

Brain-Based Mechanisms Underlying Causal Reasoning

J.A. Fugelsang(✉) and K.N. Dunbar

Abstract Since well before the time of contemporary psychological and neuroscientific research, philosophers and scientists have been fascinated with the concept of causality. Recent advances of neuroscientific techniques, specifically, neuroimaging using functional MRI, have allowed scientists to probe the brain in order to uncover the mechanisms underlying people's conceptions of causality. In this chapter, we provide an overview of a portion of this recent work, specifically as it pertains to the nature of how people interpret and reason about causality.

1 Introduction

One of the most fundamental attributes of the human mind is its ability to perceive and interpret causal relations apparent in the environment. Indeed, the detection of causal relations is a fundamental ability underlying an individual's success in the dynamic world in which we live. Since before the time of Aristotle (Fig. 1), philosophers and scientists have attempted to provide an account of how we know that one event causes another.

Contemporary theories of causation range from accounts that envisage a common mechanism for understanding causality, to accounts that treat understanding causality, and its resultant representations, as distinct across domains (Sperber et al. 1995). Researchers have recently begun to study the neural underpinnings of several tasks that tap different aspects of causal thinking. Specifically, using a variety of neuroimaging techniques, researchers have examined the perception of causality

AQ: Please check the inserted citation of figure 1.

J.A. Fugelsang
Department of Psychology, University of Waterloo, Waterloo, ON N2L 3G1, Canada
jafugels@uwaterloo.ca

E. Kraft et al. (eds.) *Neural Correlates of Thinking*,
© Springer-Verlag Berlin Heidelberg 2008

263

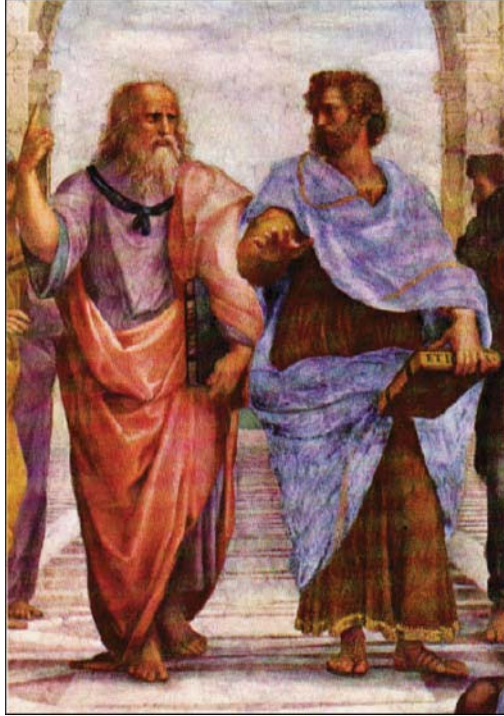


Fig. 1 Aristotle (as depicted in this photograph of a painting by Raphael with his mentor Plato) is often credited as the first academic to formally discuss the concept of causality. He defined four distinct types of cause: the material, formal, efficient, and final types

(Blakemore et al. 2001; Fonlupt 2003; Fugelsang et al. 2005), learning of causal associations (Turner et al. 2004), theory and data interactions in causal thinking (Fugelsang and Dunbar 2005), the representation of stored causal association in semantic memory (Satpute et al. 2005), and the processing of causal inferences in text (Mason and Just 2004). In this chapter, we will provide a brief review of this literature and provide some insights into the possible mechanisms that underlie causal thought. In addition, we will provide some thoughts regarding the degree to which we think neuroimaging can inform the development of theories of causality. We will group our discussions of the literature in terms of (1) studies examining the perception of causality while viewing dynamic displays, (2) learning and reasoning with statistical associations, and (3) accessing stored representation of causal knowledge in semantic memory.

2 Perceptual Causality

Within the physical domain, interactions between moving stimuli, such as collisions, are often reported as involving causal relationships. This can occur even with very simple stimuli such as two moving balls, represented by light patches, on a computer screen. For example, as depicted in the top panel of Fig. 2, if ball A moves toward ball B, stops when it contacts ball B, and B then moves away, the motion of ball B is reported by the majority of observers to have been caused by ball A. If, however, there is a temporal gap (middle panel) or a spatial gap (bottom panel), observers report the relationship as noncausal. This is despite the fact that no inherent causal interaction has occurred, just the simple kinematics as described above. The presence of a small gap or delay (an incontinuity) between the two stimulus movements reduces the likelihood with which stimulus interactions are rated as causal. This collision event has been termed the “launching effect” and is the best-known example of what is called *perceptual causality* (Michotte 1963).

A number of studies have examined the underlying neural processes associated with perceptual causality. The first reported study to this effect was reported by Blakemore et al. (2001); see also Fonlupt (2003) for an additional analysis of those data. They contrasted causal events where a blue ball *collided* with a red ball which subsequently moved, with noncausal events where a blue ball either moved across the screen and passed *under* a stationary red ball or rolled across the screen and

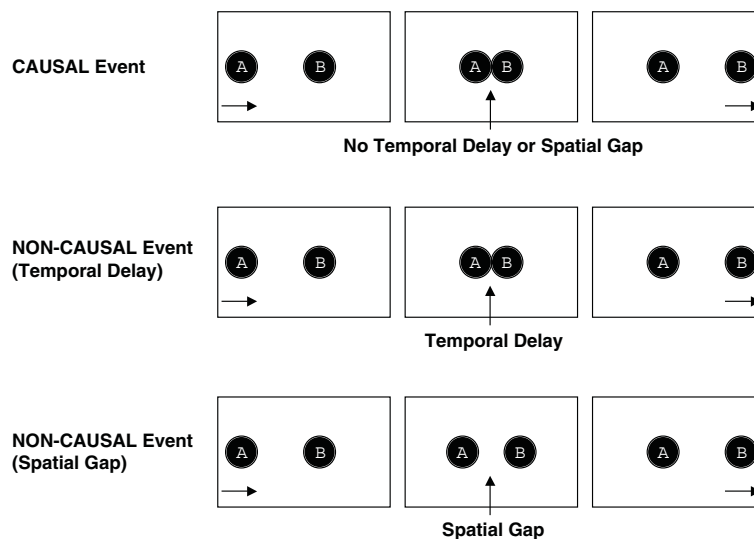


Fig. 2 Causal, temporal delay, and spatial gap events. The three panels depict the motion of a ball A toward a second ball B, and the subsequent motion of ball B

changed color to red after 1 s. They found significant activations in V5, medial, and superior temporal lobes bilaterally, as well as regions in the left superior temporal and intraparietal sulcus. As these regions are strongly implicated in tasks involving complex visual analyses, they argued that the visual system is specifically designed to recover the causal structure of dynamic visual events in the environment. In a related study, Fugelsang et al. (2005) examined the extent to which causal stimuli differentially recruit neural regions associated with spatial and temporal contiguity when those cues to causality are manipulated. Consistent with Blakemore et al. (2001) and Fonlupt (2003), we found similar activations in the temporal lobes when contrasting the causal stimuli to the stimuli with a spatial gap. When causal stimuli were contrasted with stimuli containing a temporal gap, however, activations were predominantly in the frontal and parietal cortices. Importantly, when causal stimuli were contrasted with both noncausal stimuli (those containing spatial and temporal gaps), activations were predominantly found in the frontal and parietal cortices in the right hemisphere (Fig. 3). The frontal activity found is consistent with the frontal activity found by Fonlupt (2003).

The frontal activations may be the product of a variety of processes. Perhaps the most parsimonious explanation is that causal stimuli may recruit additional higher-order executive/attentional resources above and beyond those afforded by the visual system. The preferential recruitment of regions in the prefrontal cortices for causal stimuli suggests that such stimuli may capture visual attention (de Fockert et al. 2004) and result in more attentional resources devoted to such stimuli (Smith and

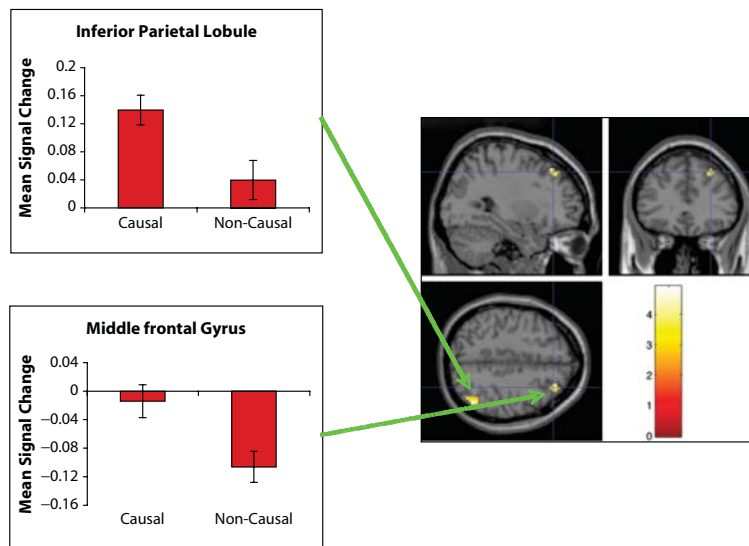


Fig. 3 Significant brain activations and region of interest analyses for causal collision events relative to noncausal events in Fugelsang et al. (2005)

Jonides 1998). Indeed, this allocation of attentional resources and subsequent recruitment of prefrontal cortex may be one of the hallmarks of causality. We will see this throughout the studies reviewed in this chapter.

3 Learning and Reasoning with Statistical Associations

Many causal relations need to be learned through the association of events and outcomes. For example, an individual may learn that he or she has an allergy to peanuts on the basis of the association of allergic reactions arising when foods are eaten that contain peanut products, and not arising when foods are eaten that do not contain peanut products. Several papers have recently been published that look at various aspects of this associative learning process (Corlett et al. 2004; Fletcher et al. 2001; Turner et al. 2004). One of the key components of associative models is that learning depends on surprise. For example, surprising outcomes are thought to enhance attention to stimuli, and thus promote learning. Fletcher et al. (2001) found support for this hypothesis in that initially novel and surprising stimuli produced maximal activation in dorsolateral prefrontal cortex when participants are learning associations. This heightened activation attenuated through learning, but was re-evoked when surprise violations of the learned association were present.

These data are consistent with recent work on theory and data interaction in complex causal reasoning conducted by Fugelsang and Dunbar (2005). We presented participants with a task requiring them to interpret data relative to plausible and implausible causal theories. The plausibility of the causal theories was manipulated by presenting participants with a brief introductory statement that depicted a causal theory that contained either a plausible or implausible causal mechanisms. Data were then provided to participants in a trial-by-trial format where they viewed multiple trials of data for each causal theory provided. These data were presented in combinations of a cause (a *red pill* or a *blue pill*) and an effect (*happiness* or *neutral* outcome) co-occurring. Figure 4 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. Importantly, the trials cumulatively presented data that were either consistent with the initial theory, or inconsistent with the theory.

We analyzed the imaging data in two stages. We were first interested in looking at the degree to which different regions of the brain would be selectively responsive to reasoning with scenarios that contained plausible as opposed to implausible causal mechanisms. Second, we examined the degree to which the consistency of the relationship between the plausibility of the causal mechanism and the data modulated the recruitment of dissociable neural regions. Considering first the effect of mechanism plausibility, like our work on perceptual causality, regions of the right superior frontal gyrus were more activated when subjects were reasoning about candidates that contained plausible causal mechanisms than candidates that contained implausible causal mechanisms. We interpreted these findings to suggest

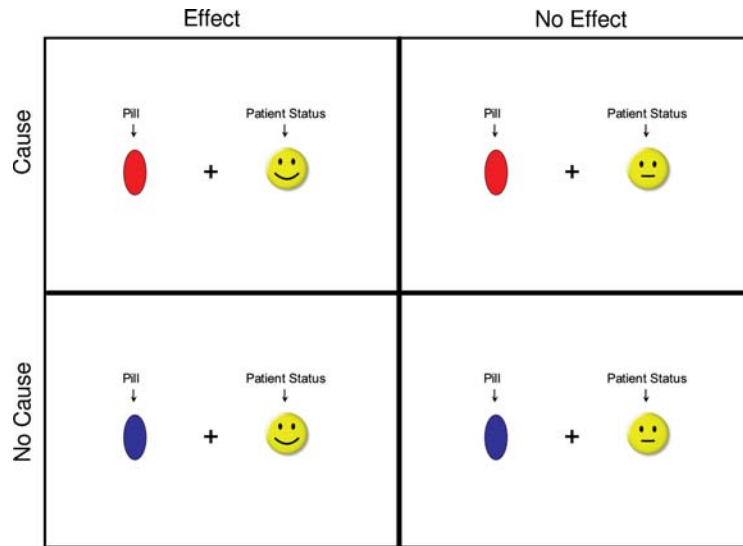


Fig. 4 Example stimuli representing the four possible combinations of the candidate cause (red pill versus blue pill) and effect (happiness versus neutral emotion) used by Fugelsang and Dunbar (2005)

that plausible causal mechanisms, like collisions that conform to expectations in the Michotte task, capture attention and thus are subject to executive processing (Curtis and D'Esposito 2003; Smith and Jonides 1999).

Concerning our second level of analyses, we found that the consistency between theory and data influenced the degree to which disparate neural tissue associated with learning or conflict monitoring/error detection. Specifically, as can be seen in Fig. 5, when data were consistent with the theory (regardless of its plausibility), activations were found in the caudate and parahippocampal gyrus. When the statistical data were *inconsistent* with a theory, however, the anterior cingulate cortex and precuneus were selectively recruited. A further important finding emerged when we looked at the effects of data consistency for plausible and implausible theory separately. Specifically, when participants viewed data that were inconsistent with a *plausible* theory, further activations on the left prefrontal cortex were also found to occur in concert with the activation in anterior cingulate cortex and precuneus. There are several possible interpretations of these findings. Our preferred interpretation is that participants likely perceived data as error when they were *inconsistent* with a *plausible* causal theory. In addition, the selective dorsolateral prefrontal recruitment in concert with the anterior cingulate cortex 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

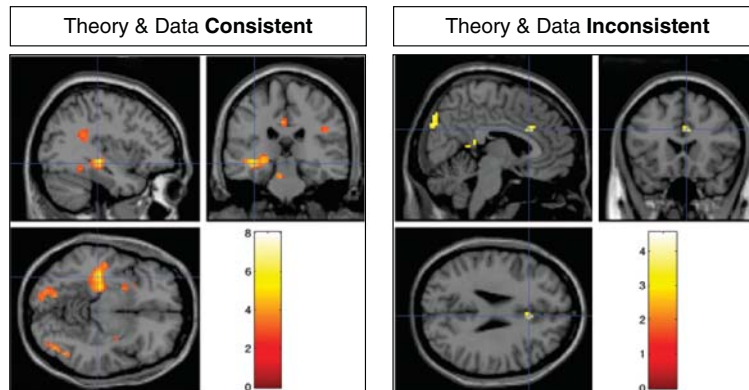


Fig. 5 Unique task-associated brain activations occurring when viewing data *inconsistent* or *consistent* with a causal theory

the *inhibition* of a response. Another possible interpretation, one that would be consistent with the findings of Fletcher et al. (2001), is that data inconsistent with a causal theory are surprising, and thus preferentially recruit attentional resources.

4 Accessing Stored Causal Relations

As we discussed in the previous sections, there is considerable evidence that (1) the visual system is specially tuned to extract some types of causal relations present in the environment and (2) other types of causal relationships are likely learned through observation. An important avenue of research has been examining how these stored associations are represented and accessed in semantic memory. Are causal associations, like wind and erosion, stored and processed the same as noncausal associations, like bread and butter? Satpute et al. (2005) examined this issue by presenting participants with causal word pairs (along with additional control stimuli), and requiring them to make either associative or causal judgments on these word pairs. Their basic hypothesis was that assessing the causal nature of relations required additional processing than simply judging mere associations. Specifically, they argued that judging causality likely requires a process called *dynamic role binding* (Hummel and Holyoak 2003). Evaluating causal relations may require forming and holding an explicit representation of the specific events bound to the roles of cause and effect. For example, in the wind/erosion example presented previously, a participant needs to evaluate the specific cause and effect roles of both items in order to make the causal judgment.

They found that causal judgments, in contrast to associative judgments, preferentially recruited regions in the left dorsolateral prefrontal cortex and the precuneus.

These activations are consistent with the idea that assessing causality requires additional neurocognitive attentional resources in order to evaluate the causal roles of the word pair in working memory. Interestingly, there are cases in which causal relatedness results in a reduction of working memory resources, and subsequently a reduction in prefrontal cortex activation. For example, pairs of sentences like “Joey’s big brother punched him again and again” and “The next day his body was covered in bruises” are typically read faster, and recruit less prefrontal cortex than sentences such as “Joey went to a neighbor’s house to play” and “The next day his body was covered in bruises” (Mason and Just 2004). Presumably, causal relations in such sentence pairs help bind the sentences together and thus require less generation of inferences. Here, the presence of a clear causal relationship serves to reduce the number of inferences required on the part of the reader.

5 Single or Multiple Causal Representations

The final area of research we wish to touch on concerns the investigation of single versus multiple representations of causality. Throughout the different sections of this chapter, we have covered a variety of tasks that all involve causal processing in some form or another. Do all of these disparate tasks invoke the same underlying processes when judging causality? There has been considerable debate in the literature regarding the extent to which judgments of causality are the product of single or multiple underlying processes (Scholl and Tremoulet 2000; Schlottmann 2000). For example, the possibility that some events can be directly perceived as causal (e.g., using the Michotte paradigm as discussed previously) suggests that there may be multiple representations or processes that support judgments of causality. Does perceptual causality represent a unique form of human causal processes that can be dissociated from that based on more associative processes? Take first the concept of perceptual causality. The findings that the perception of causality appears very early in human life (Leslie and Keeble 1987) and is culturally invariant (Morris and Peng 1994) have been taken to suggest that the visual system may be specially tuned to recover *physical* causal structure from the environment. This can be contrasted with *casual inference*, which demands the learning of causal associations based on covariation experience. This ability to *learn* causal associations develops somewhat later in life (Gopnik et al. 2001). A series of experiments lead by Roser et al. (2005) investigated whether causal perception (using the standard Michotte paradigm) could be dissociated from a task requiring causal inference. An obvious difficulty arises with determining the extent to which causal perception and causal inference are subserved by the same or different underlying processes using traditional behavioral measures. To date, the majority of research testing for the existence of unique processes supporting causal perception and inference has come from observers’ subjective reports which are highly subjective and open to postperceptual interpretation in the “normal” brain. This difficulty, however, can be overcome by using split-brain patients who have undergone surgery to isolate the two cerebral

hemispheres. By analyzing the degree to which each of the isolated cerebral hemispheres processes causality, we can determine the extent to which perceptual and inferential processes involved in understanding causality can be dissociated.

Perhaps the most obvious functional hemispheric asymmetry in humans is that of linguistic versus visual-spatial processing. For example, decades of work with patients and countless functional imaging studies have shown that the right hemisphere possesses an advantage for tasks that require visuospatial processing (Corballis 2003; Corballis et al. 2002), whereas the left hemisphere processes an advantage for linguistic processing (Milner 1962). Taking this as our starting point, we predicted that the right hemisphere would exhibit an advantage for perceptual causality, whereas the left hemisphere would exhibit an advantage for inferential causality. Two patients who underwent callosotomy surgery were presented with causal collision events using the standard Michotte paradigm, and a task requiring causal inference which we adapted from Gopnik et al. (2001). The inference task consisted of a short series of four stimulus interactions wherein the participants simply had to judge which of two “switches” (green or red) caused a “lightbox” to turn on. The data were consistent with our hypotheses in that the ability to draw causal inference and the ability to judge causal relationships during the standard Michotte paradigm were governed by different hemispheres of the divided brain (Fig. 6). Specifically, the right hemispheres were significantly more sensitive to the causal perception task than the inference task, whereas the left hemispheres were significantly more sensitive to the inference task than the perception task. These data were taken to support

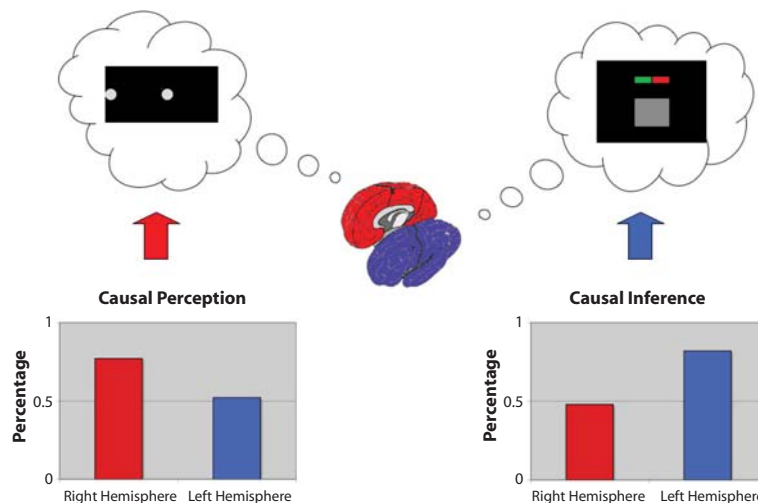


Fig. 6 Causal perception and inference task and the data observed by Roser et al. (2005). The inference tasks consisted of the sequential presentation of four stimulus interactions between a red and green switch and a lightbox. Refer to Roser et al. (2005) for a complete description of the task

the hypothesis that causation is a nonunitary construct, and that causal perception and inference can be processed independently.

We have recently extended this work to examine the degree to which dissociable perception and inference mechanisms are invoked in the “normal” brain using functional MRI. To do this, participants were asked to respond to a simple collision similar to those used in our previous work (Fugelsang et al. 2005). Importantly, however, we also manipulated the cover story which participants used to evaluate the collision events. For half the trials participants were told to “. . . imagine that the two circular objects are billiard balls.” For the other half of the trials, participants were told to “. . . imagine that the two circular objects are positively charged particles that repel each other when they come close to being in contact with each other.” It was the intention that this manipulation would change the task from a perceptual to an inferential task in that the requirement to imagine the objects as “positively charged particles” requires one to infer basic characteristics about the objects that may be in conflict with the perceptual experience. Preliminary analyses of these data have revealed a pattern consistent with that observed with the experiments with split-brain patients. Specifically, when making judgments on the stimuli when they were to be thought of as billiard balls, regions in the right superior frontal and inferior parietal cortices were recruited. In contrast, when participants were making judgment on stimuli when they were to be thought of as positively charged particles, homologous regions in the left frontal and parietal cortices were recruited in concert with those in the right hemisphere. These data are taken to support a multidimensional interpretation of causality that involves an interplay between basic perception and/or inference depending on the nature of the task.

6 Future Directions

In the preceding sections we have outlined a number of research programs that have recently contributed to our understanding of the nature of human casual thinking. As is evident from the diverse fields of research, finding a single region in the brain that uniquely represents causal thinking in humans is likely an unrealistic goal. This may speak more to the diversity of the research areas and the kinds of questions being asked than to the nature of causality. Although great progress has been made, many questions remain to be answered which will benefit from creative functional imaging experiments.

References

- Blakemore S, Fonlupt P, Pachot-Clouard M, Darmon C, Boyer P, Meltzoff A, Segebarth C, Decety J (2001) How the brain perceives causality: an event-related fMRI study. *Neuroreport* 12:3741–3746
- Corballis PM (2003) Visuospatial processing and the right-hemisphere interpreter. *Brain Cognit* 53:171–176

- Corballis PM, Funnell MG, Gazzaniga MS (2002) Hemispheric asymmetries for simple visual judgments in the split brain. *Neuropsychologia* 40:401–410
- Corlett PR, Aitken MR, Dickinson A, Shanks DR, Honey GD, Honey RA, Robbins TW, Bullmore ET, Fletcher PC (2004) Prediction error during retrospective reevaluation of causal associations in humans: fMRI evidence in favor of an associative model of learning. *Neuron* 44:877–888
- Curtis CE, D'Esposito M (2003) Persistent activity in the prefrontal cortex during working memory. *Trends Cogn Sci* 7:415–423
- de Fockert J, Rees G, Frith C, Lavie N (2004) Neural correlates of attentional capture in visual search. *J Cogn Neurosci* 16:751–759
- Fletcher PC, Anderson JM, Shanks DR, Honey R, Carpenter TA, Donovan T, Papadakis N, Bullmore ET (2001) Responses of human frontal cortex to surprising events are predicted by formal associative learning theory. *Nat Neurosci* 4:1043–1048
- Fonlupt P (2003) Perception and judgment of physical causality involve different brain structures. *Cogn Brain Res* 17:248–254
- Fugelsang J, Dunbar K (2005) Brain-based mechanisms underlying complex causal thinking. *Neuropsychologia* 43:1204–1213
- Fugelsang J, Roser M, Corballis P, Gazzaniga M, Dunbar K (2005) Brain mechanisms underlying perceptual causality. *Cogn Brain Res* 24:41–47
- Goel V, Dolan RJ (2003) Explaining modulation of reasoning by belief. *Cognition* 87:B11–B22
- Gopnik A, Sobel DM, Schulz LE, Glymour C (2001) Causal learning mechanisms in very young children: two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Dev Psychol* 37:620–629
- Hummel JE, Holyoak KJ (2003) A symbolic-connectionist theory of relational inference and generalization. *Psychol Rev* 110:220–264
- Leslie AM, Keeble S (1987) Do six-month old infants perceive causality? *Cognition* 25:265–288
- Mason RA, Just MA (2004) How the brain processes causal inferences in text. *Psychol Sci* 15:1–7
- Michotte A (1963) *The perception of causality*. Basic Books, New York
- Milner B (1962) Laterality effects in audition. In Mountcastle VB (ed) *Interhemispheric relations and cerebral dominance*. Johns Hopkins, Baltimore, pp 177–198
- Morris NW, Peng K (1994) Culture and cause: American and Chinese attributions for social and physical events. *J Pers Soc Psychol* 67:949–971
- Roser M, Fugelsang J, Dunbar K, Corballis P, Gazzaniga M (2005) Dissociating processes supporting causal perception and causal inference in the brain. *Neuropsychology* 19:591–602
- Satpute AB, Fenker DB, Waldmann MR, Tabibnia G, Holyoak KJ, Lieberman MD (2005) An fMRI study of causal judgments. *Eur J Neurosci* 22:1233–1238
- Schlottmann A (2000) Is perception of causality modular? *Trends Cogn Sci* 4:441–442
- Scholl BJ, Tremoulet PD (2000) Perceptual causality and animacy. *Trends Cogn Sci* 4:299–309
- Smith EE, Jonides J (1999) Storage and executive processes in the frontal lobes. *Science* 283:1657–1661
- Sperber D, Premack D, Premack AJ (1995) *Causal cognition: a multidisciplinary debate*. Oxford University Press, Oxford
- Turner DC, Aitken MR, Shanks DR, Sahakian BJ, Robbins TW, Schwarzbauer C, Fletcher PC (2004) The role of the lateral frontal cortex in causal associative learning: exploring preventative and super-learning. *Cereb Cortex* 14:872–880

