CHAPTER 4
Architecture of Counterfactual Thought in the Prefrontal Cortex

Aron K. Barbey, Frank Krueger, and Jordan Grafman

Remembering the past and predicting the future depend on the ability to shift from perceiving the immediate environment to an alternative, imagined perspective. Mental models of imagined past events or future outcomes not yet at hand support *counterfactual thinking* ("What would happen if X were performed in the past or enacted in the future?") (Byrne, 2002; Kahneman & Miller, 1986). The capacity for counterfactual thought enables learning from past experience (Byrne, 1997), supports planning and predicting future events (Barbey & Sloman, 2007; Brase & Barbey, 2006), provides the basis for creativity and insight (Costello & Keane, 2000; Sternberg & Gastel, 1989; Thomas, 1999), and gives rise to emotions and social ascriptions (e.g., guilt, regret, and blame) that are central for managing and regulating social behavior (Davis, Lehman, Wortman, Silver, & Thompson, 1995; Landman, 1987; Miller & Turnbull, 1990; Niedenthal, Tangney, & Gavanski, 1994; Zeelenberg, van der Pligt, & Manstead, 1998). The neural representation of counterfactual inference draws upon neural systems for constructing mental models of the past and future, incorporating prefrontal and medial temporal lobe structures (Fortin, Agster, & Eichenbaum, 2002; Tulving & Markowitsch, 1998). In this chapter, we develop an integrative cognitive neuroscience framework for understanding counterfactual reasoning on the basis of *structured event complexes* (SECs) in the human prefrontal cortex (PFC).

We begin by reviewing the biology and structure of the human PFC and introduce a cognitive neuroscience framework for the representation of event knowledge within the PFC. We then survey recent neuroscience evidence in support of the SEC framework and establish the role of distinct PFC subregions in the representation of specific forms of event knowledge. After reviewing the cognitive and neural foundations of the SEC framework, we show how this approach accounts for counterfactual reasoning. We identify three major categories of counterfactual inference (concerning action versus inaction, the self versus other, and upward versus downward thinking) and review neuroscience evidence for their representation within distinct regions of the medial PFC. We propose that mental models for goal-directed social behavior additionally recruit the lateral PFC, which represents behavior-guiding principles for counterfactual inference concerning obligatory, prohibited, and permissible courses of action. We survey recent evidence from the decision neuroscience literature to support the representation of behavior-guiding principles for counterfactual inference within distinct regions of the lateral PFC. Finally, we draw conclusions about the importance of SECs for learning from past experience, for planning and predicting future events, for creativity and insight, and for the management and regulation of social behavior.

Neurobiology of the Human Prefrontal Cortex

Structured event complexes (SECs) are representations composed of goal-oriented sequences of events involved in executing, planning, and monitoring action. We briefly review the biology and structure of the human PFC, providing evidence to support our proposal that the PFC stores cognitive representations intimately concerned with goal-directed action.

The PFC can be divided into ventromedial and dorsolateral regions, each of which is associated with posterior and subcortical brain regions. The ventromedial PFC (vmPFC) has reciprocal connections with brain regions that are associated with emotional processing (amygdala), memory (hippocampus), and higher order sensory processing (temporal visual association areas), as well as with dorsolateral PFC (dlPFC). The dlPFC has reciprocal connections with brain regions that are associated with motor control (basal ganglia, premotor cortex, supplementary motor area), performance monitoring (cingulate cortex), and higher order sensory processing (association areas, parietal cortex). The vmPFC is well suited to support functions involving the integration of information about emotion, memory, and environmental stimuli, and the dlPFC is well suited to support the regulation of behavior and control of responses to environmental stimuli.

Prefrontal cortex neurons are particularly able to fire over extended periods of time (Levy & Goldman-Rakic, 2000) and across events (Bodner, Kroger, & Fuster, 1996; Fuster & Alexander, 1971). This indicates that the PFC can maintain stimulus representations across time, enabling a subject to engage in behavior to achieve long-term goals. In addition, pyramidal cells in the macaque PFC are more spinous—and therefore can handle more excitatory inputs—than other cortical pyramidal cells (Elston, 2000). This is one structural explanation for the PFC's ability to integrate inputs from many sources and to implement complex behaviors. The monkey's PFC contains cells that respond to both internally generated and observed behaviors—these have been termed *mirror neurons* (Gallese et al., 1996). Similar regions have been shown to be activated in humans when observing and performing actions (Grafton, Arbib, Fadiga, & Rizzolatti, 1996). These data support a role for the PFC in the representation of action. Furthermore, Williams and colleagues have suggested that abnormal development of the PFC might lead to impaired social behavior (Williams, Whiten, Suddendorf, & Perret, 2001), which can also be caused by PFC damage later in life.

It is thought that the dlPFC evolved from motor regions and developed much later than the vmPFC (Fuster, 1997). Motor regions store motor programs; therefore, it seems reasonable that the functions of the "newer" PFC regions would be related to those of older PFC regions, providing a representational basis for goal-directed action.

In summary, the connectivity of PFC regions, physiological properties of its neurons, and evolutionary principles are strongly suggestive of its role in the integration of sensory and memory information and in the representation and control of actions and behavior. Along with the extended firing of neurons, specialized neural systems were developed that enabled the parsing and encoding of these behaviors into sequentially linked but individually recognizable events. At the broadest level, events are parsed into subcomponents consisting of an activity that signals the onset of the event, followed by a series of activities performed to achieve the desired goal, and a final activity resulting in event completion. Events are further characterized by their semantic content, temporal duration, and the number of component activities they entail (Zacks & Tversky, 2001; Zacks, Tversky, & Iyer, 2001).

We propose that the structure of event knowledge can be conceptualized as a "representation" or a unique form of knowledge that, when activated, corresponds to a dynamic brain state signified by the strength and pattern of neural activity in a local brain region. In this sense, over the course of evolution, the PFC became capable of representing knowledge of more complex behaviors. We label these representational units within the PFC *structured event complexes* (SECs).

Structured Event Complex Theory

A SEC represents event knowledge consisting of agents, objects, actions, mental states, and background settings that are temporally structured and semantically organized according to their causal roles (e.g., as cause, effect, enabler, or preventer). The SEC theory is a representational framework that motivates specific predictions regarding the properties and localization of SECs within the PFC (Fig. 4.1). We review principal elements of the SEC theory before turning to an assessment of its neurobiological predictions.

Neural Architecture

Structured event complexes are encoded and activated on the basis of simulation mechanisms (Barsalou, Niedenthal, Barbey, & Ruppert, 2003; Barsalou, Simmons, Barbey, & Wilson, 2003;

STRUCTURED EVENT COMPLEX PREFONTAL CORTEX		
HEMISPHERIC LATERALIZATION		
LEFT PFC		**RIGHT PFC**
Single event integration		**Across events integration**
- Meaning and feature between single adjacent events to code for boundaries between events		- Meaning and features across events to obtain goal of sequence
CATEGORY SPECIFICITY		
LATERAL PFC		**MEDIAL PFC**
Behavioral norms and scripts		**Planning and action**
- Event sequences representing social norms and scripts		- Event sequences representing mechanistic plans and actions
COMPLEXITY		
ANTERIOR PFC		**POSTERIOR PFC**
Complex SECs		**Simple SECs**
- Detailed information about event sequences		- Spare information about event sequences

Figure 4.1 SEC framework. The representational forms of the structured event complex (SEC) and their proposed localizations within the prefrontal cortex (PFC).

Damasio, 1989). A large body of neuroscience evidence demonstrates that experience in the physical and social world activates feature detectors in relevant features maps of the brain. During visual processing of a face, for example, some neurons fire for edges and planar surfaces, whereas others fire for color, configural properties, and movement. The global pattern of activation across this hierarchically organized distributed system represents the entity in vision (Palmer, 1999; Zeki, 1993). Analogous patterns of activation on other sensory modalities represent how the face might sound and feel. Activation in the motor system similarly represents responses to the face, such as the formation of a facial expression, and approach/avoidance behavior. A similar mechanism underlies the introspective states that arise while interacting with an entity. For example, activation patterns in the amygdala and orbitofrontal areas represent emotional reactions to social stimuli. Much neuroscience evidence documents the structure of feature maps across modalities and the states that arise in them.

When a pattern becomes active in a feature map during perception or action, conjunctive neurons in an association area capture the pattern for later cognitive use. For example, conjunctive neurons in the visual system capture the pattern active for a particular face. A population of conjunctive neurons together codes a particular pattern, with each individual neuron participating in the coding of many different patterns. Damasio (1989) called these association areas *convergence zones* and proposed that they exist at multiple hierarchical levels in the brain, ranging from posterior to anterior. Most locally, convergence zones near a modality capture activation patterns within it. Association areas near the visual system, for example, capture patterns there, whereas association areas near the motor system capture patterns in this local region. Downstream in more anterior regions, higher association areas, including temporal and frontal regions, integrate activation across modalities.

According to the SEC framework, event knowledge is represented by higher order convergence zones localized within particular regions of the PFC (see Fig. 4.1). Once a set of conjunctive neurons within the PFC captures feature maps (representing components of event knowledge, social norms, ethical and moral rules, and temporal event boundaries), the set can later activate the pattern in the absence of bottom-up stimulation. For example, on entering a familiar situation and recognizing it, an SEC that represents the situation becomes active. Typically not all of the situation is perceived initially. A relevant person, setting, or event may be perceived, which then suggests that a particular situation is about to play out. It is in the agent's interests to anticipate what will happen next so that optimal actions can be executed. The agent must draw inferences that go beyond the information given (Griffin & Ross, 1991). The SEC that becomes active constitutes a rich source of social inference supporting the planning, execution, and monitoring of action. The SEC can be viewed as a distributed pattern representing components of event knowledge (i.e., as a complex configuration of multimodal components that represent the situation). Because part of this pattern matched the current situation initially, the larger pattern became active in memory. The remaining parts of the pattern—not yet observed in the situation—constitute inferences, namely educated guesses about what might occur next. Because the remaining parts co-occurred frequently with the perceived parts in previous situations, inferring the remaining parts from the perceived parts is reasonable. As a partially viewed situation activates an SEC, the SEC completes the pattern that the situation suggests.

To the extent that the SEC is entrenched in memory, pattern completion is likely to occur at least somewhat automatically. As a situation is experienced repeatedly, its simulated components and the associations linking them increase in potency. Thus, when one component is perceived initially, these strong associations complete the pattern automatically. Consider the example of meeting with a colleague. Her face, clothing, and bodily mannerisms initially match modality-specific simulations in one or more SECs that have become entrenched in memory. Once one of these wins the activation process, it provides inferences via pattern completion, such as actions that the colleague is likely to take, actions that the perceiver typically takes, affective states that are likely to result, and so forth. The unfolding of such inferences—realized as an SEC—produces social prediction (for a cognitive neuroscience review of simulation mechanisms in reasoning, see Barbey & Barsalou, in press; Barsalou, Barbey, Simmons, & Santos, 2005; Patterson & Barbey, in press).

Sequence Structure

Structured event complexes integrate modality-specific components of event knowledge, providing the semantic and temporal structure underlying goal-directed action. Components of event knowledge are integrated on the basis of their causal roles (e.g., as cause, effect, enabler, preventer). At the broadest level, SECs link event subcomponents consisting of an activity that signals the onset of the event (e.g., "hearing the morning alarm clock"), followed by a series of activities performed to achieve the desired goal (e.g., "waking up," "getting out of bed," etc.), and a final activity resulting in event completion (e.g., "arriving to work"). The temporal structure of SECs further obeys cultural and individual constraints, reflecting sociocultural norms of appropriate behavior (e.g., in the United States, people typically shower in the morning daily) and personal preferences concerning the temporal order and frequency of performed activities (e.g., the individual preference to shower in the morning and at night daily). The semantic and temporal structure of event knowledge supports goal-directed action in dynamic environments, enabling the on-line modification of specific activities (e.g., due to changing circumstances) and the simulation of only those event components necessary for goal achievement in the present context (e.g., beginning at various stages in the event sequence, returning to earlier stages, skipping unnecessary activities due to time pressure, etc.).

Goal Orientation

The semantic and temporal structure of SECs derives from event goals, which provide the basis for the selection, temporal ordering, and execution of activities underlying an event. Some SECs are *well structured*, with clearly defined goals, and cognitive and behavioral action sequences that are available for goal achievement. For example, individuals with a well-structured SEC for "eating in a restaurant" are quite confident that once they have been seated at a table and have read the menu, someone will appear to take their order.

In contrast, some SECs are *ill structured*, requiring the individual to adapt to unpredictable circumstances by constructing novel or ad hoc goals, and selecting appropriate action sequences on-line (Barsalou, 1991). For example, if someone sees that a person entering a bank is wearing a ski-mask and carrying a gun, one can make sense of these events by completing the activated "bank robbery" SEC to access further components of event knowledge (concerning relevant agents, objects, actions, mental states, and background settings).

Binding

Multiple SECs are activated to support the events of our daily life; therefore, it is likely that these representations (like events within an SEC) can be activated in sequence, or additionally in a cascading or parallel manner. Event components interact and give rise to SECs through at least three binding mechanisms: sequential binding, proposed for linking multiple SECs within the PFC (Weingartner, Grafman, Boutelle, Kaye, & Martin, 1983); temporal binding among anatomically integrated regions representing event subcomponents in posterior cortex (Engel & Singer, 2001); and third-party binding of anatomical regions whose activity is synchronized via the hippocampus (O'Reilly & Rudy, 2000; Weingartner et al., 1983).

Hierarchical Structure

Given the slow development of the PFC during childhood, individual events are probably initially represented as independent memory units. For example, SECs associated with "kitchen" and "school cafeteria" cluster around the event "eat meal," whereas "car" and "school bus" cluster around the event "travel to new location." Later in development, these primitive SECs expand into large multievent units, based on repeated exposure and goal-directed action. In addition, the boundaries of event sequences become more firmly established, leading to a well-structured SEC. Thus, in adulthood, SECs will range from specific episodes to more abstract SECs that can be applied to a variety of situations (Barsalou & Wiemer-Hastings, 2005). For example, the domain "eat meal" includes specific episodes representing evenings at a particular restaurant, SECs representing the actions and themes of how to behave at different types of restaurants, in addition to more abstract SECs representing actions and themes related to "eating" that apply to a broad range of situations (e.g., at "restaurants," "parties," "picnics," "baseball games," etc.).

COUNTERFACTUAL THOUGHT

We propose that SECs provide the basis for counterfactual reasoning about past and future events and develop a process model of the regulatory functions these representations serve.

Counterfactual thinking involves mentally undoing the present state of affairs and imagining alternative realities "if only" different decisions were made or actions taken (Byrne, 2002; Kahneman & Miller, 1986). We propose that counterfactual thought depends on mental models of alternative possibilities represented in the form of SECs. For example, the counterfactual inference that "If we chose to sail the Mediterranean rather than continue writing, then you would not be reading this chapter" draws upon SEC knowledge, including the representation of relevant agents (e.g., the authors), objects (e.g., a sailboat), actions (e.g., sailing), mental states (e.g., freedom), and background settings (e.g., the Mediterranean Sea). Simulations of the representational elements of SEC knowledge provide the basis for evaluating the consequences of alternative courses of action, with the simulation of the authors "sailing the Mediterranean" resulting in the failure to complete this chapter.

A growing body of research demonstrates the importance of counterfactual inference for generating predictions about the future, supporting representations of unknown, future possibilities critical for planning and decision making (e.g., "How well would the Cubs perform next season if the manager would have acquired key players in the off season?") (Barbey & Sloman, 2007; Brase & Barbey, 2006). Predictions about the future are supported by modifying a factual event (e.g., "If the manager acquired key players in the off season...") and considering likely future consequences (e.g., "... would the team perform well next year?").

The SEC framework advocates a theory of *motivated thinking* (De Dreu, Beersma, Stroebe, & Euwema, 2006; Dunning, 1999), proposing that *drives*, *needs*, *desires*, *motives*, and *goals* structure and organize components of event knowledge and profoundly influence judgment and decision making. According to this framework, the primary role of counterfactual thought is to support emotions and social ascriptions that are central for managing and regulating social behavior. In particular, counterfactual reasoning enables the representation of guilt, regret, and blame, which are central for adaptive social behavior (Davis et al., 1995; Landman, 1987; Miller & Turnbull, 1990; Niedenthal et al., 1994; Zeelenberg et al., 1998). For example, the counterfactual inference that "The university would offer her a higher salary (in the past or future) if she were a man" gives rise to feelings of guilt and regret (for the observed gender inequity) that promote behavioral change and that enable the assessment of blame (held by university policy makers) to support planning and decision making (e.g., to apply for positions at other universities). Counterfactual inference therefore enables an assessment of the consequences of alternative decisions or actions sequences central for the representation of guilt, regret, and blame.

The neural representation of emotions and social ascriptions (e.g., guilt, regret, and blame) is distributed throughout the mPFC and is integrated with posterior knowledge networks via binding mechanisms in the medial temporal lobe (Fortin et al., 2002; Moll & de Oliveira-Souza, 2007; Tulving et al., 1998). This distributed pattern of multimodal information (e.g., representing agents, objects, actions, mental states, and background settings) gives rise to mental models for counterfactual inference.

Process Model

According to the SEC framework, counterfactual thought is deeply connected to drives, needs, desires, motives, and goals, and it provides the basis for regulatory mechanisms that keep behavior on track, particularly within social interactions. We propose that counterfactual thought depends on SEC representations and operates according to the following interactive process model (see Fig. 4.2).

i. Counterfactual thoughts are activated when a problem is encountered or anticipated in the future. Failure to achieve the desired goal or the anticipation of goal failure in the future typically initiates counterfactual thinking (e.g., due to negative emotions, the desire for rewards associated with goal achievement, etc.).
ii. Counterfactual thoughts are generated from causal implications represented by SECs in the form of events (agents, objects, actions, mental states, and background settings) that lead to a desired goal state (for a review of psychological theories of causal representation and reasoning, see Barbey & Wolff, 2006, 2007; submitted; Chaigneau & Barbey, 2008; Patterson & Barbey, in press; Sloman, Barbey, & Hotaling; in press).
iii. Structured event complexes activate corresponding behavioral intentions (e.g., to perform a particular action), mindsets (e.g., to focus on a particular class of events), motivations (e.g., to modulate one's desire for a particular outcome), and/or self-inferences (e.g., to monitor one's public image) that initiate corrective behavior.

To the extent that such behavior alleviates the original problem, this mechanism is effective in regulating behavior in terms of goal pursuit (for a review of medical health applications, see Gilbar & Heyroni, 2007; Wrosch, Bauer, Miller, & Lupien, 2007).

Categories of Counterfactual Inference

The proposed role of SECs in counterfactual thought motivates the prediction that this form of inference will depend on core elements of SECs, which fundamentally represent *actions* performed by *agents* leading to an observed *outcome*. The psychological literature on counterfactual thought supports this prediction, identifying three major categories of counterfactual inference corresponding to core components of SEC representations.

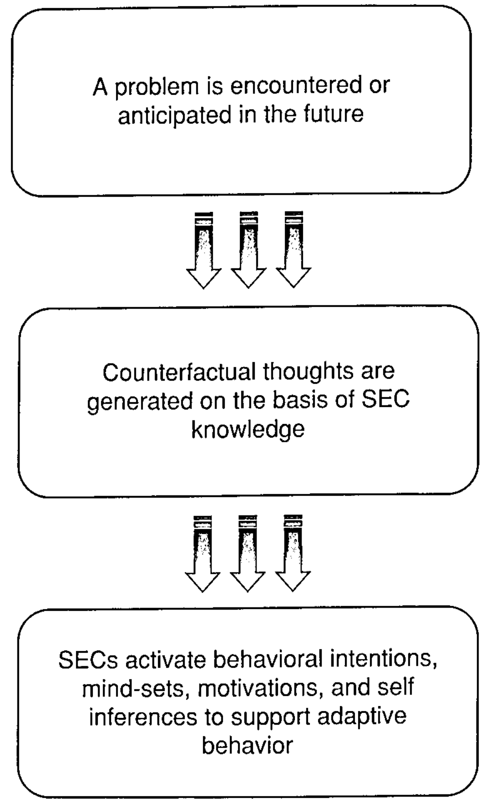

Figure 4.2 **A process model of the regulatory functions underlying counterfactual thought. SEC, structured event complex.**

Action Knowledge

One broad distinction represents counterfactual thought about *action versus inaction*, or the addition versus subtraction of an action from the present state (Roese, Hur, & Pennington, 1999). For example, the counterfactual inference that "She should never go out the night before an exam" represents the addition of an action ("going out"), whereas the inference that "He should always read the instructions carefully" represents the removal of an action ("reading carefully"). This form of counterfactual thought is central for evaluating consequences of carrying out or failing to perform specific actions (in the past or future).

Agent Knowledge

A second major category of counterfactual inference represents reasoning about the *self versus other* (Mandel, 2003). For example, the counterfactual inference that "Problems would be avoided if I attended the meeting" represents features of the self, whereas the inference that "Your skills would improve if you played more often" embodies features of others. Counterfactual reasoning about the self versus other provides the basis for adaptive social behavior and inferring the connection between specific mental states and particular patterns of behavior (e.g., "She would not have left early if she wanted to talk with you").

Outcome Knowledge

A third major category represents the comparison of a current outcome to a *better or worse alternative* (Roese & Olson, 1995). For example, the counterfactual inference that "She should accept the job with the higher salary" represents an upward inference about a better alternative, whereas the observation that "Other people with her qualifications earn much less than she does" represents a downward inference about a worse alternative. Counterfactual reasoning about better versus worse outcomes is critical for learning from the past and assessing alternative courses of action in the future.

The reviewed categories of counterfactual inference embody core features of SEC knowledge, enabling adaptive social behavior on the basis of actions, agents, and event outcomes.

Neuroscience Review

We review neuroscience findings in support of the SEC framework, providing evidence to confirm the representational role of the human PFC and to support the role of SECs in counterfactual thought. The representational aspects of SECs and their proposed localizations within the PFC are summarized in Figure 4.1 (for a review of further evidence in support of the SEC framework, see Barbey & Grafman, in press; Krueger et al., 2009).

Category Specificity of the Prefrontal Cortex

The subdivision of the PFC into neuroanatomically distinct regions designed to process specific forms of knowledge supports the proposal that SEC representations are stored within particular regions of the PFC on a content-specific basis (see Fig. 4.1). Converging evidence is provided by lesion studies demonstrating selective impairments for social and reward-related behavior in vmPFC lesion patients (Dimitrov, Phipps, Zahn, & Grafman, 1999; Milne & Grafman, 2001), and impairments for mechanistic planning in dlPFC patients (Burgess, Veitch, de Lacy Costello, & Shallice, 2000; Goel & Grafman, 2000).

Our research group conducted a PET study providing further evidence to support the representation of domain-specific SECs for *nonemotional* versus *emotional* event knowledge within the PFC (Partiot, Grafman, Sadato, Wachs, & Hallett, 1995). The employed nonemotional task asked subjects to "imagine silently the sequence of events and feelings concerned with preparation and dressing before (their) mother comes over for dinner." In contrast, subjects in the emotional task were asked to "imagine silently the sequence of events and feelings concerned with preparation and dressing to go to (their) mother's funeral." Consistent with the domain-specific predictions of the SEC framework, distinct patterns of neural activity were observed when subjects assessed nonemotional versus emotional scripts. Nonemotional scripts activated the right superior frontal gyrus (Brodmann's area [BA] 8), bilateral middle frontal gyri (BA 8 and 9), and medial frontal gyri (BA 6 and 10), whereas emotional scripts recruited the left anterior cingulate (BA 24 and 32), bilateral medial frontal gyri (BA 8 and 9), and anterior medial temporal lobe (BA 21).

Employing fMRI, we further demonstrated that *social* versus *nonsocial* SECs depend on a distinct representational topography within the PFC (Wood, Romero, Makale, & Grafman, 2003). We applied a modified go/no-go paradigm in which subjects classified individual words (e.g., "menu," "order") or phrases (e.g., "read the menu," "order the food") according to one of two focal categories (social versus nonsocial). Social activities recruited the left superior frontal gyri (BA 8 and 9), whereas nonsocial activities engaged the right superior frontal gyrus (BA 8), left medial frontal gyrus (BA 6), and the bilateral anterior cingulate (BA 25). Despite the large body of evidence to support the role of the orbitofrontal cortex (OFC) in social processing (Fuster, 1997; Milne & Grafman, 2001), activation in this region was not observed. Further inspection of the functional images demonstrated signal dropout in the OFC, limiting conclusions drawn concerning the role of this region in the storage of social SECs.

To further investigate this issue, we conducted a lesion study in which patients with PFC lesions and matched controls performed the classification task of Wood et al. (2003; Wood, Tierney, Bidwell, & Grafman, 2005). Subjects classified individual words (e.g., "menu," "order") or phrases (e.g., "read the menu," "order the food") as representing social versus nonsocial events. Patients with damage to the right OFC demonstrated cognitive impairments in the accessibility of script and semantic representations of social (rather than nonsocial) activities, providing evidence to support the role of the OFC in social processes.

In a subsequent functional magnetic resonance imaging (fMRI) study, we applied multidimensional scaling to assess the psychological structure of event knowledge and its neural representation within particular regions of the PFC (Wood, Knutson, & Grafman, 2005). Multidimensional scaling revealed three psychological dimensions underlying event knowledge (*engagement, social valence,* and *experience*). To investigate the neural correlates of the identified psychological dimensions, we conducted an fMRI experiment in which subjects classified each event according to whether it represented a social activity. Parametric analyses of event-related fMRI data were conducted to investigate brain regions whose activity was modulated by the three psychological components of event knowledge. The results demonstrated that the psychological structure of event knowledge is broadly organized along dimensions that are represented within distinct regions of the human PFC, with the *experience* dimension recruiting the medial PFC (BA 10), the *engagement* dimension activating the left OFC (BA 47), and the *social valence* dimension engaging the amygdala and right OFC (BA 11 and 47).

In summary, the reviewed studies provide evidence to support our proposal that category-specific SECs are stored within distinct regions of the PFC.

Structured Event Complexes and Counterfactual Thought

We propose that counterfactual thought depends on mental models represented in the form of SECs and review evidence demonstrating that SECs for counterfactual inference are functionally localized within distinct regions of the medial and lateral PFC.

Medial Prefrontal Cortex

Counterfactual reasoning is characterized by three major forms of inference that each recruit distinct regions of the mPFC (see Fig. 4.3). According to this framework, counterfactual thinking depends on category-specific SECs within the mPFC, which provide key representational elements within a larger network of anatomically connected prefrontal and posterior regions supporting counterfactual thought.

Action versus Inaction

According to the SEC framework, counterfactual reasoning about action versus inaction preferentially recruits the dorsomedial PFC (dmPFC). Several neuroscience studies have implicated the dmPFC in the continuous internal monitoring of action (Botvinick, Cohen, & Carter, 2004). Barch and colleagues (2001) report an extensive meta-analysis of functional imaging studies that included data from a broad range of action-monitoring tasks (e.g., involving the inhibition

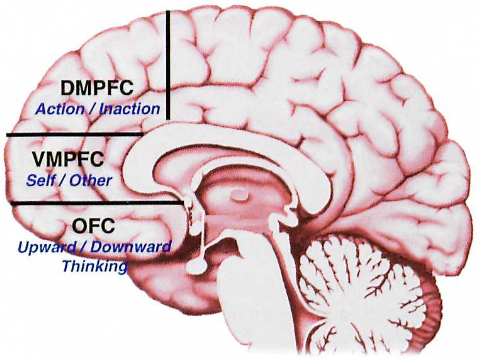

Figure 4.3 **Neural predictions of the SEC theory of counterfactual thought. The dorsomedial prefrontal cortex (DMPFC) represents counterfactual reasoning about *action versus inaction*, the ventromedial prefrontal cortex (VMPFC) represents counterfactual thoughts directed toward the *self versus other*, and the orbitofrontal cortex (OFC) represents *upward versus downward* counterfactual thinking.**

of prepotent responses) that recruited the dmPFC. Along the same lines, Walton, Devlin, and Rushworth (2004) observed activity in the dmPFC when participants monitored the outcome of self-selected actions. These finding suggest that the dmPFC is critical for monitoring the addition versus subtraction of actions for counterfactual reasoning.

Self versus Other

We propose that counterfactual thought involving the self versus others recruits the vmPFC. A large body of neuroscience evidence supports this proposal, demonstrating that the vmPFC represents components of self-knowledge, person knowledge, and mentalizing. Beldarrain et al. (2005) demonstrated impairments in self-generated counterfactual thought in vmPFC lesion patients. Converging evidence is provided by Macrae et al. (2004), who observed activation in the vmPFC when participants evaluated the self-relevance of specific personality traits. Ochsner et al. (2004) similarly found activation in the vmPFC when participants monitored their emotional states.

Recruitment of the vmPFC is also observed in studies that assess person knowledge more broadly (applying to others as well as the self). Mitchell, Heatherton, and Macrae (2002) reported activation in this region when participants judged whether a presented adjective applied to a person (rather than an inanimate object). Consistent with these findings, Schmitz et al. (2004) observed activation in the vmPFC when participants thought about themselves or a close friend.

Finally, extensive neuroscience evidence implicates the vmPFC in the process of representing another person's psychological perspective (i.e., "mentalizing"). For example, Fletcher et al. (1995) and Goel et al. (1995) reported activation in the vmPFC when participants read social scripts in which the psychological perspectives of fictional characters were inferred.

Upward versus Downward Thinking

We propose that counterfactual reasoning about upward (better) versus downward (worse) outcomes recruits the OFC, which is widely implicated in the processing of event outcomes associated with rewards or penalties. Elliott, Dolan, and Frith (2000) propose that the OFC is involved in monitoring reward and serves to guide behavior in terms of the value of possible outcomes. Walton et al. (2004) found that the activity in the OFC was elicited by the need to monitor the outcomes of externally guided actions. Similarly, Coricelli et al. (2005) found that activity in the OFC correlated with the amount of anticipated regret associated with a decision. In sum, the reviewed findings suggest that the OFC provides the basis for counterfactual reasoning about upward (better) versus downward (worse) outcomes.

Lateral Prefrontal Cortex

It is likely that mental models for goal-directed social behavior involving future thinking additionally recruit the lateral PFC, which represents behavior-guiding principles for counterfactual inference concerning obligatory, prohibited, and permissible courses of action. Emerging evidence from the social and decision neuroscience literatures demonstrates *(1)* the involvement of the vlPFC when reasoning about necessary (obligatory or prohibited) courses of action, *(2)* the recruitment of the dlPFC for drawing inferences about possible (permissible) states of affairs, and *(3)* activation in the alPFC for higher order inferences that incorporate both categories of knowledge (Fig. 4.4). The simulation architecture underlying these forms of inference further predicts the recruitment of broadly distributed neural systems, incorporating medial prefrontal and posterior knowledge networks representing modality-specific components of experience.

Ventrolateral Prefrontal Cortex

An increasing number of social neuroscience studies have shown that social norms for necessary (obligatory or prohibited) courses of action are represented by the vlPFC (BA 44, 45, and 47; Fig. 4.4b). Fiddick, Spampinato, and Grafman (2005) observed activity within bilateral vlPFC (BA 47) for social exchange reasoning, employing stimuli consisting primarily of social norms for obligatory and prohibited courses of action.

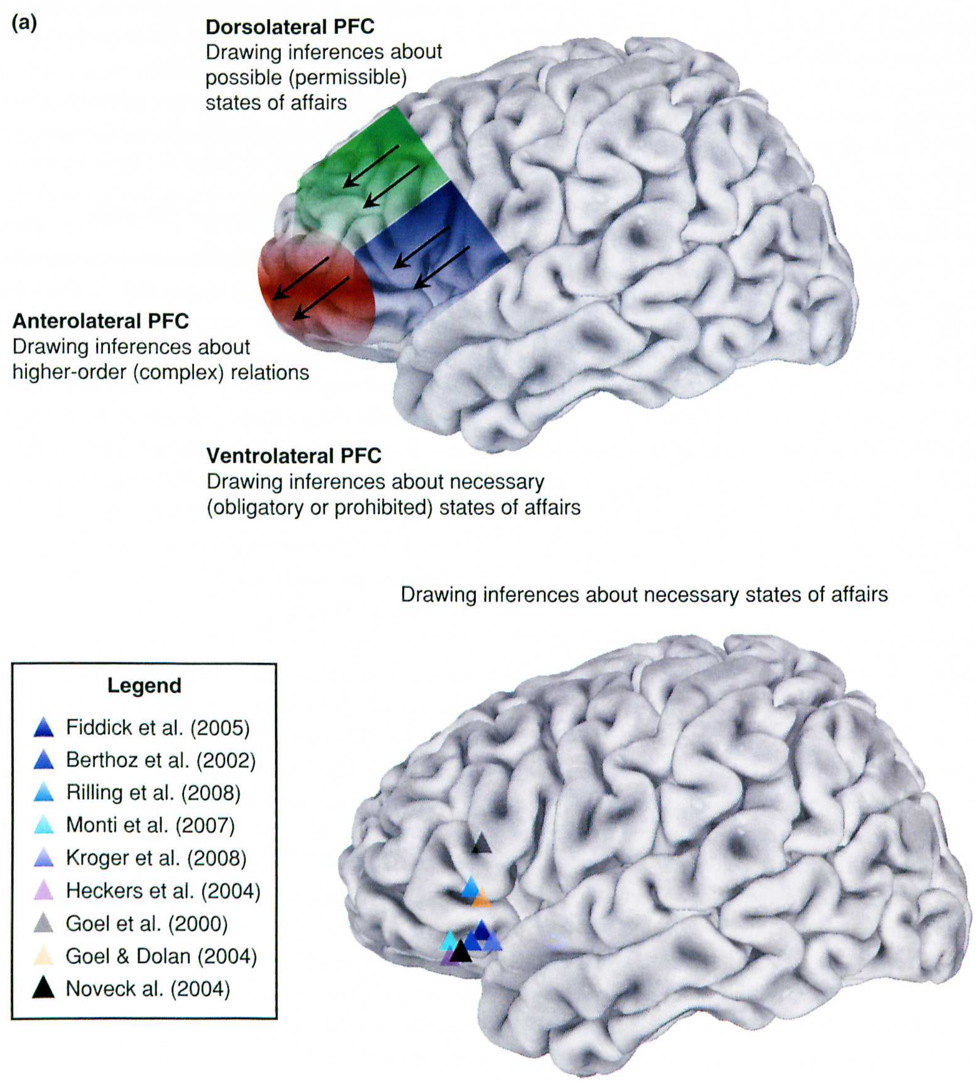

Figure 4.4 **An evolutionarily adaptive neural architecture for goal-directed social behavior. Panel *a* summarizes the functional organization of the lateral PFC, and panels *b*, *c*, and *d* illustrate supportive evidence.**

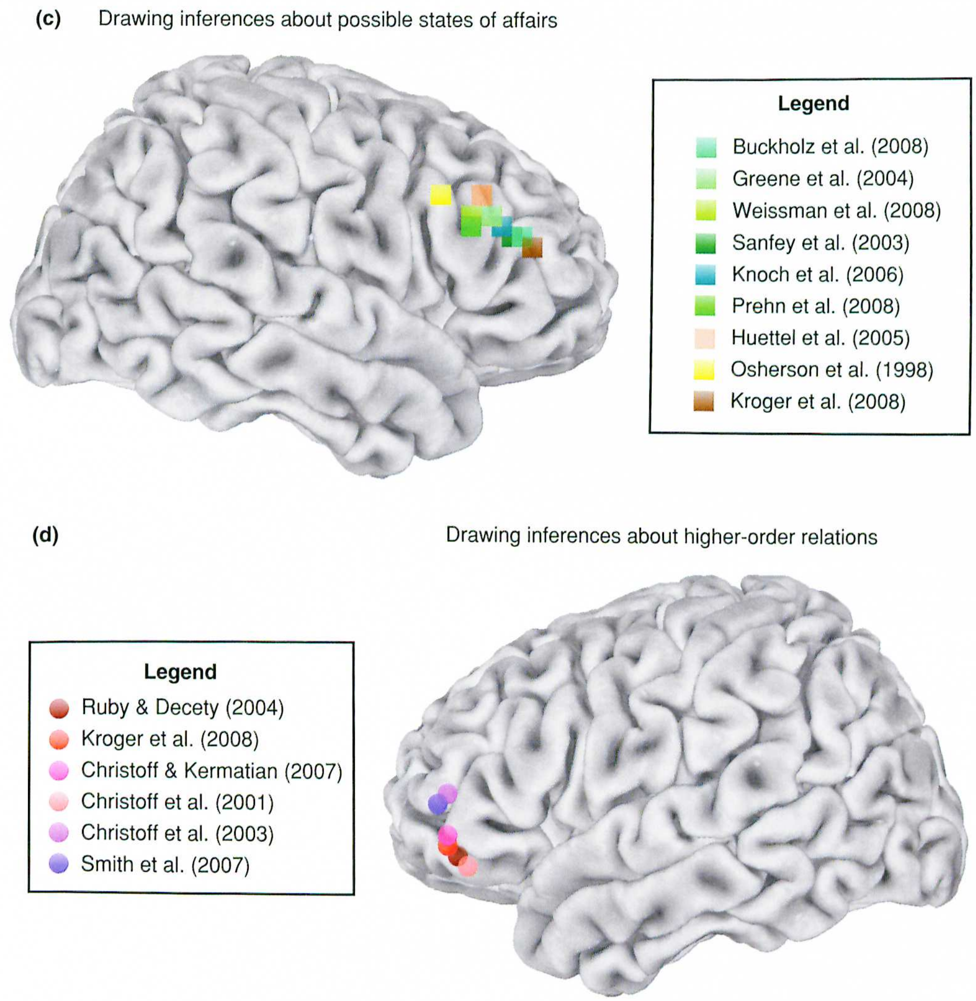

Figure 4.4 (*continued*)

Converging evidence is provided by Berthoz et al. (2002), who demonstrated recruitment of left vlPFC (BA 47) when participants detected violations of social norms stories representing obligatory and prohibited courses of action (e.g., the decision to "spit out food made by the host"). Similarly, Rilling et al. (2008) reported activation within left vlPFC (BA 47) when participants detected the violation of obligatory and prohibited norms of social exchange in a prisoner's dilemma game (i.e., the failure to cooperate).

The vlPFC is also involved when drawing conclusions that necessarily follow from the truth of the premises, that is, for *deductive inference*. Although a consensus has not yet been reached, an increasing number of studies report consistent findings when common sources of variability are controlled (regarding the linguistic content, linguistic complexity, and deductive complexity of reasoning problems). For example, a recent series of experiments by Monti et al. (2007) controlled for these sources of variability and provided evidence that the left vlPFC (BA 47) mediates representations of the logical structure of a deductive argument (e.g., If *P* or *Q*, then *Not-R*/*P*/Therefore, *Not-R*), supporting the representation of behavior-guiding principles for necessary forms of behavior within this region. Furthermore, a recent study by Kroger and colleagues (2008) controlled for the complexity and type of calculations that were performed and also observed activation within the left vlPFC (BA 44 and 45) for deductive reasoning (see also Heckers, Zalesak, Weiss, Ditman, & Titone, 2004). Converging evidence is provided by Goel and colleagues (Goel, Buchel, Frith, & Dolan, 2000; Goel & Dolan, 2004), who have consistently observed activation within the left vlPFC (BA 44 and 45) for deductive conclusions drawn from categorical syllogisms (e.g., All humans are mortal/Some animals are human/Therefore, some animals are mortal). Finally, Noveck, Goel, and Smith (2004) demonstrated recruitment of left vlPFC (BA 47) for drawing deductive conclusions from conditional statements (e.g., If *P* then *Q*/*P*/Therefore, *Q*), consistent with the role of this region for representing behavior-guiding principles in the form of a conditional. It is likely that such conditionals are utilized when charting future behavioral options.

Dorsolateral Prefrontal Cortex

Accumulating evidence demonstrates that the dlPFC (BA 46 and 9) represents behavior-guiding principles for evaluating the permissibility or fairness of observed behavior (Fig. 4.4c). An early study by Sanfey et al. (2003) reported activity within the right dlPFC (BA 46) when participants evaluated the fairness of an offer in an ultimatum game. Knoch et al. (2006) further demonstrated that deactivating this region with repetitive transcranial magnetic stimulation reduced participants' ability to reject unfair offers in the ultimatum game, suggesting that the dlPFC is central for guiding *future* behavior based on evaluations of fairness and permissibility. Converging evidence is provided by Buckholtz et al. (2008), who observed activity within the right dlPFC (BA 46) when participants assigned responsibility for crimes and made judgments about appropriate (e.g., equitable or fair) forms of punishment in a legal decision-making task. The work of Greene et al. (2004) further suggests that this region is involved in normative evaluations involving conflicting moral goals. These authors employed moral scenarios similar to the famous trolley problem (Thomson, 1976) and assessed trials in which participants acted in the interest of greater aggregate welfare at the expense of personal moral standards. This contrast revealed reliable activation within the right dlPFC (BA 46), suggesting that this region is critical for evaluating the permissibility or fairness of behaviors that conflict with personal moral standards (for additional evidence, see Prehn et al., 2008; Weissman, Perkins, & Woldorff, 2008). Applying moral standards might be relevant if future planning demands a choice from among alternative courses of social action.

In contrast to deductive inference, conclusions about possible courses of action reflect uncertainty concerning the actions that "should" be taken and/or the consequences that "might" follow; these are referred to as *inductive inferences*. Volz et al. (2004) found that activation within the right dlPFC (BA 9) increased parametrically with the degree of uncertainty held by the participant (see also Huettel, Song, & McCarthy, 2005). Furthermore, Osherson et al. (1998) observed preferential recruitment of the right dlPFC (BA 46) when performance on an inductive reasoning task was directly compared to a matched deductive inference task, supporting the role of this region for reasoning about possible (rather than necessary) states of affairs.

Anterolateral Prefrontal Cortex

The alPFC (BA 10 and 11)—and the orbitofrontal cortex (OFC) more broadly—is central for social cognition (Fig. 4.4d). Studies of patients with lesions confined to the OFC have reported impairments in a wide range of social functions, including the regulation and control of social responses, the perception and integration of social cues, and perspective taking (Bechara et al., 2000; LoPresti et al., 2008; Rolls et al., 1994). Stone et al. (2002) further demonstrate that patients with orbitofrontal damage produced selective impairments in reasoning about social contracts, supporting the proposed role of the PFC in social exchange. Bechara et al. (2000) observed profound deficits in the ability of orbitofrontal patients to represent and integrate social and emotional knowledge in the service of decision making. Converging evidence is provided by LoPresti et al. (2008), who demonstrated that the left alPFC (BA 11) mediates the integration of multiple social cues (i.e., emotional expression and personal identity), further suggesting that this region supports the integration of multiple classes of social knowledge. Other fMRI evidence was provided by Moll and de Oliveira-Souza (2006), who reported bilateral recruitment of the OFC (BA 11) during a social decision-making task when participants had to evaluate the social contributions of a charitable organization and chose not to make a donation.

Progressively anterior subregions of the lateral PFC (BA 10 and 11) have also, more generally, been associated with higher order processing requirements for thought and action (Badre, 2008; Botvinick, 2008; Koechlin & Summerfield, 2007). Ramnani and Owen (2004) reviewed contemporary research and theory investigating the cognitive functions of the alPFC, concluding that this region is central for integrating the outcomes of multiple cognitive operations, consistent with the predicted role of the alPFC for representing higher order inferences that incorporate both necessary and possible states of affairs (for representative findings, see Christoff & Kermatian, 2007; Christoff et al., 2001; Christoff, Ream, Geddes, & Gabrieli, 2003; Kroger et al., 2008; Smith, Keramatian, & Christoff, 2007). Although future planning may simply require the application of an overlearned ritual or routine, it can also require the on-line explicit construction of an action series, albeit with different degrees of social involvement. In thinking about (or simulating) the future, the typical thought process would frequently include both counterfactual thinking and obligatory/permissive conditions that would eventually be integrated into the construction of the structured event complex used in the future to execute an activity.

Conclusion

We have introduced a "representational" theory of PFC function in accord with the structure, neurophysiology, and connectivity of the PFC, and the modern cognitive neuroscience view that elements of knowledge are represented within functionally localized brain regions. The reviewed evidence in support of the SEC framework confirms the importance and uniqueness of the human PFC for representing knowledge in the form of cognitive events and action sequences.

We have further advocated for the representational basis of SECs in counterfactual thought, reviewing evidence to support the role of specific regions of the medial and lateral PFC in the representation of particular forms of counterfactual inference. According to this framework, SECs in the mPFC represent components of event knowledge (agents, objects, actions, mental states, and background settings) that are essential for constructing mental models of past or future events and assessing the consequences of alternative courses of action (Fig. 4.3). We have also surveyed a broad range of neuroscience evidence demonstrating that the lateral PFC mediates behavior-guiding principles for specific classes of counterfactual inference, with the vlPFC recruited when drawing inferences about necessary (obligatory or prohibited) courses of action, engagement of the dlPFC when reasoning about possible (permissible) behavior, and the alPFC recruited when both categories of inference are utilized (Fig. 4.4).

Our findings underscore the importance of SECs for high-level cognition more broadly, supporting their role in the construction of mental models and the simulation of alternative possibilities for learning from past experience (Byrne, 1997), for planning and predicting future events (Barbey & Sloman, 2007; Brase & Barbey, 2006), for creativity and insight (Costello & Keane, 2000; Sternberg & Gastel, 1989; Thomas, 1999), and for adaptive social behavior (e.g., supported by regulatory mechanisms based on representations of guilt, regret, and blame; Davis et al., 1995; Landman, 1987; Miller & Turnbull, 1990; Niedenthal et al., 1994; Zeelenberg et al., 1998).

In conclusion, we believe SECs are the key to understanding the human ability to represent mental models of events, which guide the selection of goal-directed action sequences and the on-line updating of behavior based on past outcomes or anticipated future events. When stored as memories, SECs provide a link between past, current, and future activities, enabling explanatory and predictive inferences that enable adaptive behavior and issue significant advantages for our species. Our review demonstrates that there is now substantial evidence to suggest that studying the nature of SEC representations is a competitive and promising way to characterize the components of event knowledge stored within the human PFC. Although SECs must coordinate with representations and processes stored in other brain regions—requiring hippocampal and related structure binding processes for the sense of an episode to emerge in consciousness—the elusive scientific characterization of knowledge stored within the PFC remains the key missing part of the puzzle. We believe that the evidence collected so far has brought us one step closer to such an understanding of the contribution of the PFC to future planning.

Acknowledgment

The authors are supported by the NINDS Intramural Research Program.

References

Barch, D. M., Braver, T. S., Akbudak, E., Conturo, T., Ollinger, J., & Snyder, A. (2001). Anterior cingulate cortex and response conflict: Effects of response modality and processing domain. *Cerebral Cortex* 11, 837–848.

Barbey, A. K., & Barsalou, L. W. (in press). Reasoning and problem solving: models. In L. Squire, T. Albright, F. Bloom, F. Gage, & N. Spitzer (Eds.), *New encyclopedia of neuroscience*. Vol. 8, (pp. 35–43). Oxford, England: Elsevier.

Barbey, A. K., & Sloman, S. A. (2007). Base-rate respect: From ecological rationality to dual processes. *Behavioral and Brain Sciences*, *30*, 241–297.

Barbey, A. K., & Wolff, P. (2006). Causal reasoning from forces. *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 2439). Mahwah, NJ: Erlbaum.

Barbey, A. K., & Wolff, P. (2007). Learning causal structure from reasoning. *Proceedings of the 29th Annual Conference of the Cognitive Science Society* (pp. 713–718). Mahwah, NJ: Erlbaum.

Barbey, A. K., & Wolff, P. (submitted). Composing causal relations in force dynamics.

Barsalou, L. W. (1991). Deriving categories to achieve goals. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 27, pp. 1–64). San Diego, CA: Academic Press.

Barsalou, L. W., Barbey, A. K., Simmons, W. K., & Santos, A. (2005). Embodiment in religious knowledge. *Journal of Cognition and Culture*, 5, 14–57.

Barsalou, L. W., Niedenthal, P. M., Barbey, A. K., & Ruppert, J. (2003). Social embodiment. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 43, pp. 43–92). San Diego, CA: Academic Press.

Barsalou, L. W., Simmons, W. K., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*, *7*, 84–91.

Barsalou, L. W., & Wiemer-Hastings, K. (2005). Situating abstract concepts. In D. Pecher & R. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thought* (pp. 129–163). New York: Cambridge University Press.

Beldarrain, M., Garcia-Monco, J., Astigarraga, E., Gonzalez, A., & Grafman, J. (2005). Only spontaneous counterfactual thinking is impaired in patients with prefrontal cortex lesions. *Cognitive Brain Research*, *24*, 723–726.

Berthoz, S., Armony, J. L., Blair, R. J. R., & Dolan, R. J. (2002). An fMRI study of intentional and unintentional (embarrassing) violations of social norms. *Brain*, 125, 1696–1708.

Bodner, M., Kroger, J., & Fuster, J. M. (1996). Auditory memory cells in dorsolateral prefrontal cortex. *Neuroreport*, *7*, 1905–1908.

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences*, *8*, 539–546.

Brase, G., & Barbey, A. K. (2006). Mental representations of statistical information. In A. Columbus (Ed.), *Advances in psychological research* (Vol. 41, pp. 91–113). New York: Nova Science Publishers.

Buckholtz, J. W., Asplund, C. L., Dux, P. E., Zald, D. H., Gore, J. C., Jones, O. D., & Marois, R. (2008). The neural correlates of third-party punishment. *Neuron*, *60*, 930–940.

Burgess, P. W., Veitch, E., de Lacy Costello, A., & Shallice, T. (2000). The cognitive and neuroanatomical correlates of multitasking. *Neuropsychologia*, *38*, 848–863.

Byrne, R. M. J. (1997). Cognitive processes in counterfactual thinking about what might have been. In D. Medin (Ed.), *The psychology of learning and motivation, advances in research and theory* (Vol. 37, pp. 105–154). San Diego, CA: Academic Press.

Byrne, R. M. J. (2002). Mental models and counterfactual thinking. *Trends in Cognitive Sciences, 6*, 405–445.

Chaigneau, S., & Barbey, A. K. (2008). Assessing psychological theories of causal meaning and inference. In *Proceedings of the Thirtieth Annual Science Society*. Hillsdale, NJ: Erlbaum.

Christoff, K., & Keramatian, K. (2007). Abstraction of mental representations: Theoretical considerations and neuroscientific evidence. In S. A. Bunge & J. D. Wallis (Eds.), *The neuroscience of rule-guided behavior*. New York: Oxford University Press.

Christoff, K., Prabhakaran, V., Dorfman, J., Zhao, Z., Kroger, J. K., Holyoak, K. J., & Gabrieli, J. D. E. (2001). Rostrolateral prefrontal cortex involvement in relational integration during reasoning. *NeuroImage, 14*, 1136–1149.

Christoff, K., Ream, J. M., Geddes, L. P. T., & Gabrieli, J. D. E. (2003). Evaluating self-generated information: Anterior prefrontal contributions to human cognition. *Behavioral Neuroscience, 117*, 1161–1168.

Coricelli, G., Critchley, H. D., Joffily, M., O'Doherty, J. P., Sirigu, A., & Dolan, R. J. (2005). Regret and its avoidance: A neuroimaging study of choice behavior. *Nature Neuroscience, 8*, 1255–1262.

Costello, F. J., & Keane, M. T. (2000) Efficient creativity: Constraint guided conceptual combination. *Cognitive Science, 24*, 299–349.

Damasio, A. R. (1989). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition, 33*, 25–62.

Davis, C. J., Lehman, D. R., Wortman, C. B., Silver, R. C., & Thompson, S. C. (1995). The undoing of traumatic life events. *Personality and Social Psychology Bulletin, 21*, 109–124.

De Dreu, C. K. W., Beersma, B., Stroebe, K., & Euwema, M. C. (2006). Motivated information processing, strategic choice, and the quality of negotiated agreement. *Journal of Personality and Social Psychology, 6*, 927–943.

Dimitrov M., Phipps M., Zahn T., & Grafman J. (1999). A thoroughly modern Gage. *Neurocase, 5*, 345–354.

Dunning, D. (1999). A newer look: Motivated social cognition and the schematic representation of social concepts. *Psychological Inquiry, 10*, 1–11.

Elliott, R., Dolan, R. J., & Frith, C. D. (2000). Dissociable functions in the medial and lateral orbitofrontal cortex: Evidence from human neuroimaging studies. *Cerebral Cortex, 10*, 308–317.

Elston, G. N. (2000). Pyramidal cells of the frontal lobe: All the more spinous to think with. *Journal of Neuroscience, 20*, RC95.

Engel, A. K., & Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. *Trends in Cognitive Sciences, 5*, 16–25.

Fiddick, L., Spampinato, M. V., & Grafman, J. (2005). Social contracts and precautions activate different neurological systems: An fMRI investigation of deontic reasoning. *NeuroImage, 28*, 778–786.

Fletcher, P. C., Happe, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S. J., & Frith, C. D. (1995). Other minds in the brain: A functional imaging study of "theory of mind" in story comprehension. *Cognition, 57*, 109–128.

Fortin, N. J., Agster, K., & Eichenbaum, H. B. (2002). Critical role of the hippocampus in memory for sequences of events. *Nature Neuroscience, 5*, 458–562.

Fuster, J. M. (1997). *The prefrontal cortex: Anatomy, physiology, and neuropsychology of the frontal lobe*. New York: Raven.

Fuster, J. M., & Alexander, G. E. (1971). Neuron activity related to short-term memory. *Science, 173*, 652–654.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain, 119*, 593–609.

Gilbar, O., & Hevroni, A. (2007). Counterfactuals, coping strategies and psychological distress among breast cancer patients. *Anxiety, Stress, and Coping, 20*, 382–392.

Goel, V., Buchel, C., Frith, C., & Dolan, R. (2000). Dissociation of mechanisms underlying syllogistic reasoning. *NeuroImage, 12*, 504–514.

Goel, V., & Dolan, R. J. (2004). Differential involvement of left prefrontal cortex in inductive and deductive reasoning. *Cognition, 93*, B109–B121.

Goel, V., & Grafman, J. (2000). The role of the right prefrontal cortex in ill-structured problem sovling. *Cognitive Neuropsychology, 17*, 415–436.

Goel, V., Grafman, J., Sadato, N., & Hallett, M. (1995). Modeling other minds. *Neuroreport, 6*, 1741–1746.

Grafton, S. T., Arbib, M. A., Fadiga, L., & Rizzolatti, G. (1996). Localization of grasp representations in humans by PET: II. Observation compared with imagination. *Experimental Brain Research, 112*, 103–111.

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron, 44*, 389–400.

Griffin, D. W., & Ross, L. (1991). Subjective construal, social inference, and human misunderstanding. *Advances in Experimental Social Psychology, 24*, 319–359.

Heckers, S., Zalesak, M., Weiss, A. P., Ditman, T., & Titone, D. (2004). Hippocampal activation during transitive inference in humans. *Hippocampus, 14*, 153–162.

Huettel, S. A., Song, A. W., & McCarthy, G. (2005). Decisions under uncertainty: Probabilistic context influences activation of prefrontal and parietal cortices. *The Journal of Neuroscience, 25*, 3304–3311.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93*, 136–153.

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., & Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science, 314*, 829–832.

Kroger, J. K., Nystrom, L. E., Cohen, J. D., & Johnson-Laird, P. N. (2008). Distinct neural substrates for deductive and mathematical processing. *Brain Research, 1243*, 86–103.

Krueger, F., Barbey, A. K., & Grafman, J. (in press). The medial prefrontal cortex represents social event knowledge. *Trends in Cognitive Sciences.*

Levy, R., & Goldman-Rakic, P. S. (2000). Segregation of working memory functions within the dorsolateral prefrontal cortex. *Experimental Brain Research, 133*, 23–32.

Macrae, C. N., Moran, J. M., Heatherton, T. F., Banfield, J. F., & Kelley, W. M. (2004). Medial prefrontal activity predicts memory for self. *Cerebral Cortex, 14*, 647–654.

Mandel, D. R. (2003). Counterfactuals, emotions, and context. *Cognition and Emotion, 17*, 139–159.

Miller, D. T., & Turnbull, W. (1990). The counterfactual fallacy: Confusing what might have been with what ought to have been. *Social Justice Research, 4*, 1–19.

Milne, E., & Grafman, J. (2001). Ventromedial prefrontal cortex lesions in humans eliminate implicit gender stereotyping. *Journal of Neuroscience, 21*, RC150 (1–6).

Mitchell, J. P., Heatherton, T. F., & Macrae, C. N. (2002). Distinct neural systems subserve person and object knowledge. *Proceedings of the National Academy of Sciences USA, 99*, 15238–15243.

Moll, J., & de Oliveira-Souza, R. (2007). Moral judgments, emotions, and the utilitarian brain. *Trends in Cognitive Sciences, 11*, 319–321.

Monti, M. M., Osherson, D. N., Martinez, M. J., & Parsons, L. M. (2007). Functional neuroanatomy of deductive inference: A language-independent distributed network. *NeuroImage, 37*, 1005–1016.

Niedenthal, P. M., Tangney, J. P., & Gavanski, I. (1994). "If only I weren't" versus "If only I hadn't": Distinguishing shame and guilt in counterfactual thinking. *Journal of Personality and Social Psychology, 67*, 585–595.

Noveck, I. A., Goel, V., & Smith, K. W. (2004). The neural basis of conditional reasoning with arbitrary content. *Cortex, 40*, 613–622.

Ochsner, K. N., Knierim, K., Ludlow, D. H., Hanelin, J., Ramachandran, T., Glover, G., & Mackey, S. C. (2004). Reflecting upon feelings: An fMRI study of neural systems supporting the attribution of emotion to self and other. *Journal of Cognitive Neuroscience, 16*, 1746–1772.

O'Reilly, R. C., & Rudy, J. W. (2000). Computational principles of learning in the neocortex and hippocampus. *Hippocampus, 10*, 389–397.

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

Palmer, S. E. (1999) *Vision science: Photons to phenomenology*. Cambridge, MA: MIT Press.

Ramnani, N., & Owen, A. M. (2004). Anterior prefrontal cortex: insights into function from anatomy and neuroimaging. *Nature Reviews Neuroscience*, 5, 184–194.

Partiot, A., Grafman, J., Sadato, N., Wachs, J., & Hallett, M. (1995). Brain activation during the generation of non-emotional and emotional plans. *Neuroreport, 6*, 1397–1400.

Patterson, R., & Barbey, A. K. (in press). Causal simulation theory: An integrative cognitive neuroscience framework for causal reasoning. In J. Grafman & F. Krueger (Eds.), *Neural basis of belief systems*. New York: Psychology Press.

Prehn, K., Wartenburger, I., Meriau, K., Scheibe, C., Goodenough, O. R., Villringer, A., van der Meer, E., & Heekeren, H. R. (2008). Individual differences in moral judgment competence influence neural correlates of socio-normative judgments.