*, *4*, 257–267.

Mynatt, B. T., Doherty, M. E., & Tweney, R. D. (1977). Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Quarterly Journal of Experimental Psychology*, *29*, 85–95.

Newstead, S. E., Pollard, P., Evans, J. St. B. T., & Allen, J. T. (1992). The source of belief-bias effects in syllogistic reasoning. *Cognition*, *45*, 257–284.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomena in many guises. *Review of General Psychology*, *2*, 175–220.

Novick, L., & Cheng, P. (2004). Assessing interactive causal inference. *Psychological Review*, *111*, 455–485.

Oakhill, J. V., Johnson-Laird, P. N., & Garnham, A. (1989). Believability and syllogistics reasoning. *Cognition*, *31*, 117–140.

Pearl, J. (2000). *Causality*. Cambridge, UK: Cambridge University Press.

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

Peterson, C. R., & Beach, L. R. (1967). Man as an intuitive statistician. *Psychological Bulletin*, *68*, 29–46.

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

Raichle, M. E. (2000). The neural correlates of consciousness: An analysis of cognitive skill learning. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 1305–1318). Cambridge, MA: MIT Press.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences USA*, *98*, 676–682.

Ranganath, C., Yonelinas, A. P., Cohen, M. X., Dy, C. J., Tom, S. M., & D'Esposito, M. (2003). Dissociable correlates of recollection and familiarity within the medial temporal lobes. *Neuropsychologia*, *42*, 2–13.

Rees, G., Frith, C. D., & Lavie, N. (1997). Modulating irrelevant motion perception by varying attentional load in an unrelated task. *Science*, *278*, 1616–1619.

Rees, G., & Lavie, N. (2001). What can functional imaging reveal about the role of attention in visual awareness? *Neuropsychologia*, *39*, 1343–1353.

Skudlarski, P., Constable, R. T., & Gore, J. C. (1999). ROC analysis of statistical methods used in functional MRI: Individual subjects. *Neuroimage*, *9*, 311–329.

Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science*, *283*, 1657–1661.

Spellman, B. A. (1996). Acting as intuitive scientists: Contingency judgments are made while controlling for alternative causes. *Psychological Science*, *7*, 337–342.

Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain*. New York, NY: Thieme Medical Publishers, Inc.

Todd, P. M., & Gigerenzer, G. (2000). Précis of simple heuristics that make us smart. *Behavioral & Brain Sciences*, *5*, 727–741.

Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1130.

Tweney, R. D. (1989). A framework for the cognitive psychology of science. In B. Gholson, A. Houts, R. A. Neimeyer, & W. Shadish (Eds.), *Psychology of science and metascience* (pp. 342–366). Cambridge, MA: Cambridge University Press.

Tweney, R. D., & Doherty, M. E. (1983). Rationality and the psychology of inference. *Synthese*, *57*, 139–161.

van Veen, V., & Carter, C. S. (2002). The anterior cingulate as a conflict monitor: fMRI and ERP studies. *Physiology & Behavior*, *77*, 477–482.

van Veen, V., Cohen, J. D., Botvinick, M. M., Stenger, V. A., & Carter, C. S. (2001). Anterior cingulate cortex, conflict monitoring, and levels of processing. *Neuroimage*, *14*, 1302–1308.

Wagner, A., Maril, A., & Schacter, D. L. (2000). Interactions between forms of memory: When priming hinders new episodic learning. *Journal of Cognitive Neuroscience*, *12*, 52–60.

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

White, P. A. (1989). A theory of causal processing. *British Journal of Psychology*, *80*, 431–454.

White, P. A. (2002). Perceiving a strong causal relation in a weak contingency: Further investigation of the evidential evaluation model of causal judgment. *The Quarterly Journal of Experimental Psychology A*, *55*, 97–114.

Woods, R. P., Grafton, S. T., Holmes, C. J., Cherry, S. R., & Mazziotta, J. C. (1998). Automated image registration. I. General methods and intrasubject, intramodality validation. *Journal of Computer Assisted Tomography*, *22*, 139–152.

Worsley, K. J., & Friston, K. J. (1995). Analysis of fMRI time-series revisited—again. *Neuroimage*, *2*, 173–181.

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error-detection: Conflict monitoring and the error-related negativity. *Psychological Review*, *111*, 931–959.