BEHAVIORAL AND BRAIN SCIENCES (2007) **30**, 439–480
Printed in the United States of America
doi: 10.1017/S0140525X07002579

Précis of *The Rational Imagination: How People Create Alternatives to Reality*

Ruth M. J. Byrne
School of Psychology and Institute of Neuroscience, Trinity College Dublin, University of Dublin, Dublin 2, Ireland
rmbyrne@tcd.ie
http://www.tcd.ie/Psychology/Ruth_Byrne/

Abstract: The human imagination remains one of the last uncharted terrains of the mind. People often imagine how events might have turned out "if only" something had been different. The "fault lines" of reality, those aspects more readily changed, indicate that counterfactual thoughts are guided by the same principles as rational thoughts. In the past, rationality and imagination have been viewed as opposites. But research has shown that rational thought is more imaginative than cognitive scientists had supposed. In *The Rational Imagination*, I argue that imaginative thought is more rational than scientists have imagined. People exhibit remarkable similarities in the sorts of things they change in their mental representation of reality when they imagine how the facts could have turned out differently. For example, they tend to imagine alternatives to actions rather than inactions, events within their control rather than those beyond their control, and socially unacceptable events rather than acceptable ones. Their thoughts about how an event might have turned out differently lead them to judge that a strong causal relation exists between an antecedent event and the outcome, and their thoughts about how an event might have turned out the same lead them to judge that a weaker causal relation exists. In a simple temporal sequence, people tend to imagine alternatives to the most recent event. The central claim in the book is that counterfactual thoughts are organised along the same principles as rational thought. The idea that the counterfactual imagination is rational depends on three steps: (1) humans are capable of rational thought; (2) they make inferences by thinking about possibilities; and (3) their counterfactual thoughts rely on thinking about possibilities, just as rational thoughts do. The sorts of possibilities that people envisage explain the mutability of certain aspects of mental representations and the immutability of other aspects.

Keywords: conditional; counterfactual; creativity; deduction; "if only" thoughts; imagination; rationality; reasoning; simulation

1. The counterfactual imagination

In 1958 Martin Luther King, Jr. was stabbed and almost died. A decade later he remarked:

> The tip of the blade was on the edge of my aorta . . . It came out in the *New York Times* the next morning that if I had sneezed I would have died . . . And I want to say tonight, I want to say that I'm happy I didn't sneeze. Because if I had sneezed, I wouldn't have been around here in 1960 when students from all over the South started sitting-in at lunch counters . . . If I had sneezed I wouldn't have had the chance later that year in August to try to tell America about a dream that I had had . . . I'm so happy that I didn't sneeze. (King 1968)

In *The Rational Imagination: How People Create Alternatives to Reality* (Byrne 2005), I offer an explanation for how the mind creates such "if only . . ." thoughts. Thoughts about what might have been can seem irresistible in daily life. They emerge at a very young age and they seem to exist in most cultures. I suggest that their explanation relies on the idea that imaginative thought and rational thought share a lot in common.

There are surprising similarities in what everyone imagines. Some aspects of reality that people mentally represent seem to be "mutable," that is, they are readily "undone" in a mental simulation (Kahneman & Miller 1986). The remarkable regularities in what most people change suggest that there are "fault-lines" in reality (Kahneman & Tversky 1982). Consider a scenario in which an individual, Paul, dies in a car accident on his way home from work. He left work at his regular time, although sometimes he would leave early to take care of chores. He did not drive home by his regular route, but instead drove along the shore to enjoy the view. The accident occurred when a truck charged through an intersection. Paul's family often said "if only . . ." during the days that followed the accident. How do you think they continued this sentence? Most participants who were told that Paul did not drive home by his regular route tended to think his family would say "if only Paul had driven home by his regular

RUTH BYRNE is Professor of Cognitive Science at the Institute of Neuroscience and School of Psychology at Trinity College Dublin, University of Dublin. She is also the Vice Provost of Trinity College Dublin. Byrne is the author of more than 80 scientific publications and has co-authored or co-edited four books. Her research examines the nature of human reasoning and the creative imagination of counterfactual alternatives to reality.

 0140-525X/08 $40.00

route" (Kahneman & Tversky 1982). Other participants who were told that Paul had driven home by his regular route but had left the office earlier than usual tended to complete the sentence by saying "if only Paul had left at his regular time." Thoughts about what might have been also seem to amplify certain emotions, such as regret, guilt, shame, relief, hope, and anticipation. The emotion seems to depend on a comparison between how the event actually turned out and how it could have or should have turned out differently. The same is true for social attributions of culpability such as blame, responsibility, and fault (Mandel 2003a; Roese & Olson 1995).

Why do people imagine alternatives to some aspects of reality more readily than to other ones? This question is the key to understanding how the counterfactual imagination works. My aim in *The Rational Imagination* is to answer it. I focus on the perceived fault-lines in reality, which may correspond to core categories of mental life, such as space, time, cause, and intention (Byrne 1997; Miller & Johnson-Laird 1976). For example, people tend to imagine alternatives to actions more than they do to inactions (Kahneman & Tversky 1982). They tend to imagine alternatives to events that are within their control more than to those outside their control (Girotto et al. 1991). Their thoughts about how an event might have turned out differently lead them to judge that a strong causal relation exists between an antecedent and the outcome (Wells & Gavanski 1989). And their thoughts about how an event might have turned out the same lead them to judge that a weak causal relation exists between the antecedent and the outcome (Boninger et al. 1994). In a simple, non-causal temporal sequence, their thoughts about how events could have turned out differently focus on the most recent event rather than on earlier ones (Miller & Gunasegaram 1990).

We gain an important glimpse of the counterfactual imagination when we look at the things that people focus on when they create alternatives to reality. Most people focus on the perceived fault-lines of reality and in Chapter 1 of the book I sketch these fault-lines. We gain an equally important glimpse when we look at the things that people do *not* focus on and I also sketch these characteristics in Chapter 1. People do not tend to create "miracle-world" counterfactuals, such as, for example, "if the Romans had had machine guns . . ." (McMullen & Markman 2002). They do not mentally alter natural laws; for example, "she would not have fallen if there had not been gravity" (Seeleu et al. 1995). They do not tend to focus on impossibilities; for example, "if Kangaroos had no tails they would topple over" (Lewis 1973). They do not tend to imagine "remote" counterfactual alternatives; instead, they focus on counterfactuals that are close to the current facts, such as a runner who just misses first place in a race. They judge "close counterfactuals" to be plausible (Tetlock & Parker 2005). People also tend to judge a counterfactual to be plausible when it is consistent with their beliefs. Consider the assertion: "'If Kennedy had listened to his Hawk advisers, he would have engaged in a nuclear strike during the Cuban missile crisis." Experts who believe that nuclear weapons could be deployed judge the assertion to be more plausible than experts who believe they would never be used (Tetlock & Lebow 2001). Most people tend to imagine plausible counterfactual alternatives.

Most people do not imagine how the world might be different after every event: they tend to think "if only" after something bad happens (Roese 1997). People can think about how things might have been better or worse. For example, after an argument with a friend they can think about a better outcome – "if I had not said such harsh things, we would still be friends" – which might help them to learn from their mistakes and prepare for the future (Roese 1994). But such thoughts can make them feel bad, so, instead, they may think about a worse outcome – "if I had told her how I felt, she would never have forgiven me" or "at least I did not say anything cruel" – which may help them to feel good (McMullen & Markman 2000). However, these kinds of thoughts can encourage complacency. People can also think about how things might have turned out exactly the same: for example, "even if I had apologised to her, she would not have forgiven me." In other words, they can exert some control over whether they imagine alternatives that are better, worse, or the same. Their choice may depend on their motivation – for example, to console a victim (Seeleu et al. 1995) – or on their goals (Roese et al. 2005).

Counterfactual thoughts are central to everyday mental life. But some people can become plagued by thoughts about what might have been. Individuals who have experienced the grief of the death of a spouse or child are sometimes haunted by thoughts of how they could have prevented the death, for example, "if only I had checked on him more often during the night." Of course, everyone thinks about what might have been after a traumatic life event. But those people who continue to think about what might have been experience greater distress (Davis et al. 1995). Counterfactual thoughts may be implicated in clinical disorders such as depression and anxiety (Roese & Olson 1995). An understanding of how people imagine alternatives to reality may advance ways to help people for whom the counterfactual imagination has become dysfunctional.

The book is organised around the fault-lines of reality; each chapter examines one of them. I start by sketching the characteristics of the counterfactual imagination in Chapter 1 and then I examine its relation to rational thought in Chapter 2. In Chapter 3, I offer an explanation of the tendency to think "if only" about actions more than inactions, and in Chapter 4 I discuss the way the counterfactual imagination deals with forbidden actions. Chapter 5 addresses the relation of "if only" thoughts to causal thoughts and Chapter 6 examines the other side of the causal coin, "even if" thoughts. Chapter 7 deals with the tendency to imagine alternatives to the most recent event in a temporal sequence. In Chapter 8, I sketch two implications of this view of the counterfactual imagination: one implication for the nature of individual differences, and the other for understanding creative thought in general. The final chapter, Chapter 9, considers the consequences of the idea of a rational imagination.

2. Imagination and rational thought

In the past, rationality and imagination have been viewed as opposites, sharing little in common. Logical thought and creativity have even been considered to be mutually exclusive. Contrary to this viewpoint, the argument I make

throughout *The Rational Imagination* is that rationality and imagination share a lot in common. In Chapter 2, I focus on one view of the nature of human rationality – that people envisage certain kinds of possibilities – and I sketch some of the principles that guide the possibilities that people think about when they make inferences. Later in the book I show how the same principles, and corollaries to them, explain how people create counterfactual alternatives to reality.

People try to think rationally in many situations. For example, you try to reach a rational conclusion when you estimate the likelihood that the next flight you take will be hijacked, or when you work out the risk that you may have contracted new variant Creutzfeldt-Jakob disease (CJD) from your exposures to meat infected by bovine spongiform encephalopathy (BSE). The idea that people are capable of rational thought has been debated (Gigerenzer & Selten 2001; Stanovich 1999).

Clearly, people can engage in hypothetical thought and it is an impressive achievement of human cognition. But most people make mistakes. Their mistakes have led some theorists to suggest that people do not have a general capacity for rational thought. Human reasoning may be based on an assortment of biases, rules of thumb, or "heuristics" (Evans 1989). Or it may depend on a fixed set of rules that govern a specific domain (Fiddick et al. 2000; Holyoak & Cheng 1995). Some theorists have suggested that people *do* have a general capacity for rational thought. Reasoning may depend on a "mental logic" of inference rules (Braine & O'Brien 1998; Rips 1994). Or it may depend on a simple semantic principle: An inference is valid if there are no counterexamples to it (Johnson-Laird & Byrne 1991). The formulation of a theory based on this last view has shed light on the imagination, and I outline it further in the next section.

2.1. Rational thought

Hypothetical thought has been studied in many guises and one good illustration is conditional reasoning. Conditionals such as "if Alicia went to the stables then she rode Starlight" have been considered to be the key to how people make suppositions and conjecture relationships between events (Johnson-Laird & Byrne 1991). One view is that the human ability to think rationally about hypothetical situations and conditional relations rests on the capacity to *imagine* possibilities (Johnson-Laird 1983). There are several principles that guide the possibilities that people consider when they understand a conditional (Johnson-Laird & Byrne 2002). The first principle is that people tend to think about true possibilities, such as "Alicia went to the stables and she rode Starlight," but they do not think about false possibilities, such as "Alicia went to the stables and she did not ride Starlight." The second principle is that initially people think about just a few of the possibilities. There are several possibilities that are consistent with the conditional, such as "Alicia did not go to the stables and she did not ride Starlight" and "Alicia did not go to the stables and she rode Starlight" (perhaps because the pony was in a nearby field). But usually people mentally represent the conditional by thinking about just a single true possibility. Multiple possibilities tend to exceed working memory capacity. Of course, knowledge can help people to think about more possibilities, it can help them to enrich those possibilities, and it can also eliminate possibilities (Johnson-Laird & Byrne 2002).

People can readily make some inferences based on the possibilities they have kept in mind, but they find other inferences more difficult to make. Suppose you know "if Mark left at 9 a.m. then he caught the airplane" and you are told, "Mark left at 9 a.m." What, if anything, do you think follows from these premises? Most people find it easy to make the *modus ponens* inference, "he caught the airplane." They have understood the conditional by envisaging the single true possibility, "Mark left at 9 a.m. and he caught the airplane." Now suppose you are told "Mark did not catch the airplane." What, if anything, follows from this premise and the conditional? Many people say nothing follows. They have difficulty making the inference to the conclusion, "he did not leave at 9 a.m." To make this *modus tollens* inference, they need to remember that there are alternative possibilities and they must think through what they are. As a result, most people find the *modus ponens* inference easier than the *modus tollens* one. On this account, people make mistakes, and can appear irrational because of the limits to the possibilities they can consider (Johnson-Laird & Byrne 1991). But they are rational, at least in principle, because they possess an underlying competence to think of all the relevant possibilities so that they can search for counterexamples.

2.2. Rational thought and imaginative thought

My claim in the book is that the principles for rational thought also guide imaginative thought. I suggest that a bridge from rationality to imagination can be built on counterfactual conditionals. Conditionals yield a good example of deductive rationality; counterfactual thoughts are a good example of everyday imagination. Counterfactual conditionals combine both rational and imaginative elements. Counterfactuals are special. Consider a counterfactual conditional in the subjunctive mood, "if Oswald had not killed Kennedy then someone else would have." It seems to mean something very different from a factual conditional in the indicative mood, "if Oswald did not kill Kennedy then someone else did" (Lewis 1973). Attempts to understand counterfactuals have led to important developments in linguistics (Athanasiadou & Dirven 1997), artificial intelligence (Ginsberg 1986), and philosophy (Lewis 1973). A counterfactual such as "if only he had been handsome, I would have married him" evokes two possibilities: an imagined possibility in which the man is handsome and the speaker marries him, and a reality in which the man was not and she did not. The way that people interpret such assertions goes beyond the simple truth of their components (Quine 1972). A counterfactual seems to presuppose that its antecedent is false, and so, according to a simple truth functional account, every counterfactual must be true. But people judge some counterfactuals to be plausible and others to be implausible. Philosophers have suggested instead that the truth of a counterfactual may depend on its truth in a "possible world" (Stalnaker 1968).

The critical psychological difference between factual and counterfactual conditionals may lie in the possibilities that people think about. Consider the conditional, "if Iraq

had weapons of mass destruction then the war was justified." It is consistent with several possibilities: Iraq had weapons of mass destruction and the war was justified; Iraq did not have weapons of mass destruction and the war was not justified; and Iraq did not have weapons of mass destruction and the war was justified (for other reasons). Most people mentally represent the factual conditional initially by thinking about just the single possibility that Iraq had weapons of mass destruction and the war was justified. Now consider the counterfactual conditional, "if Iraq had had weapons of mass destruction then the war would have been justified." It is consistent with essentially the same possibilities as the factual conditional. You might have understood the counterfactual by thinking about the conjecture, Iraq had weapons of mass destruction and the war was justified. But, you might also have thought about the facts that the speaker uttering the counterfactual is likely to have presupposed, that Iraq did not have weapons of mass destruction and the war was not justified. In other words, you might have thought about more than one possibility when you understood the counterfactual. The factual and counterfactual conditionals are consistent with essentially the same possibilities but a key difference between them is that people think about different possibilities when they understand them. The difference between factual and counterfactual conditionals illustrates a third principle underlying the possibilities that people think about: some ideas require people to think about dual possibilities, as Table 1 shows.

Of course another critical difference between factual and counterfactual conditionals is that the possibilities correspond to the putative facts for the factual conditional, whereas they correspond not only to putative facts but also to counterfactual possibilities for the counterfactual conditional. People keep track of the status of different possibilities. They think about two possibilities when they understand a counterfactual conditional, and they note one as the "facts" and the other as "imagined" possibilities (Johnson-Laird & Byrne 2002). A fourth principle is that even though people tend to keep in mind just true possibilities, they can think about what might have been because they can envisage possibilities that once were true but are so no longer. Often in everyday thoughts, people temporarily suppose something to be true even when they know it is false. They keep track of what is actually true and what is imagined. They can think about entirely imaginary situations, for example, to understand and create fantasy and fiction, in daydreams as well as in theatre, film, and literature. They can think about what is true in these imagined situations and what is false in them.

Table 1. *Summary of principles that guide the possibilities that people envisage*

1. **True possibilities:** People keep in mind true possibilities.
2. **Few possibilities:** People keep in mind few possibilities.
3. **Dual possibilities:** Some ideas require people to think about two possibilities.
4. **Counterfactual possibilities:** People think about possibilities that once may have been true possibilities but can be true no longer.
5. **Mutability of dual possibilities:** People readily imagine a counterfactual alternative to a possibility if it is mentally represented with a second possibility.
6. **Forbidden possibilities:** People think about the forbidden possibility as well as the permitted possibility when they understand an obligation.
7. **Temporal possibilities:** The possibilities people think about encode the temporal order of events in the world.

People tend to think about two possibilities for some ideas and this factor has significant consequences for how they imagine what might have been. It helps to explain why people mentally change only some aspects of their mental representation of reality. A basic principle of mutability is that an event can be readily changed if it brings to mind an alternative. For example, suppose you go to a concert and you sit in seat 426. You change seats after a while to seat 739 to get a better view. At the interval, an announcement is made that there is a prize of a trip around the world for the person seated in seat 426 (Johnson 1986). You may wish you had not changed seats. The prize seems to have been almost in your grasp. But suppose instead you had gone to the concert and you sat in seat 739 from the outset. At the interval, an announcement is made that there is a prize of a trip around the world for the person seated in seat 426. You may wish you had won, but you are unlikely to believe that the prize was almost in your grasp. In the first scenario, there is a ready-made alternative and people are influenced by the availability of alternatives. They may base their judgments on the ease with which instances can be brought to mind (Kahneman & Tversky 1982). The fifth principle is that people readily imagine a counterfactual alternative to a possibility if it is mentally represented from the outset with a second possibility. The possibility can be mutated easily; for example, it can be deleted from the mental representation and replaced with the second possibility as its counterfactual alternative.

The principles that guide the possibilities that people envisage when they reason help to explain how people imagine alternatives to reality. In Chapters 3 to 7 of *The Rational Imagination* I show how the application of these principles explains the regularities people exhibit in their imagination of counterfactual alternatives. I begin in Chapter 3 by showing how these principles explain why most people imagine counterfactual alternatives to an action more than to a failure to act in many situations.

3. Imagining how actions might have been different

Suppose you hear about a new disease, such as Severe Acute Respiratory Syndrome (SARS). It can be fatal and you suspect you could be exposed to it. But the vaccine can also have serious consequences. What will you decide to do? The decision is a difficult one. Many people choose to do nothing, even when the chances of death from a vaccine are less than death from a disease (Ritov & Baron 1990). Take a moment to imagine the families of the victims. Which families do you think are more likely to say, "if only . . ." – the families of individuals who died from the disease, or of individuals who died from

the vaccine? Most people think the families who will feel worse are the families of the individuals who died from the vaccine. The example illustrates that thoughts about what might have been often fall towards one side of the perceived fault-line between action and inaction. In Chapter 3, I first sketch some characteristics of the tendency shown by most people to focus on actions rather than failures to act when they imagine counterfactual alternatives to reality. Next, I outline some clues from reasoning that help identify why this perceived fault-line is observed. The clues come from studies of how people reason about counterfactual conditionals. Then, I show how these clues help to explain the tendency people have to focus on actions. I also explain an important exception to this tendency: situations in which people sometimes focus on failures to act.

People tend to regret actions that lead to a bad outcome more than inactions that lead to a bad outcome (Kahneman & Tversky 1982). Most people also judge that individuals feel better about their actions than their inactions when the outcome is good (Landman 1987). The tendency to focus on actions occurs in many everyday situations (Catellani & Milesi 2001; Zeelenberg et al. 1998a). To delete an action from a mental representation may require less mental effort than to add one (Dunning & Parpal 1989). Actions seem to be mentally "re-wound" in reminiscences and re-played differently (Hofstadter 1985). Of course, failures to act can be real causes of outcomes (Hart & Honore 1959). For example, the Herald of Free Enterprise passenger car ferry capsized off the Belgian port of Zeebrugge en route to the English port of Dover in 1987, ending on her side half submerged in shallow water. One hundred and eighty-eight people were killed; the worst death toll for a British ship in peacetime since the sinking of the Titanic in 1912. The cause of the tragedy was the failure of staff to close the bow doors (through which cars drove on to the ferry). However, in many circumstances people may tend not to view omission as a real cause (Ritov & Baron 1990). Actions seem to call to mind the alternatives of not acting, or acting differently, whereas inactions do not readily call to mind actions. Why do people focus on actions in their "if only . . ." thoughts? A clue to the answer comes from the study of counterfactual conditionals.

3.1. Clues from reasoning: Counterfactual conditionals

A key principle is that people think about some ideas by keeping in mind two possibilities. The two possibilities they envisage affect the way people think in many situations; for example, the two possibilities affect the way they reason about counterfactual conditionals. Suppose you are told: "if Mark had left at 9 a.m. then he would have caught the airplane." You may interpret the subjunctive conditional as a counterfactual and think initially about two possibilities: the conjecture, Mark left at 9 a.m. and he caught the airplane, and the presupposed facts, Mark did not leave at 9 a.m. and he did not catch the airplane. Suppose you then discover that Mark did not catch the airplane. What would you conclude? People are able to infer readily that Mark did not leave at 9 a.m. (Byrne & Tasso 1999). The *modus tollens* inference is difficult to make from a factual conditional "if Mark left at 9 a.m. then he caught the airplane," because people think about just a single possibility initially: Mark left at 9 a.m. and he caught the airplane. But as the theory predicts, the inference is easier to make from the counterfactual because of the enriched representation.

Suppose you discover that, in fact, Mark left at 9 a.m. What would you conclude from this premise and the counterfactual conditional? People are able to infer readily that Mark caught the airplane. These data for *modus tollens* and *modus ponens* support the suggestion that people think about both possibilities: the conjecture, Mark left at 9 a.m. and he caught the airplane, and also the presupposed facts, Mark did not leave at 9 a.m. and he did not catch the airplane. The evidence that people understand counterfactual conditionals by keeping in mind two possibilities provides an important clue to solve the puzzle of why people imagine counterfactual alternatives to their actions more than their inactions.

3.2. The rational imagination: Why people imagine alternatives to actions

People think about two possibilities when they understand an action. When you understand the decision to act you think about the situation before the action and the situation after the action. Consider the scenario in which two individuals, Mary and Laura, consider their stock options. Mary owns shares in company A. During the past year she considered switching to stock in company B, but she decided against it. She now finds out that she would have been better off by $1,000 if she had switched to the stock of company B. Laura owned shares in company B. During the past year she switched to stock in company A. She now finds out that she would have been better off by $1,000 if she had kept her stock in company B (Kahneman & Tversky 1982). When people think about Laura's action they think about both the pre-action situation, Laura owned shares in company B, and the post-action situation, Laura owns shares in company A:

> Laura owned shares in company B (pre-action)
> Laura owns shares in company A (post-action)

However, when they think about Mary's inaction, they think only about one situation – the shares that she owns in company A:

> Mary owns shares in company A (pre-decision and post-decision)

They do not need to think about two possibilities for an inaction – the situation before the decision and the situation after the decision – because these two situations are the same. An important corollary to the principles that guide the possibilities that people think about is that people keep in mind two possibilities when they understand an action. Inactions are mentally represented more economically than actions (Byrne & McEleney 2000). This difference may underlie the perception that actions are a departure from the status quo (Kahneman & Miller 1986). Because people envisage more information for actions than for inactions, they can mentally change actions more easily than inactions.

Intriguingly, in certain circumstances people imagine a counterfactual alternative to an *inaction*. People usually think about a single state of affairs when they understand an inaction. But, of course, they can switch from thinking

about one possibility to thinking about two possibilities. In *The Rational Imagination* I consider several ways in which this switch to thinking about two possibilities for an inaction occurs. For example, people think about two possibilities to mentally represent a failure to act when they know that there were good reasons to act (Zeelenberg et al. 2002; Zhang et al. 2004). They also think about two possibilities when they consider the consequences of a failure to act from a long-term perspective (Gilovich & Medvec 1995). When most people look back over their lives, it is their inactions that they tend to regret, the things they failed to do, such as failing to spend time with family and friends, failing to avail of educational opportunities, or failing to pursue hobbies. I outline in the book how people switch to thinking "if only" about a failure to act from a long-term perspective when they can imagine a better outcome (e.g., if he had switched to a different college, he would have been happy), compared to the outcome they imagine when they think "if only" about an action (e.g., if he had stayed in his original college, he would have remained unhappy) (Byrne & McEleney 2000). People can think of the good that may have resulted had more time been spent with family and friends, had educational opportunities been availed of, or had hobbies been pursued. In these situations, if only the person had acted, a better outcome might have occurred than the outcome that did in fact occur. A second corollary to the principles that guide the possibilities that people think about is that people can switch from thinking about one possibility to thinking about two possibilities in some circumstances. The impact of an imagined better outcome is illustrated in the example of Jane who went to a travel agency to look for holidays in the Caribbean. There was a wonderful package deal at a very affordable price. She wondered whether to book it immediately but she decided to think about it overnight. The next day when she returned to the travel agency the holiday had been sold and there were no other deals that looked as attractive. Jane is likely to regret her inaction. If she had acted, she would have been able to go on a fabulous holiday.

People tend to focus on actions when they think about what could have happened. Even more strikingly, they tend to focus on forbidden actions. In Chapter 4 of *The Rational Imagination*, I offer an account of how people think about what *should* have happened and what *should not* have happened.

4. Thinking about what *should* have happened

When people think about what they could have done differently, they sometimes think about what they *should* have done differently. Everyday judgments are often based on beliefs about obligations, and about what is permitted and what is forbidden. Should scientists be allowed to clone humans? Are you morally obliged to recycle your office waste-paper? Ought manufacturers to identify products containing genetically modified ingredients? In Chapter 4, I first sketch some characteristics of the tendency shown by most people to focus on forbidden actions. Next, I outline some clues from reasoning that help identify why this perceived fault-line is observed. The clues come from studies of how people reason about obligations. Then, I show how these clues help to explain the tendency people have to focus on forbidden actions, and for obligations to be immutable.

Consider Steven, who was delayed on his way home by a series of events, and arrived too late to save his wife from dying of a heart attack (Girotto et al. 1991). The events that delayed Steven included a road blocked by a fallen tree, visiting a bar for a drink, and having an asthma attack. What do you think Steven's thoughts about "what might have been" focused on after the death of his wife? When people were asked to complete his "if only . . ." thoughts, they focused on his decision to stop at a bar for a drink, saying "if only I hadn't stopped for that drink" (Girotto et al. 1991). Of course, stopping for a drink is the only event in the series that was completely within Steven's control. It is also an event that may be perceived to be socially unacceptable. Drinking in a bar may fit with social norms about how to celebrate with friends, but it seems inappropriate in the context of drinking alone while your spouse is ill at home. Most people focus their thoughts about what-might-have-been on those controllable events that are socially unacceptable, more than on controllable events that are socially acceptable (McCloy & Byrne 2000). The contents of thoughts are constrained not only by natural laws, but also by social laws such as social conventions and cultural regulations. Why do people imagine alternatives most readily to socially unacceptable events? An important clue comes from the inferences people make about obligations.

4.1. Clues from reasoning: Inferences about obligations

Knowledge of a regulation – for example, "hard-hats must be worn on the building site" – ensures that most people think about not only what is obligatory, such as wearing a hard hat, but also what is not permissible, such as not wearing a hard hat. A sixth principle about the possibilities that people envisage is that they think about the forbidden possibility as well as the permitted possibility when they understand an obligation. For the obligation, "if Jack's parents are elderly, he should look after them," people keep in mind two possibilities: (1) Jack's parents are elderly and he looks after them, and (2) Jack's parents are elderly and he does not look after them. They note that the status of this second possibility is that it is forbidden. People usually think about true possibilities. Even when they understand a counterfactual conjecture that contradicts known facts, they may temporarily suppose the conjecture to be true. But obligations are different. Their understanding requires people to think about a forbidden possibility. Obligations are unique in that they require people to consider the possibility that is explicitly ruled out by the conditional. The forbidden possibility influences the inferences they make and the counterfactual alternatives they imagine.

Most people reason well about obligations (Bucciarelli & Johnson-Laird 2005; Cheng & Holyoak 1985; Cosmides 1989; Sperber et al. 1995). Because people think about the permitted possibility and the forbidden possibility, they can make certain inferences readily. When reasoners are told "if Paul rides a motorbike, he must wear a helmet," and are also told that "Paul is not wearing a helmet," they can infer that Paul is not permitted to ride a motorbike. In other words, they readily make the otherwise

https://doi.org/10.1017/S0140525X07002658 Published online by Cambridge University Press

difficult *modus tollens* inference from a conditional with obligation content. The ease with which they make the inference highlights an important feature of reasoning. There is more than one way to make an inference. A common route to making the *modus tollens* inference is by thinking about the true possibilities, as we have seen earlier. But reasoning about obligations illustrates a second route for the inference, by thinking about what is forbidden. For the conditional, "if Paul rides his motorbike, he must wear a helmet," people think about two possibilities: (1) Paul rides his motorbike and he wears a helmet, and (2) Paul rides his motorbike and he does not wear a helmet (forbidden). When they are told that Paul did not wear a helmet, they can match this information to the forbidden possibility, and infer that it is forbidden for him to ride his motorbike. The theory accurately predicts that people can make more *modus tollens* inferences from obligation conditionals compared to conditionals with other sorts of content (Quelhas & Byrne 2003).

Counterfactual obligations communicate, through shared knowledge or context, information about what was once obligatory and what was once forbidden; and they also communicate that what was once unacceptable is no longer so. Suppose, for example, that it used to be obligatory to pay a property tax on houses which increased with increasing house size, but that the tax was abolished some years ago. Suppose your friend buys a large house. You might say, "if you had bought that house 10 years ago, you would have had to pay a very large tax bill." For a counterfactual obligation, people may think about one of the permitted possibilities conjectured by the counterfactual: your friend bought the house 10 years ago and paid a large tax bill. They may also think about the forbidden possibility: your friend bought the house 10 years ago and did not pay a large tax bill. And they may also think about the presupposed facts: your friend did not buy the house 10 years ago and did not pay a large tax bill. These insights from how people reason about obligations and forbidden events help to explain why people are drawn to forbidden events when they imagine counterfactual alternatives.

4.2. The rational imagination: Why people focus on forbidden fruit

When you learn that manufacturers in Europe must identify genetically modified constituents in food, you understand what is forbidden: they are not allowed to conceal genetically modified constituents. When you know that scientists are obliged to restrict cloning attempts to nonhuman species, you understand what is forbidden: they are not allowed to clone humans. A conditional obligation is understood by thinking about one of the permitted possibilities, and also what is forbidden. This ready access to the forbidden possibility has a large impact on how people imagine events turning out differently.

People keep in mind both the permitted possibility and the forbidden possibility, and so they can readily imagine a counterfactual alternative to the forbidden possibility. Steven, who did not return home in time to save his wife, is judged to say "if only ..." most often about his decision to visit a bar for a drink. A third corollary to the principles that guide the possibilities that people think about is that they envisage two possibilities when they understand controllable events, as Table 2 shows. People imagine alternatives to socially unacceptable actions, such as stopping at a bar for a beer in the context of an ill spouse waiting at home, because they think about two possibilities – stopping for a beer and not stopping for a beer – and they note that one of the possibilities is unacceptable (stopping for a beer), and the other is acceptable (not stopping for a beer). They imagine an alternative by mentally changing the unacceptable possibility and they say, "if only Steven had not stopped for a beer ..." The unacceptable possibility brings to mind its acceptable counterpart, just as an event that is exceptional for an individual brings to mind its normal counterpart (Kahneman & Miller 1986).

If you do what you should do socially, people do not tend to wish you had not done it. Consider a situation in which a friend of yours, Bernard, is disappointed: he had an important performance at a concert in front of thousands of people and it was not well received. You know he stayed out late at a pre-concert party the night before, something he rarely does. He tells you he went to the party because he wanted to meet one of his orchestral heroes. You may believe that his poor performance was his own fault and he should not have gone to the party (Klauer et al. 1995). But suppose, instead, your friend tells you he went to the party because it was a fund-raiser for his orchestra. You understand his obligation by keeping in mind two possibilities: "he went to the party" is acceptable in this case, and "he did not go to the party" is unacceptable. Would you think he should not have gone to the party? Would you think his poor performance was his own fault? When participants described Bernard's thoughts, they generated reliably fewer "if only ..." sentence completions that focused on the party when they were told he was obliged to go to it, compared to when they were not given a reason (Walsh & Byrne 2007).

Some theorists believe that the ease with which people reason about certain permissions or obligations shows that the mind evolved to store knowledge about social regulations in specialised modules (Cosmides 1989). The

Table 2. *Summary of some corollaries to the principles that guide the possibilities that people think about when they imagine counterfactual alternatives*

1. **Actions:** People think about two possibilities when they understand an action.
2. **Single possibilities:** People can switch from thinking about one possibility to thinking about two possibilities, for example, for inactions.
3. **Controllable events:** People think about two possibilities when they understand controllable events.
4. **Causes:** People think about a single possibility when they understand a strong causal relation (cause and outcome). They can readily access a second possibility for an enabling condition (enabler and outcome, no enabler and no outcome).
5. **Semifactual alternatives:** When people mentally represent a semi-factual assertion they think about two possibilities, A and B, and not-A and B. They mentally represent a weak causal relation by thinking about these same two possibilities.
6. **Anchored possibilities:** People think of the first element in the possibility as immutable.

modules contain rules that are specific to reasoning about certain content, such as the domains of obligation and permission (Cheng & Holyoak 1985), and they are operated on by specialised procedures (Gigerenzer & Hug 1992). An alternative view is that people *reason* about permissions and obligations rather than give fixed responses to them (Sperber & Girotto 2003). Their reasoning depends on their ability to think about permitted possibilities and especially about possibilities that are forbidden. Human beings evolved but it is not known yet which aspects of the brain resulted from selective pressures. In Chapter 4, I suggest that what evolved is the ability to understand obligations by thinking about not only the permitted possibility but also the forbidden possibility.

People tend to focus on controllable actions, especially controllable actions that are socially unacceptable. But intriguingly, people often imagine a counterfactual alternative to a controllable event even when the controllable event was not the cause of the outcome. In Chapter 5 of *The Rational Imagination* I explore the vagaries in the relationship between causal and counterfactual thoughts.

5. Causal relations and counterfactuals

In the days that followed the September 11th, 2001 attacks on the World Trade Center in New York, many media reports focused on the failure of the airport security personnel to detect passengers who carried weapons on board the aircraft. The implication of the reports was that the baggage handlers' failure allowed the attack to occur. But why was media focus drawn to this factor? Presumably no one believed that the baggage handlers *caused* the attack. And media attention could have focused on other plausible causes, such as, "if only the Al-Qaeda network did not exist. . ." and many other antecedent events, for example, "if only the hijackers had been over-powered on board. . . ." The focus on the baggage handlers illustrates a curious feature of counterfactual thoughts: Their relation to causal thoughts is not straightforward. In Chapter 5 of *The Rational Imagination* I propose that counterfactual and causal thoughts sometimes focus on different events because people can distinguish strong causes from enabling conditions. Their causal thoughts tend to focus on strong causes; their counterfactual thoughts tend to focus on enabling conditions. If the enabling condition had not occurred the outcome would not have occurred.

Consider, as an example, that you were seriously injured in a car crash that happened on a route you do not usually take home, and that occurred when a careless driver crashed into you. What would you tend to say "if only . . ." about most often: "if only I'd gone home by my usual route" or "if only the other driver had been more careful?" Most people say "if only I had gone home by my usual route," even though they identify the other driver as the cause of the accident (Mandel & Lehman 1996). This tendency has been verified in naturalistic studies of individuals who have sustained severe injuries, such as paralysis, as a result of a traumatic accident. They tend spontaneously to imagine counterfactual alternatives that depend on how they could have prevented their accident, without believing that they caused it (Davis et al. 1996). In considering causal and counterfactual thoughts, Chapter 5 follows the by-now familiar structure: First, I sketch some characteristics of the relation between counterfactual and causal thoughts. Next, I garner some clues from reasoning that help to clarify the relationship. The clues come from studies of how people reason about different causes in particular strong causes and enablers. Then, I show how these clues help to explain the tendency people have to focus on enabling conditions when they imagine a counterfactual alternative to reality.

5.1. The causal chicken and the counterfactual egg

Causal thoughts can influence the counterfactual alternatives that people imagine. Causal knowledge affects judgments about the plausibility of a counterfactual speculation. Experts committed to a causal law perceive less "wiggle room for re-writing history" (Tetlock & Lebow 2001, p. 838). And the relation is reciprocal: People sometimes imagine counterfactual alternatives to help them to work out causal relations. Consider for example, what caused the rise of the West? If Islamic armies had conquered France and Italy in the 8th century, would later European development have been side-tracked? Such counterfactual thoughts are an essential ingredient in historical reasoning, even if they are considered controversial (Tetlock & Lebow 2001).

The counterfactual thought that if an antecedent had not happened, then the outcome would not have happened, increases the judgment that the antecedent has a causal relation to the outcome. Consider a taxi driver who refused to give a couple a lift and the couple were subsequently killed in an accident in their own car as they drove across a bridge that collapsed (Wells & Gavanski 1989). Most people judge the driver's decision not to give the couple a lift to have had some causal role in the couple's death when they are told that the taxi driver crossed the bridge safely. But they tend to judge the driver's decision to have had little causal role when they are told that he was also killed as he drove across the collapsing bridge. When the driver was also killed, they cannot say "if only he had given them a lift, they would still be alive" (Wells & Gavanski 1989). Causal thoughts can influence counterfactual thoughts, and vice versa. But causal thoughts and counterfactual thoughts can diverge. An important clue to explain why people imagine counterfactual alternatives to events within their control, even though they know these events were not the cause of the outcome, comes from the study of different sorts of causal relations.

5.2. Clues from reasoning: Strong causal and enabling relations

The observation that people sometimes focus on different events in their counterfactual and causal thoughts can be explained by the possibilities that people think about. The explanation has two main components. The first part of the explanation as to why the focus of counterfactual and causal thoughts differs is that people mentally represent a causal claim and a counterfactual one by thinking about different possibilities. A causal claim can be readily paraphrased as a counterfactual; for example, "Heating the water to 100 degrees centigrade caused it to boil" can be

paraphrased by "if the water had not been heated to 100 degrees centigrade it would not have boiled." Many philosophers have supposed that to think counterfactually is to think causally (Hume 1739/2000; Mill 1872/1956). The strong causal relation is consistent with the possibility, "water was heated to 100 degrees and it boiled," as well as the possibility, "water was not heated to 100 degrees and it did not boil." The counterfactual is consistent with the same two possibilities. However, people mentally represent a causal relation and a counterfactual in different ways. For example, they keep in mind just a single possibility initially when they understand the strong causal relation. But they keep in mind both possibilities when they understand the counterfactual, one corresponding to the conjecture and the other to the presupposed facts. As a result, the counterfactual makes more information readily available to them from the outset than does the causal assertion.

The second part of the explanation for why the focus of counterfactual and causal thoughts differs is that there are different sorts of causes and people think about different possibilities when they understand them. A strong causal relation, such as the one about water boiling, is consistent with two possibilities – although people initially think about just one of them (Goldvarg & Johnson-Laird 2001). An enabling relation, such as "dry leaves made it possible for the forest fire to spread" is consistent with three possibilities: "dry leaves and the forest fire," "no dry leaves and no forest fire," and "dry leaves and no forest fire" (Goldvarg & Johnson-Laird 2001). People can distinguish between strong causal relations and enabling relations (Cheng 1997; Goldvarg & Johnson-Laird 2001). Moreover, they make different inferences from strong causal relations and enabling relations (Byrne et al. 1999).

People mentally represent the enabling relation by initially thinking about the possibility, "dry leaves and the forest fire," and they can readily access a second possibility, "no dry leaves and no forest fire." An important corollary to the principles that guide the possibilities people consider is that they think about a single possibility when they understand a strong causal relation; they are able to think readily about a second possibility when they understand an enabling relation. This second possibility provides a ready-made counterfactual alternative. They can imagine the situation turning out differently if the enabling condition had not occurred. Consequently, people tend to focus more often on enabling relations than on strong causal relations in their counterfactual thoughts.

5.3. The rational imagination: Why counterfactual and causal thoughts differ

The clue from reasoning about causes that helps solve the puzzle is the crucial distinction between a strong causal relation and an enabling relation. Counterfactual thoughts often focus on enabling conditions, for example, "if only I had taken a different route home, I would not have had the crash," whereas causal thoughts often focus on strong causes, for example, "the drunk driver caused the crash." Causes can seem immutable because people keep in mind just a single possibility. Enabling conditions seem mutable because they require people to think about two possibilities. Enablers can be mentally deleted from an imagined alternative – dry leaves removed, alternative routes home taken, airport security improved. Wishing that whatever could have been done had been done to prevent the outcome or promote a better one may be a plausible alternative.

It is easier to think about one possibility than about several possibilities. The theory predicts that people should produce more causal thoughts than counterfactual thoughts, because they mentally represent a strong causal relation by thinking about a single possibility, and they mentally represent a counterfactual by thinking about two possibilities. The data support the hypothesis: Most people tend to think spontaneously more often about why a situation turned out the way it did, than about how it might have turned out differently (McEleney & Byrne 2006). Further support comes from the observation that causal assertions are made earlier by children than counterfactual assertions (Harris 2000). An important development of childhood cognition is the ability to keep more information in working memory (Oakhill & Johnson-Laird 1985). As a result, children should be able to think about a strong causal relation, for which they need think about only a single possibility, before they can imagine a counterfactual alternative, for which they must think about two possibilities. Two-year-olds can engage in pretend play, for example, they can climb into a large brown box and pretend it is a boat at sea (Riggs & Peterson 2000). Counterfactual thinking is similar to pretence but it requires a comparison of the facts with the imagined alternative (Harris 2000). By age 2 or 3 years, children can use "nearly" and "almost" (Harris et al. 1996). By age 3 or 4 years, they can answer questions about what should have been done so that a bad outcome would not have happened (Laura should have used the pencil so her fingers would not have got ink on them). Children are able to make hypothetical inferences, for example, "If I draw on this piece of paper, which box will it go in?" (in a situation in which blank sheets go in a different box from used sheets), before they can make counterfactual inferences such as "If I had not drawn on the paper, which box would it go in?" (Riggs et al. 1998).

When people imagine how an outcome could have turned out differently, their "if only . . ." thoughts about an antecedent increase their judgment that it is causally related to the outcome. Intriguingly, when they imagine how the outcome might have turned out the same, their "even if . . ." thoughts about the antecedent decrease their judgment that it is causally related to the outcome. I address this other side of the causal coin in Chapter 6 of *The Rational Imagination*.

6. "Even if . . ."

Suppose a survivor from an airplane crash with severe injuries struggles for days through the jungle but dies just before reaching a village. You might think "if only he had managed to walk to the village, he would have been rescued" (Miller & Turnbull 1990). But suppose you wish to defend the rescue team who got as far as the village but no further. Your motivation to defend may influence the alternative you imagine (Roese et al. 2005).

You may focus on the severity of the victim's injuries and suggest "even if he had managed to walk to the village, he still would have died." "Even if . . ." conditionals have been called "semi-factual" by philosophers because they combine a counterfactual antecedent and a factual consequent (Chisholm 1946). Uniquely among thoughts about what might have been, imagined semi-factual alternatives suggest that the outcome is inevitable.

When people can imagine an alternative in which an antecedent does not occur and yet the outcome still occurs, their judgment that the antecedent caused the outcome decreases. Their judgments of fault, responsibility, and blame are also affected dramatically. Suppose your task is to change an account of a bank-robbery, but you are to alter only the actions of the bank-robber, and you are to ensure that the outcome is different, for example, the teller does not get shot (Branscombe et al. 1996). You might imagine that the bank-robber did not grab the teller and the teller did not get shot. Suppose instead that your task is to change the story, again altering only the actions of the bank-robber, but this time you are to ensure that the outcome remains the same. You might imagine that the bank-robber did not grab the teller but she still got shot. People blame the bank-robber when they imagine his actions were different and the outcome was different, more so than when they imagine his actions were different but the outcome remained the same (Branscombe et al. 1996). The effects of a counterfactual or semi-factual alternative on judgments of cause and blame are implicated in everyday decisions, including decisions made by juries. Why do "even if" thoughts decrease judgments of causality? The explanation relies on the identification of a third sort of causal relation, distinct from a strong causal relation or an enabling condition – namely, a weak causal relation. In Chapter 6, I first sketch some characteristics of "even if . . ." thoughts. Next I outline some clues from reasoning. The clues come from studies of how people reason about semi-factual conditionals. Then, I show how these clues help to explain how people imagine "even if . . ." alternatives to reality.

6.1. Clues from reasoning: "Even if . . ." conditionals and inferences

People think about a semi-factual conditional such as, "even if the airplane crash victim had made it to the village, he still would have died" by thinking initially about two possibilities. They envisage the conjecture, "he made it to the village and he died" and the presupposed facts, "he did not make it to the village and he died." A corollary to the principles that guide the possibilities that people think about is that they mentally represent an "even if . . ." conditional by thinking about two possibilities, one in which the antecedent and the outcome both occur, and one in which the antecedent does not occur but the outcome occurs.

There is clear evidence that people think about two possibilities when they understand a semi-factual "even if . . ." conditional. A semi-factual conditional such as "even if there had been lilies there would still have been roses" primes you to read the conjunction "there were no lilies and there were roses" very rapidly. You understand the semi-factual by mentally representing two possibilities from the outset: "there were lilies and there were roses" and "there were no lilies and there were roses" When you are subsequently told that in fact "there were no lilies and there were roses" you can process this information rapidly. It matches one of the possibilities you have thought about from the outset. The theory accurately predicts that people are able to read the conjunction "there were no lilies and there were roses" more quickly after they have understood an "even if" conditional compared to after they have understood a factual "if" conditional (Santamaria et al. 2005; See also Moreno-Rios et al., in press). The evidence shows that people understand semi-factual conditionals by keeping in mind two possibilities, one in which the antecedent and outcome both occur and one in which the antecedent does not occur but the outcome occurs. It provides an important clue to solve the puzzle of why people do not tend to think an antecedent caused the outcome when they can imagine a semi-factual alternative.

6.2. The rational imagination: Imagined semi-factuals and causality

Consider a scenario about a runner competing in the Olympics (based on Boninger et al. 1994):

> On the day before the 400 metre race, in a freak accident during training, you sprain your left ankle . . . Your trainer recommends that you choose between two drugs, both legal according to Olympic guidelines. One is a well-known painkiller that has been proved effective but also has some serious side effects including temporary nausea and drowsiness. The other pain killer is a newer and less well-known drug . . . the newer drug might be a more effective pain killer [but] its side effects are not yet known . . . After considerable thought, you elect to go with the more well-known drug. On the day of the race, although there is no pain in your ankle, you already begin to feel the nausea and find yourself fighting off fatigue. You finish in fourth place.

In the days after the race the runner thought "even if . . ." How do you think she completed this thought? People completed the "even if . . ." sentence by saying, for example, "even if I had taken the newer drug, I still would have lost the race." They think about two possibilities: the conjecture, "she took the newer drug and she lost the race," and the presupposed facts, "she did not take the newer drug and she lost the race." To what extent do you think the runner's decision to take the older well-known drug led to her failure to obtain an Olympic medal in the 400 metre race? In *The Rational Imagination* I propose that the two possibilities imagined for the semi-factual alternative emphasize that the antecedent, taking the well-known drug, was not necessary for the outcome. They correspond to the possibilities that people consider initially for a weak causal relation. The theory accurately predicts that judgments of the strength of the causal relation between the antecedent and the outcome are weaker when people imagine a semi-factual alternative (McCloy & Byrne 2002).

When people imagine an alternative to reality in which the outcome turns out the same even though they acted differently, the outcome can seem inevitable (Sanna et al. 2002b). The idea that an outcome was inevitable can be comforting and can even appear to absolve people of blame and responsibility. Some religions and cultures

value the related ideas of fate and destiny. Given that the ideas of individual progress and perfection gained currency just a few hundred years ago, we can speculate that our ancestors may not have been so inclined to perceive fault-lines in reality and to imagine alternatives that are better than reality. In Chapter 7 of *The Rational Imagination* I consider one last perceived fault-line of reality. I examine the tendency people show to imagine alternatives in which they change the last event in a simple sequence of events.

7. The "last chance"

During the World Cup soccer tournament in Japan in 2002, the match between Ireland and Germany went into several minutes of extra "injury" time. The score was 1–0 to Germany. Then, at the eleventh hour, Robbie Keane scored a goal. For every Irish fan who watched the match there could be no doubt: the striker had saved Ireland's World Cup hopes. Did Keane really save the day? The question illustrates a curious feature of human thought. In a temporal sequence of events, people zoom in on the most recent one. This "last chance" focus in counterfactual thought extends beyond the sporting domain. For example, counterfactual thoughts about how major events in history could have turned out differently also tend to focus on the "last chance" juncture, such as in conjectures about why the West rose to dominance (Tetlock & Parker 2005). In Chapter 7, I first sketch some characteristics of the tendency shown by most people to focus on the most recent event rather than on earlier events. Next I outline some clues from reasoning that help identify why this perceived fault-line exists. The clues come from studies of how people reason about "only if" conditionals. Then, I show how these clues help to explain the tendency people have to focus on recent events.

Consider a game in which two individuals, Lisa and Jenny, each toss a coin. If both tosses come up the same, heads or tails, they will each win $1,000. But if they come up different, neither wins. Lisa goes first and tosses a head, Jenny goes next and tosses a tail, and so neither wins. How do you think things could have turned out differently, if only . . .? When people are asked to think about how things could have turned out differently, almost all of them think, "if only Jenny had tossed a head . . ." (Miller & Gunasegaram 1990). What is more, they judge that Lisa will blame Jenny more, and that Jenny will feel more guilt than Lisa. The next section examines clues from reasoning to explain why people focus on the most recent event in their counterfactual thoughts.

7.1. Clues from reasoning: The case of "only if" conditionals

People think about possibilities in a way that preserves a record of the temporal order of the events in the world. Consider an "only if" conditional "Alicia went swimming only if she slept well." You may be surprised to find that an "if" conditional, such as "if Alicia went swimming, she slept well," and an "only if" conditional, such as "Alicia went swimming only if she slept well," are usually logically equivalent. The equivalence can be illustrated by thinking about what possibility is ruled out by "Alicia went swimming only if she slept well." Most people judge accurately that it rules out the possibility that Alicia went swimming and she did not sleep well, which is the same possibility ruled out by the "if" conditional (Jeffrey 1981). But the two conditionals are not psychologically equivalent (Evans 1977). When you know that "Alicia went swimming only if she slept well," which event do you believe occurs first, Alicia went swimming, or, she slept well? The "only if" conditional seems to work best when its second clause, "Alicia slept well," refers to a state of affairs that holds prior to the state referred to in its first clause, "Alicia went swimming" (Marcus & Rips 1979).

People make different inferences from "only if" and "if" conditionals. Suppose you know that "Alicia went swimming only if she slept well" and then you find out that "Alicia did not sleep well." What would you conclude? Many people conclude readily that, "Alicia did not go swimming." The *modus tollens* inference, which is difficult to make from an "if" conditional, is made readily from the "only if" conditional. Why is the inference so easy from the "only if" conditional? The answer is that "only if" requires people to think about more possibilities from the outset than "if" (Johnson-Laird & Byrne 1989). People understand "Alicia went swimming only if she slept well" by thinking about the possibility in which both events occurred and they also think about the possibility in which neither event occurred. As a result, they can readily make the inference. But there is more to the mental representation of "only if" than keeping in mind two possibilities.

Temporal information is conveyed implicitly by many "only if" conditionals, such as "Alicia went swimming only if she slept well." People understand the order in which the events occurred in the world and so they think about the possibility "Alicia sleeps well and she goes swimming" rather than the opposite order, "Alicia goes swimming and she sleeps well." Likewise, they think about the second possibility "Alicia does not sleep well and she does not go swimming." When they read an "only if" conditional, "A only if B," they are primed to read quickly the conjunction, "B and A" (with the clauses in the reversed order of mention to their order in the "only if" conditional). They are also primed to read quickly the conjunction "not-B and not-A." They read "not-B and not-A" reliably faster after they read "A only if B" compared to "if A, then B." The result supports the view that people think about two possibilities when they understand "only if." Equally tellingly, they do not read "A and B" more quickly after they read "A only if B" compared to "if A then B" (Santamaría & Espino 2002). A seventh principle about the possibilities people consider is that the possibilities encode the temporal order of events in the world. The evidence shows that people understand "only if" conditionals by keeping in mind two possibilities which preserve the temporal order of the events in the world. It provides an important clue to solve the puzzle of why people imagine a counterfactual alternative that changes the most recent event in a sequence of events.

7.2. The rational imagination: Why people change recent events

In the coin toss game in which Lisa tossed heads and Jenny tossed tails, most people believe the players

https://doi.org/10.1017/S0140525X07002658 Published online by Cambridge University Press

could have won if only Jenny had tossed heads. There are three counterfactual possibilities for this game but people tend to think about just one of them, the possibility that "Lisa tossed heads and Jenny tossed heads and they win." They do not think about the possibility that "Lisa tossed tails and Jenny tossed heads and they lose" because it is not an *effective* counterfactual alternative; that is, it does not change the outcome (Byrne 1997). But why do they not think about the possibility, "Lisa tossed tails and Jenny tossed tails and they win?" The answer lies in the final corollary to the principles of possibilities: people encode the first part of a possibility as immutable. When they imagine a counterfactual alternative to the coin toss, they hold constant the first player's selection, "Lisa tossed heads," and they change the second player's selection, "Jenny tossed tails." The first player's selection is presupposed (Miller & Gunasegaram 1990). It provides the background against which later events are perceived (Sherman & McConnell 1996).

Suppose you wanted a computer program to behave like a person who participated in the coin toss game. The program must generate the same sorts of counterfactual alternatives as people do and so it must focus on the most recent event. To carry out the task in the way that the theory proposes people do, the program needs to keep a record of possibilities and make changes to them. A computer program written in the programming language LISP to simulate the theory takes as input a description of the coin toss game, "if the two coins are of the same face (both heads or both tails), each player wins $1,000" (Walsh & Byrne 2004). It produces as output a counterfactual about how the events could have turned out differently: "if Jenny had tossed heads, they would have won." The program makes a record of the facts described in the story, "Lisa tossed heads and Jenny tossed tails and they both lost." It also makes a record of the set of counterfactual possibilities suggested by the story, that is, the possibilities in which the players would have won. The program compares its record of the facts to the two possibilities in which the players would have won. It generates a counterfactual alternative by changing parts of its record of the facts to be like parts of its record of the possibilities in which the players would have won. The program relies on a simple algorithm based on the key principle: Because Lisa is the first player mentioned in the facts, her selection is the anchor and it is held constant.

People do not readily think of an alternative to an anchored idea. But just as people can sometimes switch to thinking about two possibilities in situations in which they usually think about a single possibility, so too an anchored idea can be transformed into a more mutable one. The theory predicts that people should be able to imagine a counterfactual alternative even to an anchor such as the first player's selection, when they have thought about two possibilities rather than a single possibility. Imagine a television game show based on the coin toss game in which a technical hitch occurs:

> Lisa goes first and tosses a head. At this point, the game-show host has to stop the game because of a technical difficulty. After a few minutes, the technical problem is solved and the game can be restarted. Lisa goes first again, and this time she tosses a tail. Jenny goes next and she tosses a head.

The first event, Lisa tosses heads, is the anchor but the technical hitch provides an alternative to it: Lisa tosses tails. People think about two possibilities, the pre-hitch and post-hitch plays. They can imagine a counterfactual alternative in which they mentally change the first event. The theory accurately predicts that people say, "if only Lisa had tossed heads" as often as "if only Jenny had tossed tails" (Byrne et al. 2000).

The idea of a "last chance" is a powerful one. When people try to understand why someone carried out a shocking act, such as a school massacre, they often focus on the events that occurred immediately before it. Their focus reflects the belief that the bad outcome was precipitated by the event that immediately preceded it.

In Chapters 3 to 7 of *The Rational Imagination* I explain the perceived fault-lines of reality that people focus on in their imagination of counterfactual alternatives, such as actions, forbidden actions, strong and weak causes and enabling conditions, and recent events. The explanation depends on the idea that people think about possibilities, and the possibilities they think about are guided by a set of principles. In the penultimate chapter, Chapter 8, I consider some implications of this view of the imagination.

8. Individuals and creative thoughts

The idea that the imagination depends on thinking about possibilities has implications for understanding other aspects of cognition. In Chapter 8 of *The Rational Imagination* I consider two important implications. One implication is for understanding the nature of individual differences in imaginative thoughts. Most people exhibit the sorts of regularities described in the book, such as the tendency to imagine alternatives to actions or controllable events or forbidden events. However, a minority of people do the opposite. Chapter 8 offers an explanation for these individual differences. A second implication is for understanding the relation of counterfactual thoughts to other sorts of creative thoughts. The counterfactual imagination is one sort of imaginative thought. An explanation of it may contribute towards understanding other sorts of creative thoughts.

8.1. Individual differences in imaginative thoughts

Cognitive differences underlie many differences between individuals in thinking and reasoning (Sternberg 1997). People may create different counterfactual alternatives because of differences in their ability to think about possibilities of various sorts (Barrouillet & Lecas 1999; Torrens et al. 1999). Some individuals may tend to focus more on the facts and others on the imagined possibility. For example, what two shapes do you think would best fit the counterfactual, "if there had been a circle on the blackboard then there would have been a triangle?" People consider two possibilities when they understand the counterfactual, the imagined possibility of a circle and a triangle, and the presupposed facts of no circle and no triangle. The theory accurately predicts that people are torn between these two possibilities (Byrne & Tasso 1999). Some people say that the two shapes that would best fit the counterfactual are the circle and the triangle, whereas others say that the best two shapes would be shapes that are *not* the circle or the triangle. People

were similarly torn when their task was to say what two shapes would definitely go against the description.

There may be even more fundamental differences in the way different individuals interpret counterfactuals (Thompson & Byrne 2002). A minority of individuals seem to think about just a single possibility. About three-quarters of participants envisaged two possibilities when they understood a counterfactual such as "if Mark had gone to Moose Jaw then Karl would have gone to Medicine Hat." They thought about the conjecture, "Mark went to Moose Jaw and Karl went to Medicine Hat," and they also thought about the presupposed facts, "Mark did not go to Moose Jaw and Karl did not go to Medicine Hat." We established that they thought about these two possibilities by their answers to two tasks. First, when these individuals judged what someone who uttered the counterfactual meant to imply, they interpreted the intended implication as: "Mark did not go to Moose Jaw" or "Karl did not go to Medicine Hat," or both. Second, when they judged whether different situations were consistent or inconsistent with the counterfactual, they judged the situation, "Mark went to Moose Jaw and Karl went to Medicine Hat" to be *consistent* with it.

In contrast, about one-quarter of participants focused on a single possibility: the facts, "Mark did not go to Moose Jaw and Karl did not go to Medicine Hat." First, when these individuals were asked what someone who uttered the counterfactual meant to imply, they judged the person meant to imply that "Mark did not go to Moose Jaw" or "Karl did not go to Medicine Hat," or both – like the first group. But, unlike the first group, when they were asked to judge whether different situations were consistent or inconsistent with the counterfactual, this second group judged the situation "Mark went to Moose Jaw and Karl went to Medicine Hat" to be *inconsistent* with the counterfactual. Importantly, the two groups reasoned differently. For example, the single-possibility group who understood the counterfactual by thinking about just the single possibility, "Mark did not go to Moose Jaw and Karl did not go to Medicine Hat," tended not to make the *modus ponens* inference from "Mark went to Moose Jaw" to "therefore, Karl went to Medicine Hat." These differences in the imagination of possibilities may have consequences for the impact of counterfactual thoughts. Someone who understands a counterfactual "if Bert had driven fast, he would have been injured," by thinking only about the facts "Bert was not driving fast and he was not injured" may not benefit from the preparatory effects of the counterfactual alternative (Bert should not drive fast in the future). They also may not gain any solace from its emotional amplification (Bert feels relieved at his lucky escape this time).

8.2. Creative thoughts

My interest in *The Rational Imagination* is in the creation of counterfactual alternatives to reality. But an understanding of the counterfactual imagination may have implications for understanding other sorts of creative thoughts. Creative thoughts are relied upon to write a poem, paint a picture, compose a piece of music, design an experiment, or invent a new product. These sorts of activities can seem very different from the imagination of a counterfactual alternative to reality. But counterfactual imaginative thoughts may share some similarities with other sorts of creative thoughts, such as category expansion, concept combination, and insight.

One sort of creative thought is inventing new instances of a category, such as designing a new coffee mug, a better car, or a fashionable "new look." Take a moment to imagine a new alien life form. The results are typically diverse, from the giant plants in "The day of the triffids," to the fluid metal of the robots in "Terminator 2." When people draw a creature from a planet somewhere else in the universe, their drawings differ very widely (Ward et al. 2004). But they also show some very informative regularities. For example, people tended to make sure that their alien creature had sensory organs, most commonly eyes, and functional appendages, most commonly legs. People may think of possibilities corresponding to true instances of the category, and they may not think about possibilities that are not instances of the category. They may imagine an alternative creature by making a minimal change to their representation of the exemplar they have in mind. They can readily think of alternative possibilities for the size, shape, and color of creatures; for example, known birds come in many different sizes. They may be able mentally to alter these aspects of the category more readily than other aspects. They may think of just a single possibility to mentally represent other aspects, such as the presence of the sensory organ of eyes. These aspects of reality seem immutable because people do not mentally represent them from the outset by thinking about alternative possibilities.

A second sort of creative thought is concept combination. One way that people come up with new ideas is by combining several existing concepts to form a new one. Suppose you hear for the first time the combination "corn oil." What do you think it means? You may decide it is oil made from corn. Now imagine you hear for the first time, "baby oil." What does it mean? You are unlikely to decide it is oil made from babies. It is more likely you will conclude it is oil to rub on babies' skin. Now, what is "lamp oil?" It is unlikely you will say it is oil made from lamps, or oil to rub on lamps. Instead you might say it is oil to fuel lamps (Wisniewski 1996). When people combine two concepts they often transfer aspects of the first concept to change something about the second concept. What do you think a "robin snake" is? You might think it is a snake with a red breast. Alternatively, you might decide that it is a snake that eats robins. What is a "cactus fish?" You might consider that it is a fish that has prickly spines. Prickly spines are diagnostic of cactii, there are few other plants that have them (Costello & Keane 2001). People may combine concepts by identifying the most immutable aspect of the first concept and the most mutable aspect of the second concept. They may think about what snakes eat, or the shapes and colors of fish, by thinking about several possibilities from the outset. The perceived fault-lines of the concept correspond to the aspects for which people can readily think of several possibilities. Because they mentally represent some aspects of a concept, for example, what snakes eat, by more than a single possibility, they can alter those aspects readily.

A third sort of creative thought is insight. In the sciences and arts, individuals sometimes report experiencing a moment of "insight" in which a new idea "pops" into

mind. Suppose you are asked to describe how to throw a ping-pong ball so that it will go a short distance, come to a dead stop, and then reverse itself. You are not allowed to bounce the ball against any object or attach anything to it (Ansburg & Dominowski 2000). What solution would you suggest? Most people suggest throwing the ball so that it curves back to them (but this solution violates the constraint that the ball comes to a dead stop). Others suggest throwing the ball against a wall or to another person (but this solution violates the constraint not to bounce the ball against any object). Few people reach the correct solution, to throw the ball up into the air. People may think about problems such as the ping-pong one by thinking about possibilities that are limited by their previous experience with similar situations (Galinsky & Moscowitz 2000; Keane 1997). They may add the usual horizontal trajectory of a ping-pong ball in play to their mental representation of the problem. But they may not add the horizontal trajectory assumption to their mental representation of other sorts of ball in play. In a basketball version of the problem, the trajectory of the ball should become a fault-line in the mental representation of reality, that is, an aspect of the facts for which people can think of alternative possibilities. This account accurately predicts that people can solve a basketball version more readily than a ping-pong version of the same problem (Murray & Byrne 2007). Moments of insight may be moments of mutability, when a previously immutable aspect of the mental representation of reality is transformed into a fault-line. In Chapter 8 of *The Rational Imagination* I suggest that the idea that the imagination depends on thinking about possibilities has implications for understanding individual differences and for understanding other sorts of creative thoughts. In Chapter 9, I consider what it means for the counterfactual imagination to be rational.

9. The idea of a rational imagination

Human mental life would be very different from the way it is if people could not imagine counterfactual alternatives to reality. To appreciate how commonplace thoughts about what might have been are, consider what life would be like without them. There would be no conjectures about how things might have turned out differently, or the same, and so people would not experience a sense of inevitability or a sense that something else "almost" happened (Roese & Olson 1995). They would find it hard to learn from their mistakes and to plan how to avoid similar ones in future. Their experiences of hope or relief and regret or guilt or remorse would be impoverished. Their ability to ascribe blame, fault, and responsibility would be limited.

Some unfortunate individuals do appear to lose the capacity to create counterfactual alternatives as a result of brain injury to the frontal cortex (Knight & Grabowecky 1995). Provided people are able to imagine counterfactual alternatives, does it matter whether the counterfactual imagination is rational or irrational? One reason that people create alternatives to reality may be to learn from mistakes and to prepare for the future (Roese 1994). Another reason may be to work out causal relations (Roese & Olson 1995). Thoughts about what might have happened "if only . . ." or "even if . . ." may help people to make sense of their personal histories and they may help people in their attempts to make sure that events in the future turn out better. If the counterfactual imagination is an irrational process, then its usefulness and reliability is in doubt.

In *The Rational Imagination*, the claim that the counterfactual imagination is rational depends on three steps. The first step is that humans are capable of rational thought. People can make rational inferences, as shown by evidence about their deductions. But people also make mistakes. I argue that the existence of error does not undermine the idea of rational thought. People may be rational in principle, that is, they have the mental machinery to make rational inferences. But they may err in practice, that is, their performance is constrained by various factors such as working memory limitations, as well as their knowledge, beliefs, and interest in different topics (Johnson-Laird & Byrne 1991). Their competence to be rational rests on their ability to imagine alternatives, including counterexamples to conclusions. The second step towards the conclusion that the counterfactual imagination is rational depends on the idea that people make inferences by thinking about possibilities. This view places imagination at the heart of reasoning (Johnson-Laird & Byrne 2002). There is extensive experimental support for it (for a review, see Johnson-Laird 2001). I argue that the idea that human reasoning depends on the imagination of possibilities is crucial to the idea that the counterfactual imagination is rational. The third step towards the conclusion that the imagination is rational is that counterfactual thoughts rely on thinking about possibilities, just as rational thoughts do. In *The Rational Imagination* I outline the set of principles that guide the possibilities that people think about when they create counterfactual alternatives to reality and I sketch how these principles apply to key phenomena of the counterfactual imagination.

Of course, you may accept these three steps and still reject the conclusion that the counterfactual imagination is rational. I suggest that the conclusion depends on what it means for the counterfactual imagination to be rational. A strong version is that the cognitive processes that underpin the counterfactual imagination are capable of producing the best counterfactual thoughts. Of course, there are many examples of the counterfactual imagination producing irrational outputs (Miller et al. 1990; Ritov & Baron 1992). For example, when people can readily imagine a woman who was attacked doing something differently, say, walking home by a different route, or not accepting a lift, they tend to blame the victim for her fate (Branscombe et al. 1997). Their judgments of fault, responsibility, and causality are swayed by how readily they can imagine an alternative. And how readily they can imagine an alternative sometimes depends merely on the way in which the information about the facts was presented to them (Walsh & Byrne 2004). But just as the existence of invalid deductions does not indict human competence in reasoning, so too the existence of irrational judgments does not indict human competence in counterfactual imaginative thought.

How can you tell whether a counterfactual thought is the best one? There is no normative standard. A counterfactual thought could be judged to be good by whether or not it serves its purpose, for example, to help people to learn from mistakes (Roese et al. 2005). But of course, people may imagine what might have been different

even when such thoughts are not useful for preparing for the future. Alternatively, a counterfactual thought could be judged to be good if it helps people to feel better. But if people only generated counterfactual thoughts about how things could have been worse, their thoughts may act as a false panacea. Another option is that a counterfactual thought could be judged to be good by how plausible it is. For example, historical counterfactual conjectures can be illuminating or merely whimsical (Tetlock & Parker 2005). The difference may depend on whether or not people made *minimal* changes to their mental representation of reality to create the counterfactual alternative (Pollock 1986). In the book I suggest that the regularities that people exhibit, the perceived fault-lines of reality, are not limitations of the imagination; they are, in fact, the very hallmark of its rationality. The regularities that people exhibit reflect a rational exploitation of fault-lines in their representation of reality. People think about two possibilities from the outset when they understand certain ideas, such as choices, actions, controllable events, forbidden actions, and so on. Minimal changes may ground imaginative thoughts in the bedrock of rationality and ensure that the facts are recoverable from the imagined possibility (Byrne 1997).

Is the creation of counterfactual alternatives a uniquely human capacity? There is little evidence as yet to indicate whether animals reflect upon their past mistakes and create counterfactual alternatives about what might have been. There is evidence that people do so. Reality can seem to shimmer with glimpses of counterfactual alternatives. Human mental life is made undeniably richer by their existence.

ACKNOWLEDGMENTS

I wish to thank Paul Bloom, Phil Johnson-Laird, and Mark Keane for their helpful and insightful comments on an earlier draft of this précis. I am also grateful to each of my collaborators for allowing me to discuss our joint work in this context.

Open Peer Commentary

Imagination and reason

doi: 10.1017/S0140525X07002580

Joseph Agassi

Department of Philosophy, Tel-Aviv University, Ramat Aviv 69978, Israel.

agass@post.tau.ac.il **http://www.tau.ac.il/~agass/**

Abstract: Byrne's book is intended to explain why people imagine the things they do when they create alternatives to reality. Two fruitful areas of further research are: (1) How can her approach explain dreams and daydreams? (2) What is the developmental time course of the child's understanding of reality and imagination?

Ruth Byrne's (2005) book provides many examples of everyday imagination of variations on everyday experiences. Her focus on is the assessment of their rationality. In the cases that she (and the pertinent literature) discusses, the objects of the imagination are small variants on past events. Some of these variants are extreme ones and not very useful, such as imagining the probable but false event that my parents have never met each other. (This is how Clarence Darrow opens his autobiography.) Other variants are less imaginative but more useful, such as the options I could have taken that would have prevented me from getting stuck in a traffic jam. This may improve patterns of driving, and so it is eminently rational. Byrne observes that we do spend time and effort on such matters.

Byrne and the literature she refers to limit discussion to cases of imagination that vary only within a narrow compass of existing circumstances, and only within reason. One omission is that she does not discuss more extreme departures from reality, as in dreams or in daydreams. And so this book, like the rest of the literature to which it belongs, ignores Freud's theory of dreams and of daydreams, as well as the criticism of it and the alternatives to it. This is regrettable, since Freud offers a view of dreams as rational: Dreams give vent to our secret wishes and thus reduce our sense of frustration that can be debilitating. Moreover, dreams may be very pertinent to Byrne's concerns, as we often exhibit some patterns in normal thinking and in dreaming.

A second omission is developmental: Child psychologists show interest in the stages of cognitive development when children learn to realize the difference between experience and dreams or imaginations. Taking these stages seriously, we may see not only the growth of the sense of reality, but also the growth of the ability to think abstractly. These two directions of development diverge, and therefore we keep feeling the need to balance realism with abstractness and the freedom of imagination. This is why we need to check our sense of reality and keep our imagination in check. The writings of Ludwig Wittgenstein interest those who shun wild imagination; those of Karl Popper interest those who court it. Byrne's concern with the rationality of imagination may lead her to study this disparity.

Reasons to act and the mental representation of consequentialist aberrations

doi: 10.1017/S0140525X07002592

Jean-François Bonnefon

CLLE, CNRS, and Université de Toulouse, Maison de la Recherche, 31058 Toulouse Cedex 9, France.

bonnefon@univ-tlse2.fr **http://w3.univ-tlse2.fr/ltc/bonnefon/**

Abstract: If imagination is guided by the same principles as rational thoughts, then we ought not to stop at the way people make inferences to get insights about the workings of imagination; we ought to consider as well the way they make rational choices. This broader approach accounts for the puzzling effect of reasons to act on the mutability of actions.

The idea that imagination is guided by the same principles as rational thoughts is developed in *The Rational Imagination* (Byrne 2005) into a rich and fruitful framework for the analysis of counterfactual thoughts. This framework, however, would probably benefit from a broader approach to rational thinking. The book focuses on one aspect of rational thinking, namely, the way people make inferences. Rationality, however, is not only about making sound inferences: It is also (and, some would argue, mainly) about making choices that will help one to achieve his or her goals (Baron 1988/2000; Evans & Over 1996).

The way people make such choices (and the way they believe other people make their choices) is likely to yield valuable insights – just as the way they make inferences

offers insights – into the workings of counterfactual imagination. As an illustration, I attempt to show how such a broader approach to rationality can explain the puzzling effect of reasons to act on the mutability of actions and inactions, which is considered in the book on several occasions.

This explanation requires a few assumptions. The first assumption is that the effect is real. That assumption is not as trivial as it may seem, because most of the evidence for the effect is indirect, at best. Reasons to act are known to impact the *regret* people feel about actions and inactions, and this impact is commonly assumed to be mediated by counterfactual thoughts; but direct demonstrations of the effect are scarce. However, from the new data reported in *The Rational Imagination*, as well as recent data reported in Bonnefon et al. (2007), we can assume that a good reason to act does make an action less mutable.

Second, let's assume that when people think of the way they and others make rational choices, they draw on some lay notion of consequentialism. Broadly defined, consequentialism holds that decisions should be made on the basis of their consequences for the achievement of the goals of the decision maker. Now, a decision maker is aware of at least some of the consequences of her actions: Let us say that among these consequences, the ones that help her achieve her goals have positive utility and the ones that prevent her from achieving her goals have negative utility. Overall, some actions will have a negative net utility for the decision maker, whereas others will have a positive net utility. The lay notion of consequentialism that people are likely to entertain is the following: *Individuals do not knowingly undertake an action that has a negative net utility for them.* The possibility that someone would knowingly do something that ultimately compromises her personal goals – which, incidentally, may very well be altruistic goals – would be a consequentialist aberration. (That is not to say it does not happen; only that the folk conception of behavior considers is the exception rather than the rule.)

Third, let's assume that in a vast majority of cases, a "reason to act" essentially boils down to the belief that an action has a positive utility, or that the corresponding failure to act has a negative utility. Strong reasons to act signal comparatively higher positive utility, whereas weak reasons to act signal a positive utility that is not much higher than zero and that, in fact, might be treated as null. This framework does not discriminate between reasons "based on goals" and reasons "based on obligations." Obligations simply denote situations wherein it is clear to everyone, from widely shared social norms, that a failure to act would have unpleasant consequences: for example, blame, punishment, privation of rights, and so on.

From these assumptions, we can conclude that possibilities where one had a (strong) reason to act and did not, just as possibilities where one had a (strong) reason not to act but acted nevertheless, both correspond to consequentialist aberrations. Now, if counterfactual imagination is really guided by the same principles as rational thought, should not people disregard counterfactual possibilities that correspond to consequential aberrations? Such a principle would explain why people do not mutate actions when there was a reason to act, because they are not willing to consider the possibility of not acting when acting had positive utility – an irrational behavior according to their lay notion of consequentialism.

The idea that people disregard possibilities that correspond to consequentialist aberrations is also supported by data on human reasoning. Bonnefon and Hilton (2004) showed that when presented with conditional statements such as "If Didier takes up the new job, his life will improve in every respect," participants spontaneously inferred that "Didier will take up the new job," as if they discarded from their representation of the conditional all the possibilities where Didier would act against his own interests and not take up the job. Similarly, when presented with premises like "If Mary goes to the party, she buys a new dress; if Mary buys a new dress, she cannot pay the electricity bill; Mary goes to the party," participants refused to endorse the valid conclusion that "Mary buys a new dress." Again, they seemed to discard from their representation of the premises all possibilities where Mary would act against her own interests and end up unable to pay the bill.

Thus, it would appear that the principles of rational decision making, at least in the guise of some lay notion of consequentialism, can act as a filter on the possibilities people are willing to consider, both when they make inference and when they exercise counterfactual imagination. The argument that imagination is guided by the principles of rational thought is a clever one, but rational thought should not be limited to inference making. Broadening its notion of rationality, to encompass decision making as well as reasoning, would provide *The Rational Imagination* with an even more fruitful framework to investigate the parallels between rational and creative thoughts.

Counterfactuals in science and engineering

doi: 10.1017/S0140525X07002609

Sanjay Chandrasekharan and Nancy J. Nersessian

School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA 30332-0280.

sanjayan@cc.gatech.edu **nancyn@cc.gatech.edu**
http://www-static.cc.gatech.edu/~sanjayan/
http://www.cc.gatech.edu/~nersessian/

Abstract: The notion of mutation is applicable to the generation of novel designs and solutions in engineering and science. This suggests that engineers and scientists have to work against the biases identified in counterfactual thinking. Therefore, imagination appears a lot less rational than claimed in the target article.

Our research focus is the generation of novel solutions and discoveries in engineering and science. The generation of counterfactual scenarios is central to these areas, and we analyze this process as involving the simulation of mental models, often in conjunction with built models (simulative model-based reasoning) (see Nersessian 2002; in press). Kahneman and Tversky (1982) proposed simulation as the mechanism underlying counterfactual thinking, but Byrne's account in *The Rational Imagination* (Byrne 2005) does not elaborate on this aspect. It remains unclear how simulation in counterfactual thinking and simulation in science and engineering are related.

One possible way simulation in the two domains could be related is via the processing of logical implication, which involves imagining counterfactual scenarios. This use of counterfactuals in logical processing seems to show that (even) logic needs imagination, rather than that imagination is rational. The latter claim appears justified only because Byrne highly constrains her definition of counterfactual thinking, staying close to sentence-level processing. This definition covers a much more narrow space than the generation of creative problem solutions in engineering and science. Nevertheless, the notion of mutating a factual scenario seems to be applicable to both domains, so it may be fruitful to raise the following question: Do the factors that influence mutation in everyday situations also influence mutations in engineering and science? Our cautious answer is that they do, and this complicates the notion of a rational imagination.

Consider the following design problem: How can a cell phone understand context? Essentially, the phone should shift to vibration mode, or forward calls to voice mail, when the user enters places like libraries and classrooms. The phone should also block calls when the user is driving, but should allow calls if she is a passenger. A much-discussed solution uses the Global Positioning System (GPS) to discover the coordinates of

the cell phone, but it faces the thorny problem of inferring context from coordinates. A simple solution would be adding small policy-announcing devices, installed by buildings and by carmakers/owners, which "instruct" cell phones to shut up. We wondered why such devices don't exist, even though many cities have introduced fines for using cell phones while driving, and some charge heavy fines for phones ringing in opera halls. Note that such spaces usually have announcing devices for humans; for instance, big signs saying "Do not use cell phone." There are other such design problems where similar environment-based solutions have been ignored. We hypothesized that this is because adding *epistemic structures* (labels, color codes, shelf talkers, etc.) to the world is a readily available design strategy for humans, but it is less available for artifacts. To test this hypothesis, we developed problem scenarios involving humans and artifacts (cell phones, robots), where participants were asked to propose solutions to a design problem. We used two groups of student participants, one general and the other specialist (master's level engineering). Both groups achieved the same level of performance, proposing environment-based solutions for problems involving humans, but artifact-based solutions for the cell phones and robots (see Chandrasekharan 2005).

Based on counterfactual thinking research, this bias in mutating the environment could be because of two related reasons. One, participants perceive the environment as more controllable in the case of human problems, and artifacts as more controllable in the scenarios that focus on these. Second, there could be an actor/observer difference (Kahneman & Miller 1986), where, in the case of human problems, the participants take an actor perspective and "simulate" the humans, and this leads to treating the environment as controllable. For artifacts, the participants take an observer perspective, and this leads to a fixation on the artifacts and on their possible mutations. To test this hypothesis, we gave the artifact scenarios to another group of participants and asked them to think of themselves as cell phones/robots (simulate the artifacts) while solving the problems. The number of environment-based solutions increased significantly in this case.

This seems to suggest that the biases in everyday counterfactual thinking also operate in design thinking, such as a preference for changing controllable events and actions. A similar bias could exist in the case of science. For instance, think of a biochemist trying to block the expression of a complex protein. Will she *prefer* to manipulate the actions she observes (e.g., use an antagonist to block a binding), or the inactions (e.g., use an agonist to activate another action in the cell), or a third option that involves neither or a combination of these choices? In theoretical research, are foundational assumptions (such as the currently questioned Weismann barrier in genetics) treated as immutable, because they are perceived as anchors, or as similar to forbidden possibilities? Closer to our research, do clinical researchers preferentially generate pharmacological solutions, rather than biomedical engineering solutions, because the former are more available? How can the latter be made more available? Such questions have not been raised in science and engineering, even though these fields deal with counterfactual scenarios on an everyday basis. We believe applying the insights from counterfactual thinking to these areas would prove valuable.

A more general question raised by this line of inquiry is whether mutation is a general process or a specialist one. Chapter 8 of the book, and the claim that logical implications are processed using counterfactuals, seem to suggest it is general. This raises the possibility that *all* counterfactual scenarios, including non-rational ones such as hallucinations, are generated using mutation. So the biases underlying everyday counterfactuals would be involved in these as well. This would mean that imagination is not rational as claimed, but rather, that it is a general *mechanism*, similar to, say, recursion, which is used in all situations. Further, this general status of mutation raises the possibility that in science and engineering, good solutions arise because scientists and engineers have developed ways of *overcoming* these biases (such as building simulations to explore the parameter space exhaustively, or using biological design as inspiration for engineering design). In other words, science and engineering, which stand right beside deductive logic as paragons of rationality, have to *work against* the biases in counterfactual thinking. This would make imagination still less rational.

What we imagine versus how we imagine, and a problem for explaining counterfactual thoughts with causal ones

doi: 10.1017/S0140525X07002610

Winston Chang and Patricia Herrmann

Department of Psychology, Northwestern University, Evanston, IL 60208.
winston-chang@northwestern.edu
p-herrmann@northwestern.edu
http://www.wcas.northwestern.edu/psych/

Abstract: Causal and counterfactual thoughts are bound together in Byrne's theory of human imagination. We think there are two issues in her theory that deserve clarification. First, Byrne describes which counterfactual possibilities we think of, but she leaves unexplained the mechanisms by which we generate these possibilities. Second, her exploration of "strong causes" and enablers gives two different predictions of which counterfactuals we think of in causal scenarios. On one account, we think of the counterfactuals which we have control over. On the other, which counterfactuals we think of depends on whether something is a strong cause or an enabler. Although these two accounts sometimes give the same predictions, we present cases in which they differ, and we would like to see Byrne's theory provide a way of reconciling these differences.

We offer two criticisms of Ruth Byrne's treatment of causal and counterfactual thinking in *The Rational Imagination* (Byrne 2005). The first is that she does not explain how we mentally generate some possibilities and avoid others. The second is that there are contradictory explanations in her discussion of "strong causes" and enablers.

Before diving into the discussion, it will be helpful to lay out some of Byrne's terminology. *True* possibilities are those which are consistent with a set of premises; generally, the premises are a person's beliefs about the world. Therefore, when speaking of future events, a true possibility is one that could happen, and a false possibility is one that could not. When speaking of past events, a true possibility is one that actually happened, and a false one is one that did not. A *counterfactual* possibility is one that once was true but is now false.

For the purpose of this discussion, we draw another distinction, which we will call *correct* and *incorrect* possibilities. *Correct* possibilities are those that are true or were true in the past (this includes counterfactuals). *Incorrect* possibilities are those that were never true. According to Byrne's theory, people think of correct possibilities – the true and the counterfactual – but we do not think of incorrect possibilities.

The first issue is that Byrne provides no explanation for how we generate correct possibilities while avoiding incorrect ones. Simple cases where subjects are given if–then statements can be handled by an algorithm that generates three of the four possibilities. In most cases, however, the problem is more complicated: we infer possibilities from our understanding of how the world works. In Chapter 5, Byrne describes counterfactual thoughts expressed in the media in the aftermath of the September 11, 2001, attacks. Many start like this: "If only the hijackers had been prevented from getting on board ..." Here is one that people tend not to think of: "If only the Al-Qaeda network did not exist ..." Byrne describes which counterfactuals we

think of, but she does not explain how we connect these counterfactual antecedents to the consequent: "... then the attacks would not have occurred." Nor does she explain how we avoid incorrect counterfactual antecedents such as, "If only there were more police at the World Trade Center ..." or, "If only there were purple elephants ..."

Byrne catalogs which counterfactual thoughts we have and describes a general structure to them, but does not explain *how* we generate them. This is analogous to a distinction made in biology: cataloging features of animal species (some finches have long, narrow beaks and others have short, stout beaks) versus explaining how those features came about (natural selection). The former is important, but the latter is the theoretical foundation – and it is missing here. What are the mechanisms underlying counterfactual thought? An example answer in the causal domain is that we use Bayes' nets to generate counterfactuals (Pearl 2000). (Bayes' nets, however, would not make the same predictions as Byrne's theory without a lot of extra machinery.) It might be too much to expect Byrne to commit to a particular theory of underlying mechanism, but we would have liked this book to give some hints here. After all, on page 1, she states, "This book is about how people imagine alternatives to reality," not just *which* alternatives they imagine.

The second issue concerns Byrne's discussion of "strong causes" and enablers. Imagine that you take a new route on your drive home, and on your way a careless driver swerves into your path, resulting in a crash. People tend to think, "If only I had driven home by a different route ...." Byrne explains this as follows:

> People mentally represent the strong causal relation by thinking initially about just one possibility, the co-occurrence of the cause and its outcome, whereas they mentally represent the enabling relation by thinking about the enabler and its outcome, and they can readily think about the absence of both. Accordingly most people focus on enablers (and disablers) in their thoughts about what might have been because the second possibility provides a ready-made counterfactual alternative. (Byrne 2005, pp. 118–19)

In the next paragraph, she writes:

> Causes occur – lightning strikes, careless drivers swerve, and terrorists formulate campaigns of destruction – and wishing the cause did not occur may be a remote and implausible hope. But enablers can be mentally deleted in an imagined alternative: dry leaves removed, alternative routes home taken, airport security improved. Wishing that whatever could have been done had been done to prevent the outcome or promote a better one may be a plausible alternative. (p. 119)

These two paragraphs offer two different explanations:

1. Strong causes seem immutable because we think of only one possibility, while enablers seem mutable because we think of two possibilities. (First paragraph quoted above.)

2. We think of alternatives when we have control over them. (Second paragraph quoted above.)

These explanations happen to agree in Byrne's examples, but there are many cases where they pull apart. Byrne presumes that we generally cannot control strong causes whereas we can control enablers, but this doesn't seem right. If it were, then we would never view our actions as strong causes when other possibilities are readily available; for example, putting sandals on your feet instead of shoes would merely *enable* the sandals to end up there. Byrne could take a hard line and say that all controllable actions are mere enablers, but this contradicts normal language; it would transform her theory into a normative one with respect to causes and enablers.

What happens if the strong cause is within one's control and the enabler is not? Imagine that you drive drunk and crash into someone who is taking a new route home. Here the strong cause is your driving drunk, and the enabler is the person taking the new route home. According to Byrne's first explanation, you would not think of an alternative to the strong cause, but you would think of an alternative to the enabler, so you would think, "If only he hadn't taken a new route home." But according to the second explanation, you would think of alternatives that you have control over, so you would think, "If only I hadn't driven drunk." Which explanation is correct? In the stories Byrne uses, the two explanations happen to make the same predictions, so we can't tell. We would like to know what happens when the explanations disagree, as they do here. We suspect that, for counterfactual thoughts, controllability is the more important factor.

Three steps to rational imagining?

doi: 10.1017/S0140525X07002622

Jennifer Church

Department of Philosophy, Vassar College, Poughkeepsie, NY 12604.
church@vassar.edu **http://faculty.vassar.edu/church/**

Abstract: Ruth Byrne presents a three-step argument to the conclusion that counterfactual imagining is rational. Insofar as this argument is valid, the conclusion is weaker than it seems. More importantly, it does not represent the central contributions of this book – contributions that, if anything, point instead to what is *ir*rational about counterfactual imagining.

At several points in *The Rational Imagination*, Byrne (2005) presents her argument as a three-step (or three-premise) argument to the conclusion that counterfactual imagining is rational. Insofar as this argument is valid, the conclusion is weaker than it seems. Moreover, it does not represent the central contributions of this book, which, if anything, point instead to what is *ir*rational about counterfactual imagining.

1. Human reasoning is guided by rational principles. Byrne begins with the premise that human reasoning is rational – a premise that she explicates as the claim that human reasoning is based on rational capacities or principles in much the same way that human language is based on grammatical capacities or principles. Even though our actual reasoning performance (like our actual speaking performance) is often flawed, that is due to such factors as the constraints on our working memory, not to any flaws in the cognitive mechanisms that define our underlying competence. Here Byrne is aligned with Jonathan Cohen (1986), among others, who insists that human irrationality occurs at the level of performance, not competence; but she does not offer any arguments for this position.

The main alternative to this approach claims that human reasoning is based, instead, on various heuristics – heuristics that are not rational in the logical, truth-preserving sense, yet are usually useful in the contexts humans find themselves in – contexts in which time and energy are limited, contexts in which some possibilities are much more likely or much more significant than others, and so forth. This alternative has little need for a competence/performance distinction since the specified heuristics are supposed to be nature's way of factoring our limitations into our design. (The analog, in the case of language, locates the grammar of a language in the way people actually talk, rejecting the notion that there is a deeper and more competent grammar that underlies our speech, against which our flawed performance may be measured.) Whether such heuristics constitute a different, or better, type of rationality than that defined by logical principles will determine whether or not advocates of this alternative would agree with Byrne's first premise as stated.

2. Human reasoning is guided by imagined possibilities. In at least some summaries of her argument (pp. 198–99; 215), Byrne's second premise states that our reasoning depends on thinking about, or imagining, possibilities. (*Thinking* about possibilities is effectively equated with *imagining*

possibilities – particularly when the possibilities in question are mere possibilities. Nothing turns on whether the imagining in question involves imagistic representations.) This is a claim that Bryne and Johnson-Laird have been defending for some time (and one which, as she notes, has a long philosophical lineage). Here the main alternative is a version of the inference rule theory – the theory that our inferences are guided by more or less complex sets of rules that tell us how to proceed from premises to conclusions; that our reasoning consists in applying such rules rather than actively considering a range of possibilities. There are several points in Byrne's book where she argues against this alternative (e.g., pp. 51–52 and pp. 115–16), but these passages are fairly brief and inconclusive. This book does not (primarily) address that dispute.

Note how the above two premises combine in the case of deductive reasoning, according to Byrne:

> Deductive reasoning is rational because people have the underlying competence to think of all the relevant possibilities so that they could search for counterexamples. Why then do people make mistakes? Their performance is sometimes not rational because of the limits to the possibilities they can consider. (p. 29)

From what I have described so far, one might expect Byrne's project to be a detailed account of just how limits on our imagination constrain our reasoning performance – how our imaginations select from the totality of logical possibilities in order to produce our (often flawed) reasoning performance. And this is precisely what I think she actually does. Chapter by chapter, she describes particular ways in which we limit the logical possibilities that we imagine when reasoning about what actions or conditions would or would not have made a difference, what ought to have happened, what we regret, and what we deem inevitable. These are interesting, useful, and nicely documented observations, deserving of close attention and continued discussion.

Byrne's stated position, however, is something different, for the conclusion of each version of her three-step argument (and what she repeatedly calls the central idea of the book) is this:

Conclusion: *Counterfactual imagining is rational*

Given the severe restrictions on the sets of possibilities that we are said to imagine when contemplating what would have made a difference, or what we should have done, and so on, and given Byrne's equating of deductive competence with an ability to imagine all relevant possibilities (and thus all possible counterexamples), this is a surprising conclusion.

3. Counterfactual imagining is guided by the same principles as those that guide human reason and imagining possibilities. Byrne's crucial third premise, which is stated in a number of different ways (compare versions on pp. 38, 199, 208, and 215), claims that the principles that guide human counterfactual imagining – which is a subset of imagining possibilities more generally – are the *same* principles as those that guide human reasoning. Since premise 1 affirms the rationality of these principles, it is now an easy step to the conclusion that counterfactual imagining is rational.

As we have already noted, though, the most that this would establish would be our *competence* for rational counterfactual imagining; it would not ensure the rationality of our actual counterfactual imagining. And given Byrne's careful detailing of the many possibilities that most of us do not (usually) consider in counterfactual imagining (possibilities that are less controllable, possibilities that are forbidden, possibilities that are in the more distant past, etc.), it is clear that our actual performance falls far short of our underlying competence. If the third premise is understood as referring to the principles that characterize our rational competence, it would seem more appropriate to conclude that counterfactual imagining is *ir*rational – not in principle, but in fact.

On the other hand, if the principles that guide counterfactual imagining are captured in the list of principles that Byrne articulates, chapter by chapter (see Table 7.2, p. 161, for the complete list), then there is little reason to think that these principles (e.g., "People keep in mind *few* possibilities") are the principles that are constitutive of our rational competence. They may be efficient or instructive or reassuring in some of the ways that Byrne sketches (on pp. 209–12), but they do not take account of all relevant alternatives. If, when considering how things might have been different, we restrict ourselves to imagining changes in controllable factors only, or changes in only the most recent events, surely we are not fully exercising our capacity for rationality.

If, as seems plausible, human reasoning rightly relies on a combination of logical principles and pragmatic principles, then it is not surprising that counterfactual reasoning and counterfactual imagining also rely on a combination of logical and pragmatic principles. That is not a surprising conclusion, and it is not where the real interest of Byrne's book lies. The most significant contribution of this book is her description of the ways in which certain possibilities are usually *not* imagined when we reason with counterfactuals – ways in which our rationality is, understandably, limited.

Beyond rationality: Counterfactual thinking and behavior regulation

doi: 10.1017/S0140525X07002634

Kai Epstude and Neal J. Roese

Department of Psychology, University of Illinois at Urbana–Champaign, Champaign, IL 61820.

epstude@uiuc.edu **roese@uiuc.edu**

Abstract: Counterfactual thinking may be described as disciplined, realistic, and rational, but we move a step further to describe a theoretical perspective centering on behavior regulation. According to this perspective, counterfactual thinking primarily centers on coordination of ongoing behavior. In short, most "if only" thoughts in daily life center on the acquisition of goals; hence, counterfactual thinking may be illuminated by considering the large literature on goal cognition.

In her book *The Rational Imagination*, Byrne (2005) describes some of the cognitive mechanisms underlying counterfactual thinking. Her guiding theoretical framework is informed primarily by the mental models tradition (Johnson-Laird 1983) but also by norm theory (Kahneman & Miller 1986), which emphasize the lower-order building blocks of cognition. Byrne's main claim is that counterfactual thinking is inherently rational, which is to say disciplined, rooted to accurate inferences about reality rather than mere whimsy. We agree completely with this claim, but suggest that it perhaps does not go far enough. Counterfactual thinking is not only rational, but motivated (and motivating). Connected deeply to goal-oriented cognition, counterfactual thoughts contribute to the effective management of ongoing behavior. Bundled under a functional theory of counterfactual thinking (Epstude & Roese, submitted; Roese 1994; 1997; Roese & Olson 1997), we argue that taking into account the motivational and regulatory basis of the imagination helps to explain key findings about which the mental models tradition remains silent.

Consider the following empirical observations. Counterfactual thoughts are idealistic: they are more likely to focus on how the past might have been better than on how it might have been worse (Nasco & Marsh 1999; Summerville & Roese, in press). Counterfactual thoughts are situationally reactive: they are more likely to appear after failure than after success (Roese & Hur 1997). Counterfactual thoughts are problem-focused: they are more likely to focus on fixing a problem than on random life events (Roese et al. 1999). Counterfactual thoughts are

egocentric: they are more likely to focus on the actions of oneself than on those of other people (White & Roese, submitted). Finally, counterfactual thoughts involve activation of the orbitofrontal region of the brain, a region previously linked to planning and problem solving (Coricelli et al. 2005; Ursu & Carter 2005). What do these observations tell us about counterfactual thinking?

Taking these findings in hand, it becomes clear that the abundant research on goal pursuit (Carver & Scheier 1998; Fishbach & Ferguson 2007; Higgins 2006; Lewin 1935) provides us with important insights into the form, function, and effect of counterfactual thinking. Goals may be defined as cognitive representations of desired ends and the means to achieve such ends (Fishbach & Ferguson 2007). Counterfactual thoughts that occur in everyday life involve, for the most part, alternative means that "might have been" implemented so as to have obtained a desired end. We recently described how principles of motivation and goal cognition might explain counterfactual thinking (Epstude & Roese, submitted). Encountering a problem typically triggers an upward counterfactual (e.g., "If only I had studied harder, I would have passed"). Counterfactual thoughts themselves have as an inherent property such causal implications, and these directly fuel the activation of corresponding behavioral intentions ("I intend to study harder next time"), which in turn unleash corresponding corrective behavior (the student indeed studies harder the next time). To the extent that such behavior alleviates the original problem, this mechanism is effective in terms of regulating behavior in terms of goal pursuit. This regulatory mechanism is content-specific; that is, the information contained in the counterfactual directly translates into a related action.

In addition to a content-specific mechanism by which counterfactual thinking influences behavior, evidence also suggests a content-neutral mechanism. A content-neutral mechanism reflects *how* rather than *what* information is handled. For example, independent of their specific meaning, counterfactuals can exert an influence on attention and information processing, as in demonstrations of a counterfactual mind-set, which involves a heightened albeit generic tendency to consider alternatives (e.g., Galinsky et al. 2000). As another example, the negative affect that often springs from upward counterfactuals (which make the present look less desirable in contrast to a better alternative) may itself motivate behavior change (Markman et al. 2006). In addition, structural properties of counterfactual thoughts may evoke either approach or avoidance motivation (e.g., Roese et al. 1999).

The interplay between emotion and counterfactual thinking is pivotal. Regret is an unpleasant feeling state that depends on an upward counterfactual, an aching despair born of the realization that one might have made a better decision or achieved a better outcome (Roese 2005). People are motivated to manage their regret experiences even as they draw insights from their regrets (Zeelenberg & Pieters 2007). Moreover, recent studies have linked mental health dysfunction to both an excess and a deficit in counterfactual thinking and regret. The principal consequences of upward counterfactual thinking (i.e., regret) are problem-solving insights and negative emotion; hence, excessive counterfactual thinking has been found to be associated with pathology rooted to excessive problem-focused cognitions (e.g., anxiety; Kocovski et al. 2005) and excessive negative affect (e.g., depression; Markman & Miller 2006). By contrast, a deficit of counterfactual thinking is associated with a deficit of problem-focused cognition (e.g., underachievement, work difficulty, social dysfunction) along with an absence of negative affect. Along these latter lines, schizophrenia has been shown to be associated with impaired counterfactual thinking (Roese et al., in press) and deficits in goal-related cognition (Brandstätter et al. 2001). These studies suggest that there is an optimal level of counterfactual thinking and emotional reactivity to such inferences, and that both too much and too little may spell trouble for mental health.

An earlier generation of research on counterfactual thinking, dating from the 1980s and stimulated by the writings of Kahneman and Tversky (1982), treated such thoughts as instances of bias, and hence, impediments to sound judgment and shrewd action. The work of Byrne and others has illuminated counterfactual thinking in a different light, as an instance of principled and rational imagination. Counterfactual thoughts do sometimes bring bias, yet balancing this cost is the larger benefit of the effective management of daily behavior. Counterfactual thinking, we argue, is best understood as an input to course correction, as an instance of goal cognition, and as an essential component of behavior regulation.

Semifactual: Byrne's account of even-if

doi: 10.1017/S0140525X07002646

Simon J. Handley[a] and Aidan Feeney[b]

[a]*Centre for Thinking and Language, Department of Psychology, University of Plymouth, Drake's Circus, Plymouth PL4 8AA, United Kingdom;* [b]*Department of Psychology, Durham University, Queen's Campus, Stockton-on-Tees TS17 6BH, United Kingdom.*

shandley@plymouth.ac.uk **aidan.feeney@durham.ac.uk**
http://www.plymouth.ac.uk/pages/dynamic.asp?page=staff details&id=shandley
http://www.dur.ac.uk/aidan.feeney

Abstract: Byrne's approach to the semifactual conditional captures the reasoning data. However, we argue that it does not account for the processes or principles by which people arrive at representations of even-if conditionals, upon which their reasoning is said to be based. Drawing upon recent work on the suppositional conditional we present such an account.

In choosing to write a book primarily about counterfactuals, Ruth Byrne has been able to integrate the literatures on reasoning and social cognitive aspects of decision making and judgment. In so doing, she has performed an invaluable service to cognitive and social psychologists alike. Thus, there is much to applaud in *The Rational Imagination* (Byrne 2005) and there are several of its aspects upon which we could comment. Because we have commented elsewhere on Byrne's application of mental model theory to judgmental phenomena such as the action effect (see Feeney & Handley 2006), in this commentary we focus on the mental model account that she outlines of how people reason about semifactual conditionals.

According to Byrne, a subjunctive semifactual conditional such as

(1) Even if Pete had studied hard he would have failed the exam

usually conveys the conjecture that its antecedent is false and its consequent true, and is often used to assert that the antecedent could not have prevented the consequent from occurring. Even-if conditionals call for the representation of two possibilities: one where the antecedent occurs and the consequent occurs, and one where the antecedent does not occur and the consequent occurs. If we consider the foregoing example, people represent the conjecture, "He studied hard and failed the exam," and they represent the presupposed facts, "He didn't study hard and failed the exam." This mental representation is said to explain people's tendency not to affirm the consequent (i.e., that he studied hard, from being told that he failed the exam) and to infer the opposite to the standard conclusion (i.e., that he failed rather than passed the exam) from a denial of the antecedent (see Handley & Feeney 2004; 2007; Moreno-Rios et al. 2004). It also provides an explanation for one of the most intriguing characteristics of concessive conditionals – their compelling invitation to the listener to infer the consequent, a characteristic much commented on in linguistics and philosophy (Konig 1986).

It has been recognised in these literatures that the study of less common conditional forms can provide significant insights into

the way in which the ordinary conditional is represented and processed. Consequently, the study of how people reason and think about even-if is a valuable enterprise and Byrne's analysis provides an interesting, though in our view incomplete, framework. The principal problem is that the account as presented is purely descriptive, and lacks clear principles or a detailed mechanism that can explain how people arrive at a representation corresponding to the possibilities that Byrne describes. In this commentary we present an alternative account of even-if, which is grounded in linguistic pragmatics, and we contrast this with Byrne's model. The account details how *even* serves to modify the representation of the conditional connective *if*, and draws upon recent work on the suppositional conditional (Evans et al. 2003; Handley et al. 2006).

Understanding even-if requires a consideration of the function of *even* in everyday natural language. Consider the assertion in conditional 2:

(2) Even Tony distrusts George

Several philosophers (Jackson 1987; Sanford 1989) have suggested that *even* serves to deny an available presupposition, for example, that we might expect Tony to trust George. It serves to pick out an extreme position, and calls up a range of contextually determined alternatives that are less surprising; for example, that Gordon distrusts George, Hilary distrusts George, or Jacques distrusts George. In so doing, the utterance invites the listener to infer that George is a man not to be trusted.

In the study of different conditional constructions it is important to consider how the linguistic terms interact in determining meaning. Understanding how *even* interacts with *if* also requires an account of the conditional connective. According to the suppositional account, conditionals cue a mental simulation (often referred to as the Ramsey test) in which the listener imagines that the antecedent condition holds and evaluates their degree of belief in the consequent in that context (Evans et al. 2005). For example, consider the following conditional:

(3) If the United States cuts fuel emissions then global warming will be reduced

This assertion cues us to suppose that the United States cuts their emissions, and on the basis of this supposition, together with background beliefs, we can evaluate our belief that global warming will in fact be reduced. The suppositional account predicts that belief in a conditional is closely related to conditional probability (P[q/p]), a prediction confirmed in numerous recent studies (Evans et al. 2003; Over et al. 2007). Of course one might not believe that there is any sort of relationship between U.S. fuel emissions and global warming and imagine that global warming will increase irrespective of U.S. policy, which creates a perfect opportunity for asserting a concessive conditional:

(4) Even if the United States cuts fuel emissions global warming will increase

Combining the analysis of *even* with our account of *if*, the concessive in conditional 4 denies the presupposition in conditional 3, and calls up a range of alternative conditionals on a probability scale that are less surprising or unexpected, and where, in probabilistic terms, P(q/p) is higher. Often, as Jackson (1987) has pointed out, the scale will consist of the conditional as in (4) above, with its antecedent negated:

(5) If the United States doesn't cut fuel emissions global warming will increase

The combination of the conditional in (4) with the conditional in (5) leads directly to the inference that global warming will increase, irrespective of U.S. policy. It is important to note here that this inference is both logically (through *constructive dilemma*) and probabilistically valid, but that it is not the result of representing discrete possibilities, consistent with the initial conditional, as Byrne's account claims. What is important about our account, and contrasts clearly with Byrne's, is that we specify how *even* and *if* combine in cueing the activation and representation of alternative associations that can then be integrated with a representation of the original assertion in order to make inferences. Unfortunately, the details of such a mechanism are absent in Byrne's account.

A second important point of contrast between our account and Byrne's is that, because ours is rooted in linguistic pragmatics, it is naturally extended to inferences about speaker intentions. For example, we have presented people with even-if conditionals and asked them to make immediate inferences about whether the speaker intends to carry out the action specified in the antecedent (Handley & Feeney 2004; 2007). In general, people seem to infer from these conditionals that the speaker does not intend to carry out the antecedent action. So when presented with an even-if assertion of the following kind:

(6) Even if I study hard I will fail the exam

participants infer that the speaker does not intend to study. This finding suggests first that people's representations of the utterance might include information about the cost of the antecedent action, and second that even-if conditionals serve an important rhetorical function. Our intuitions here are, no doubt, related to Byrne's claims about the role that semifactual conditionals play in denying a causal link between antecedent and consequent. One can justify a decision not to study by implying that in this instance studying will not cause one to pass an exam.

In conclusion, although we welcome Ruth Byrne's highly integrative book and her original analysis of even-if and the semifactual conditional, we disagree with some of the details of her account. We very much look forward to debating these and other issues with her in the coming years.

The goals of counterfactual possibilities

doi: 10.1017/S0140525X07002658

Paolo Legrenzi

School of Advanced Studies, Dorsoduro 3488u rio novo, 30123 Venice, Italy.
legrenzi@iuav.it

Abstract: Why do humans imagine alternatives to reality? The experiments conducted by Byrne explain the mental mechanisms we use when we do just this – that is, imagine one, or more, alternative reality. But why do we do this? The general reason is to give ourselves an explanation of the world, to tell stories; at times to console ourselves, and at times to despair. A good story is not only based on a description of what happened, but also hints at, or explains, what might have happened. Depending on our aim, we construct different kinds of counterfactuals. In all cases, however, we are faced with constraints. These constraints are specific to a given domain of beliefs and use of counterfactuals.

The goal of historians consists in constructing counterfactuals that deviate as little as possible from what really happened. Conjectures of the type "if only . . ." must make the least amount of changes possible to stories, in order to isolate the essential causes of what happened. From the kind of "if only" conjecture a historian uses, we understand the structure of his or her explanations and the nature of the causes that he or she identifies for a given episode in the past (Tetlock et al. 2005). When, on the contrary, we construct alternatives to reality in order to invent new technologies or scientific theories, we are constrained by the criteria which the community we belong to accepts. As is the case with historians, we cannot work from fantasy.

Fantasy is nothing other than the imagination restrained by fewer constraints. When we create a story of fantasy, science fiction, or magic, we can violate certain principles of reality, but do not arrive at the absurd because it would not serve our purposes. For example, according to the principles of naïve physics, we are able to move at very high speeds, fly, travel backwards and forwards in time, disappear, make difficult calculations, preserve excellent memory, and transform objects.

Nevertheless, we always begin from characteristics of the world with which we are familiar. What we do is imagine powers that we do not have, but that are conceivable. So much so that these powers usually belong to other nonhuman entities: speed (missiles), going backwards in time (history books), transforming objects (technologies), calculations and memory (computers), disappearing (complete mimicry of certain animals), and so forth.

When the purpose is to console ourselves or others, in order to encourage or deceive ourselves, or to make ourselves feel guilty, we resort to the notion of control. We construct counterfactuals to demonstrate that avoiding a certain situation was beyond our power, or that it was in our power, but we are to blame for not having avoided it. To exalt in ourselves, instill hope, or deceive ourselves, we imagine alternative worlds that are "worse off" without our intervention. To make ourselves feel guilty, we imagine that if we had not existed, "better" alternative worlds would have.

In childhood, we are unable to create counterfactuals with respect to the contents of our mind. Children believe that what is inside their minds and what is inside the minds of others is the same. For this reason a child cannot believe that others have false beliefs (they have the same beliefs as the child and these beliefs are true), until he or she reaches full development between the ages of one and five (Surian et al. 2007). Adults become so sophisticated that they are able to comprehend statements such as the one made by the former Federal Reserve chairman, Alan Greenspan: "I know you believe you understand what you think I said, but I am not sure you realize that what you heard is not what I meant" (Resche 2004, p. 731).

Ruth Byrne says that "people do not tend to imagine 'miracle-world' alternatives" (Byrne 2005, p. 191). However, if the goal is to construct a religion, it is better to do just that. "Religions are costly, hard-to-fake commitments to a counterintuitive world of super-natural causes and being" (Atran 2002, p. 264). The experiments of Scott Atran on the impact and memorizability of intuitive and minimally counterintuitive beliefs show that the delayed one-week recall presents the following sequence of remembering: intuitive and ordinary, intuitive but bizarre, minimally counterintuitive, and maximally counterintuitive. The best type of counterfactual to use for religious believers is the intuitive but bizarre domain of beliefs: for example, floating pencil, dangling cat, blinking newspaper. These are the couplings typical of miracles: walking on water, restoring sight to the blind, multiplying food, curing the ill. From the viewpoint of naïve physics, such things are impossible, but believable. In earthly matters, which are matters of explanations and calculation, one prefers a probable counterfactual to an improbable one for explaining things. In unearthly matters, an improbable belief is worse than an impossible belief for building faith in a religion. It is on such grounds that Oscar Wilde (1889/1989, p. 990), another great Dubliner like Ruth Byrne, criticized the Church of England and anticipated the experiments of Atran: "The growth of common sense in the English Church is a thing very much to be regretted. It is really a degrading concession to a low form of realism. It is silly, too. It springs from an entire ignorance of psychology. Man can believe the impossible, but man can never believe the improbable."

Differential focus in causal and counterfactual thinking: Different possibilities or different functions?[1]

doi: 10.1017/S0140525X0700266X

David R. Mandel

Defence Research and Development Canada (Toronto), Toronto, Ontario, M3M 3B9, Canada.

David.Mandel@drdc-rddc.gc.ca

http://mandel.socialpsychology.org/

Abstract: In *The Rational Imagination*, Byrne proposes a mental models account of why causal and counterfactual thinking often focus on different antecedents. This review critically examines the two central propositions of her account, finding both only weakly defensible. Byrne's account is contrasted with judgment dissociation theory, which offers a functional explanation for differences in the focus of causal and counterfactual thinking.

In *The Rational Imagination: How People Create Alternatives to Reality*, Byrne (2005) offers an account of why the contents of causal and counterfactual thoughts often diverge. Her account, based on the principles of mental models theory (MMT; Johnson-Laird & Byrne 2002) has two central propositions. First, counterfactual and causal thoughts sometimes differ in content because the former tend to focus on enablers, whereas the latter tend to focus on strong causes. Enablers are necessary conditions for the occurrence of an effect ($C \Rightarrow A$), whereas strong causes are both necessary and sufficient for the effect ($A \Leftrightarrow C$). Second, enablers are consistent with three possibilities and tend to be represented by two, whereas strong causes are consistent with two possibilities and tend to be represented by one. Specifically, if antecedent A is an enabler of consequent C, then the possibilities $A \wedge C$, $A \wedge \neg C$, and $\neg A \wedge \neg C$ are consistent and the first and last possibilities will tend to be mentally represented. For causes, the first and last of the same possibilities are consistent and only the first will tend to be represented. In Byrne's account, then, the basis for divergence in the content of causal and counterfactual thinking is that the former focuses on necessary and sufficient antecedents, whereas the latter focuses on necessary (but not necessarily sufficient) antecedents.

I agree with Byrne on the basics; namely, we share common definitions of enablers and strong causes, and we agree on the possibilities that are consistent with each. Our psychological accounts of causal and counterfactual thinking, however, diverge sharply. A fundamental difference is that Byrne traces differential focus in causal and counterfactual thinking to differences in the possibilities represented by enabler and strong-cause concepts, whereas judgment dissociation theory (JDT; Mandel 2003c; 2005) traces the differential focus to functional differences in these related but nevertheless distinct forms of goal-directed cognition. In JDT, the primary function of counterfactual thinking about negative past outcomes is to identify acts or events, particularly personally controllable ones (Mandel & Lehman 1996), that would have been sufficient to prevent the actual outcome or consequentially similar outcomes had they been taken or had they occurred. In contrast, the primary function of causal thinking is to identify acts or events that were sufficient to bring about the outcome as it actually occurred under the circumstances. Byrne does not deny these functional differences – indeed, at points in her book, she alludes to them – but they remain on the periphery of her account.

Consider Byrne's proposition that counterfactual statements focus on enablers, whereas causal statements focus on strong causes. This distinction alone cannot explain the phenomenon of differential focus, because strong causes are, by definition, also enablers. That is, strong causes constitute a subset of C-enablers that are also sufficient to yield C. Hence, all strong causes should be candidates for counterfactual thinking, although some enablers would not be candidates for causal ascription. Byrne's first proposition begs the question of why counterfactual thinkers would focus on enablers that are not strong causes if the latter already meet the enabler criterion. If not functionally motivated, such behavior might seem a waste of scarce cognitive resources, perhaps even irrational.

Contrary to Byrne's account, the fact that counterfactuals meet the logical criterion for enabling seems to me largely incidental. Consider the statement, "If only the CIA hadn't botched their analyses, 9/11 would have been averted." According to

Byrne, this counterfactual signifies that "botching" was necessary for 9/11, with emphasis placed on the necessary condition for the generation of the actual disaster. According to JDT, the counterfactual means something quite different; namely, that the absence of (or a reduction in) botching would have sufficed to have prevented the disaster. The emphasis here is on foregone sufficient disablers rather than actual necessary enablers. In this view, such counterfactual conditionals represent a form of *satisficing* (Simon 1956) in which one identifies events, especially controllable acts, which would have been enough to undo a past failure. The emphasis on control in this account, sharing much in common with Collingwood's (1940) manipulation theory of causation, can also explain why counterfactuals often focus on factors other than strong causes even though the latter satisfy the enabling criterion – namely, because manipulability is, at best, a weak constraint on causal ascriptions. The theoretical focus on sufficient disablers rather than necessary enablers is also supported by literature indicating that people are biased toward sufficiency testing for adaptive reasons (Friedrich 1993; Klayman & Ha 1987) and tend to interpret causatives in terms of sufficiency (Mandel 2003c; Mandel & Lehman 1998; Wolff 2007).

Briefly, let me say a few words about Byrne's second central proposition, which links the distinction between enablers and strong causes to mentally represented possibilities. Her prediction that enablers (and, by extension, counterfactuals) conjure up possibilities $A \wedge C$ and $\neg A \wedge \neg C$, whereas strong causes conjure up only the former, fits the data. However, the reason for this prediction is unexplained. Indeed, the opposite prediction seems to me more plausible: If temporal order is preserved, as it tends to be in causal reasoning (Einhorn & Hogarth 1986), then only one of the two models is congruent with enabling ($\neg A \wedge \neg C$ as $\neg A \Rightarrow \neg C$), whereas both are congruent with strong causes ($A \wedge C$ as $A \Rightarrow C$ and $\neg A \wedge \neg C$ as $\neg A \Rightarrow \neg C$). As I have proposed elsewhere (Mandel 2003b), the reason why past-tense counterfactual conditionals appear to evoke two possibilities, whereas indicative conditionals tend to evoke only one, may be because the former are better than the latter at eliminating uncertainty. It is conversationally implied in counterfactual statements that both A and C, in fact, did not transpire. Thus, $\neg A \wedge \neg C$ is more than a mere possibility; it is an assumed fact. In contrast, indicative conditionals do not point to facts; only possibilities. Given that possibilities, not facts, constitute the basic units of mental representation in MMT, Byrne's account cannot accommodate this type of explanation.

In summary, Byrne provides a good overview of the mental models perspective on counterfactual thinking. In my own estimation, her book succeeds in presenting that account, even if the account itself reveals its own limitations.

NOTE

1. The author of this commentary carried out this research on behalf of the Government of Canada, and as such the copyright of the commentary belongs to the Canadian Crown and is not subject to copyright within the United States.

Counterfactuals need not be comparative: The case of "As if"

doi: 10.1017/S0140525X07002671

Keith D. Markman[a] and Matthew N. McMullen[b]

[a]*Department of Psychology, Ohio University, Athens, OH 45701;* [b]*Department of Psychology, Montana State University–Billings, Billings, MT 59101.*

markman@ohio.edu **mmcmullen@msubillings.edu**

http://markman.socialpsychology.org/

http://mcmullen.socialpsychology.org/

Abstract: Byrne (2005) assumes that counterfactual thinking requires a comparison of facts with an imagined alternative. In our view, however, this assumption is unnecessarily restrictive. We argue that individuals do not necessarily engage in counterfactual simulations exclusively to evaluate factual reality. Instead, comparative evaluation is often suspended in favor of experiencing the counterfactual simulation as if it were real.

Ruth Byrne's *The Rational Imagination* (2005) is an outstanding work that provides the clearest answers thus far to long-standing questions about counterfactuals such as "What is most mutable?" and "Why is this so?" Thus, in the spirit of broadening the perspective offered by Byrne, we will merely choose to quibble with a single phrase that appears in the book, specifically, "counterfactual thinking requires a comparison of the facts with the imagined alternative" (p. 122). In our view, this assumption, shared by most if not all scholars who study counterfactual thinking and conditional reasoning, is unnecessarily restrictive. Rather, we have suggested (e.g., see Markman & McMullen 2003; 2005; McMullen 1997; McMullen & Markman 2000), and continue to maintain, that individuals do not necessarily engage in counterfactual simulations exclusively to evaluate factual reality. Instead, comparative evaluation may be, and often is, suspended in favor of experiencing the counterfactual simulation as if it were real.

Markman and McMullen (2003) proposed a Reflection and Evaluation Model (REM) to account for how counterfactuals can have either contrastive (i.e., displaced away from a counterfactual standard) *or* assimilative (i.e., displaced toward a counterfactual standard) effects on affect and judgments. The model asserts that two psychologically distinct modes of mental simulation operate during counterfactual thinking: *reflection*, an experiential, "as if" mode in which counterfactual possibilities are vividly simulated, and *evaluation*, a comparative mode in which counterfactual standards are used as a reference point against which to evaluate factual reality. Reflection renders standard-consistent cognitions accessible, thereby yielding assimilation, whereas evaluation yields contrast because counterfactuals are used as a standard against which to compare factual reality.

To illustrate, consider how an individual who just missed being killed in a plane crash may experience a great deal of anguish by dwelling on what might have happened (assimilation) (see also Oettingen 1996), but at another time feel quite lucky by focusing on the fact that they are, by good fortune, alive today (contrast). Similarly, imagining having won the lottery might elicit positive affect by one's fantasizing about having a great deal of money (assimilation) so long as one suspends their evaluation of the fact that they do not have a great deal of money. On the other hand, evaluation of this fact in light of the counterfactual should lead one's perception of their present state to seem even more impoverished (contrast).

Byrne (2005) describes an interesting study conducted by Thompson and Byrne (2002) that examined whether there are differences among individuals in their willingness or ability to consider more than one possibility when confronted with a counterfactual conditional. Although the majority of participants kept in mind two possibilities – the presupposed facts and the counterfactual conjecture – about a quarter of the participants kept in mind only a single possibility – the presupposed facts – when they contemplated the conditional. In our view, however, this work neglects to mention a third way that individuals may entertain counterfactual assertions: keeping only the *false* possibility in mind. Moreover, engaging in such an "as if" type of simulation should have assimilative effects on subsequent responses and judgments. Is there empirical evidence to support such a claim?

An early and clear demonstration was provided by McMullen (1997), who asked participants to recall a somewhat negative event in their own lives and imagine how things could have turned out better (upward counterfactual) or worse (downward counterfactual) than they actually did. Participants in the reflection condition were then instructed to "vividly imagine what might have happened instead," whereas those in the evaluation

condition were instructed to "vividly imagine the event and what might have happened instead." Providing clear evidence for affective assimilation, participants in the reflection condition reported positive affect after making upward counterfactuals and negative affect after making downward counterfactuals, whereas this pattern was reversed in the evaluation condition, thereby evidencing affective contrast. Of course, it may be argued that comparison must still be the default effect of a counterfactual (cf. Roese et al. 2005), and that affective assimilation is only possible under specific conditions. However, McMullen's data are not consistent with that conclusion. When instructed to generate a counterfactual and then vividly imagine that counterfactual, participants exhibited an affective assimilation effect; contrast effects only emerged for those who explicitly evaluated their factual events. This suggests that counterfactuals are not contrasted with their corresponding factual events by default, but rather the explicit evaluation of the factual event must first be made.

How can a counterfactual, which is by definition an alteration of a factual event, not by default act as a contrast to that factual event? Put another way, have we chosen to define counterfactuals too broadly? Are they truly counterfactuals if individuals treat possibilities as fantasies (either positive or negative) rather than as standards of comparison? Consider that McMullen's (1997) participants were simply instructed to "think of how something different could have happened rather than what actually happened." Only following this counterfactual generation was the reflection/evaluation manipulation initiated. Thus, equivalent simulations were shown to yield both affective assimilation and contrast. In turn, one might argue that an obviously unreal alternative should be subjectively perceived as distinct from a real event, but the work of Johnson and Raye (1981) and Anderson (1984) suggests that the representations of real and imagined events share many characteristics, and that imagined events are sometimes confused with real events. Similarly, Gilbert et al. (1990) suggested that, in order to be comprehended, a proposition, even an obviously false one, must first be accepted and treated *as if* it were true. Hence, perhaps a counterfactual, before it can be used as a standard of comparison, must first be comprehended in a manner that treats that counterfactual as if it were true, and only then can it be used as an evaluative standard. In this way, then, an initial assimilation effect would later give way to a contrast effect. In all, we hope that drawing attention to "as if" processing of mental simulations can broaden the theoretical landscape so artfully portrayed in Byrne's *The Rational Imagination*.

Imagination as a source of rationality in development

doi: 10.1017/S0140525X07002683

Henry Markovits

Département de Psychologie, Université du Québec à Montréal, C.P. 8888, Succ Centre-Ville, Montréal, QC H3C 3P8, Canada.

henrymarkovits@gmail.com; markovits.henry@uqam.ca

http://www.psycho.uqam.ca/NUN/d_pages_profs/d_Markovits/Index.html

Abstract: Byrne's book makes a strong case for the important role of imagination as a creator of possibilities that are used to understand complex relations, while remaining rational. I suggest that imagination also serves a critical developmental role by creating possibilities that are not rational, and that act to modify the nature of the cognitive processes that are used to define rationality.

Ruth Byrne has provided a thorough and important analysis of the relationship between rationality and imagination (Byrne 2005). This is a very useful extension of more typical analyses of the nature of rational thought which often concentrate on the ability to give the right answer, and mostly neglect the ability to go beyond the problem parameters that is an important aspect of our intuitive understanding of imagination. As Byrne points out, people spend a great deal of cognitive energy in reflecting on what could have been in different situations and on the resulting possibilities. This, in itself, is not particularly new; the novel aspect of her analysis is the use of a framework that explicitly situates imagination within the more constrained focus of rational thought. In other words, the possibilities that people imagine are not simply generated at random, but are constrained by the nature of the cognitive processes that define rationality.

Nonetheless, I would claim that the relationship between imagination and rationality that is sketched out in this book is at least partially flawed, because it does not allow for any real development of either imagination or rationality. To make this point clear, I must start by simplifying what is a complex analysis. The essence of Byrne's argument is of course derived from mental model theory. This claims that people have standard semantic representations of reality that allow them to represent not only what is, but a subset of what is possible. These representations are essentially rational, because they are defined in such a way that, when they are applied consistently and used with the standard algorithms defined by the mental model theory, they will give responses to inferences that are either textbook ones or, in some more general sense, optimal. Possibilities that are generated must remain consistent with these semantic representations. This in turn allows for a fairly tight definition of imagination that constrains the possibilities implied by imagination to those that are consistent with the rational semantic representations that underlie the interpretation of relational terms, such as if–then.

This general model proposes an essentially static relationship which accounts for the influence of rationality upon imagination, but does not allow the inverse effect. The key problem is that possibilities are processed within the constraints imposed by whatever processes define rational thinking. This is a basic limitation, particularly when considered in a developmental context. One point of view, which implicitly underlies Byrne's analysis, is that the processes that define rational thought are essentially innate, and that there is no need to suppose any underlying developmental change. If this is not the case, then there has to be a mechanism by which the nature of these processes change. And, one of the key intuitive candidates must be imagination. In other words, if whatever constitutes rational thinking in adults is derived from a developmental process, then imagination might be considered to be one factor in creating more complex forms of thinking (and possibly more rational ones).

This is, in fact, the hypothesis that was put forward by Piaget (1981) in a series of imaginative experiments examining the relationship between the range of possibilities that children of different ages were able to generate in a given context and the complexity of their cognitive processes. Older and cognitively more advanced children were able to generate a larger and qualitatively more diverse set of possibilities than were younger ones. My students and I have found similar results when looking at the relationship between different levels of conditional reasoning and the kinds of possibilities that are generated by children and adults (Janveau-Brennan & Markovits 1999; Markovits & Vachon 1990; Venet & Markovits 2001). Most interestingly, these relate not only to the quantity of such possibilities, but also to their nature. In other words, if one looks at imagination developmentally, there is clear evidence of a qualitative shift that goes from more constrained, experientially based possibilities to more general and abstract forms of possibilities. Piaget (1981) argued that the relation between the cognitive processes that determine reasoning and children's imagination was bidirectional. Although many possibilities are indeed implied, and limited, by the

cognitive processes used by a given subject in a way that is consistent with Byrne's model, Piaget also claimed that some were generated because children observed or deduced possibilities that were in fact inconsistent with these cognitive processes. These possibilities become sources of disequilibrium that can only be made "rational" by adjusting the characteristics of the cognitive processes. In other words, this idea assumes that imagination can sometimes go beyond available cognitive processes and result in a reconfiguration of what is considered to be rational.

Direct evidence for any such process remains anecdotal, although anyone who has listened to a child work out a complicated problem will find it quite convincing. There are, however, examples of this kind of process in the development of science that make very useful analogies. For example, Newtonian mechanics postulated that velocities are linearly additive, which is of course a very intuitively rational concept. The Michelson-Morley experiment provided empirical data that was simply inconsistent with Newtonian theory (Michelson & Morley 1887). In other words, the results of this experiment were not possible within what was considered to be rational at that time. It was not until Einstein's special theory of relativity that an explanation of this result was (eventually) accepted: In this theory, velocities are not linearly additive, which has replaced Newtonian rationality. Einstein derived this theory, not by empirical work, but by a series of thought experiments, that is, by imagination, that allowed him to go beyond Newtonian rationality.

Thus, what I suggest here is that Byrne's analysis fairly represents a major part of the work done by the imagination, which is to examine possibilities in a way that is consistent with what a person's "rational" processes allow as being possibly true. However, it neglects the potentially critical role of the imagination in constructing possibilities that are not rational, but that suggest the necessity of revising our definition of what is rational.

Thinking developmentally about counterfactual possibilities

doi: 10.1017/S0140525X07002695

Kevin J. Riggs[a] and Sarah R. Beck[b]

[a]*Psychology Department, London Metropolitan University, London E1 7NT, United Kingdom;* [b]*School of Psychology, University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom.*

k.riggs@londonmet.ac.uk **s.r.beck@bham.ac.uk**

Abstract: Byrne implies that working memory development underpins children's ability to represent counterfactuals as possibilities at 3 to 4 years of age. Recent findings suggest that (1) developments in the ability to consider alternatives to reality in children of this age are underpinned by improvements in inhibitory control, not working memory, and (2) children do not develop an understanding of counterfactuals as possibilities until mid-childhood.

Based on an impressive amount of data from a sustained research program over many years, Byrne (2005) presents an exciting theoretical framework for how we think about alternatives to reality. We focus on two points: (1) counterfactuals as dual possibilities, and (2) the cognitive processes involved in counterfactual thinking; and we comment on them from a developmental perspective.

Counterfactuals as dual possibilities. Based on her work with adults, Byrne claims that counterfactuals are represented as dual possibilities. Given that children first start to pass explicit tests of counterfactual thinking at around age 3 or 4 years, it seems reasonable to infer that children also represent counterfactuals as possibilities at this age. However, in one recent study 4-year-olds who could correctly answer a question that referred to a counterfactual event of the type, "What if X had happened, how would the world be?" were unable to acknowledge that at a previous point in time either the counterfactual or actual event *could* have occurred (Beck et al. 2006). Our interpretation of this finding is that when children first start to think counterfactually they think only about what did not happen, but do not relate it to, or represent, the actual world explicitly. Thus, early counterfactual thinking might not involve thinking about possibilities, even though it does involve thinking about what might have been.

Cognitive processes. Throughout the book Byrne suggests that differences in working memory may be responsible for both individual differences and developmental changes in counterfactual thinking. Byrne argues, quite reasonably, that pre-school children find counterfactual conditionals, "*What if X had not happened, how would the world be?*" more difficult than simple causal conditionals, "*What if X happens, how will the world be?*" (see Riggs et al. 1998) because counterfactuals make the greater working memory demands (we also know that working memory develops substantially in the pre-school years). While the case for adult variation in counterfactual thinking (and working memory) is supported by empirical evidence, there is little or no evidence to suggest that working memory underpins early developments in counterfactual thinking.

Recently, we tested Byrne's idea (Beck et al. submitted). We asked 3- and 4-year old children counterfactual conditional questions of the sort used by Riggs and colleagues (1998) and also gave them a battery of executive function tasks. We found that once receptive vocabulary and age were taken into account, working memory did not predict counterfactual thinking ability. Rather, we found that inhibitory control predicted success on counterfactual conditional tasks, independently of age, language, and working memory (though, interestingly, inhibitory control did not predict success on counterfactual syllogistic reasoning tasks).

Current evidence suggests that developments in counterfactual thinking continue after the pre-school years and comes from the literature on regret (Guttentag & Ferrell 2004). Children's evaluations of who will feel regretful are not influenced by counterfactual possibilities until they are at least 7 years old. Given that children do not make the comparison between how things are and how things could have been suggests to us that they are not holding both possible worlds in mind. We agree with Byrne that working memory should be implicated in counterfactual thinking when one holds in mind both the counterfactual and the actual possibility; and for this reason we predict that developments in working memory underpin the ability to understand counterfactual emotions.

In short, we are in agreement with Byrne that mature counterfactual thinking requires representing dual possibilities. But we do not think this ability develops at around 3 or 4 years of age. At this age there are developments in the ability to consider alternatives to reality, but this ability appears to be related to improvements in inhibitory control, not working memory. Recent evidence suggests that representing counterfactuals as possibilities (what we might think of as genuine or adult-like counterfactual thinking) develops later, in middle childhood, which may well be driven by developments in working memory (though inhibitory control may also play a role).

Byrne's framework will prove to be immensely helpful to developmental psychologists who question when and how children engage in imaginative reasoning, including object substitution pretence, counterfactual conditional reasoning, and syllogistic reasoning with false premises. Furthermore, the book raises a number of other topics for future developmental research programs. However, we also believe that a developmental perspective on many of these issues will provide a richer and ultimately more comprehensive account of what it means to be able to consider alternatives to reality.

When imagination is difficult: Metacognitive experiences at the fault lines of reality

doi: 10.1017/S0140525X07002701

Lawrence J. Sanna

Department of Psychology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3270.

sanna@unc.edu **http://www.unc.edu/~sanna**

Abstract: Imagination and rational thought may be guided by identical principles, and Byrne's (2005) analysis expertly synthesizes a diverse literature on counterfactual thinking. Further attention should be paid to metacognitive experiences, like ease or difficulty of thought generation, which accompany the imaginative process. Only by considering metacognitive experiences along with the content of what people imagine can we fully understand imagination.

People's capacity to imagine what might have been, as described by Byrne (2005), is governed by rational principles that are identical to other forms of thinking. On the one hand, this is a particularly intriguing idea because prior theory has been at odds in suggesting that imagination is somehow irrational, or that imagination and rational thought are directed by incompatible rules. Byrne demonstrates this is not so by delineating common principles whereby "fault lines" in reality – where counterfactual thinking is most probable – produce thoughts of an "if only" nature. Her book provides a lucid integration of diverse literatures, the major points of which I am in agreement with. Byrne is to be commended for a highly thoughtful and readable book, which serves as a welcome breakthrough in conceptualizing imaginative reasoning.

On the other hand, my main point is to focus greater attention on the mostly neglected but critically important role of people's metacognitive experiences in influencing imaginative (and rational) thinking. This includes various subjective experiences that accompany the imaginative process, such as ease or difficulty of thought generation or recall, processing fluency, or emotions like surprise (Sanna & Schwarz 2006; 2007; Schwarz et al. 2007). To fully understand imagination, metacognitive experiences must also be accounted for, because they are informative in their own right and can qualify or even reverse the implications that people draw from what they are imagining.

A hindsight bias example illustrates this (Sanna et al. 2002a). *Hindsight bias*, people's belief that they knew it all along (Fischhoff 1975), results from thinking about known outcomes but it may be eliminated – or lessened – by thinking about counterfactuals (Guilbault et al. 2004; Hawkins & Hastie 1990). After reading about a British–Gurkha war that the British won, some people were asked to imagine 2 or 10 reasons supporting this outcome, whereas others were asked to imagine 2 or 10 reasons supporting the counterfactual outcome (Sanna et al. 2002a, Experiment 1). If only content mattered, hindsight bias should be greater when imagining 10 than 2 reasons supporting the known outcome (British victory); the bias should be lessened when imagining 10 than 2 reasons supporting the counterfactual (Gurkha victory). But exactly the opposite happened (Fig. 1): Imagining more reasons favoring the known outcome decreased hindsight bias, whereas imagining more reasons favoring the counterfactual outcome increased hindsight bias (see also Sanna et al. 2002b) (Fig. 1).

Thus, it is not just *what* people imagine that counts. A key to understanding these results is that people's self-reports indicated that imagining 2 reasons was experienced as easy and 10 reasons was difficult, irrespective of whether they focused on known or counterfactual outcomes. Known outcomes were seen as unlikely when it was difficult to think of reasons for a British victory – after all, if there were many reasons for a British victory, it should not be so hard to think of 10. Conversely, people inferred that counterfactual outcomes were unlikely when it was difficult to think of reasons for a Gurkha victory. In each case, people's

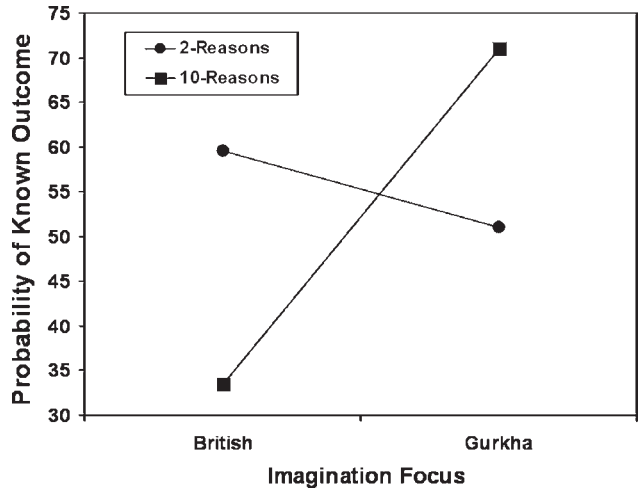

Figure 1 (Sanna). Mean probability of known outcome (British victory) on a 0–100% scale. British focus represents imagining the known outcome; Gurkha focus represents imagining the counterfactual outcome. All people were asked, "If we hadn't already told you who had won, what would you have thought the probability of the British winning would be?" Adapted from Sanna et al. (2002a, Experiment 1).

inferences were consistent with imagination content only when reason generation was easy, whereas inferences were *opposite* to the implications of imagination content when reason generation was difficult.

Other examples of metacognitive experiences include *processing fluency* – that is, ease or difficulty with which new information is processed – and emotions like surprise. People presented with general knowledge questions and answers (e.g., "How high is the Eiffel Tower?" – "300 meters") more likely imagined that they knew the outcome (answer) all along when questions and answers were presented in easy-to-read rather than difficult-to-read colors (Werth & Strack 2003); and people who first identified faces more likely imagined that naive others would identify the faces sooner than they themselves just did (Harley et al. 2004). Emotions such as high surprise can inform people that outcomes were unexpected, and low surprise, that outcomes were expected (Ortony et al. 1988); and moods can inform people that things are fine or problematic (Sanna et al. 1999). Each of these experiences has implications for imagining what happened and what might have been in ways not predicted by the content of imagination alone (Sanna & Schwarz 2006; 2007; Schwarz et al. 2007).

Consistent with the simulation heuristic (Kahneman & Tversky 1982) and norm theory (Kahneman & Miller 1986), Byrne's principles and corollaries (summarized on pp. 200 and 203) recognize that features like actions, controllable events, and so on, more easily bring to mind counterfactuals, and that, when two (or more) possibilities are available from the outset, counterfactuals are more likely. Byrne's synthesis greatly enriches the field by providing a framework in which to understand disparate findings that span various literatures. But metacognitive experiences encompass much more than this (Sanna & Schwarz 2007). In this sense, Byrne's analysis did not go far enough. Metacognitive experiences are part and parcel of the imaginative process. In fact, the principles outlined in Byrne's book (e.g., actions are more mutable) may exert their influences precisely because of the information people derive from metacognitive experiences. And the accompanying metacognitive experiences can actually *change* the meaning and inferences drawn from thinking counterfactually.

Because people may truncate thought generation early (Bodenhausen & Wyer 1987), in many real-life circumstances

counterfactuals could be imagined before any experienced difficulty, under ease or fluency.[1] But it would be erroneous to conclude that reactions can thus be predicted on the basis of content alone. For example, one potential irony is that difficulty or disfluency might occur precisely when thinking about alternatives is most needed, as when people are particularly motivated to understand what went wrong and wind up searching for many counterfactuals. This may leave people *less* able to learn from past mistakes, and unlikely to take steps to improve. Thus, only by considering metacognitive experiences along with the content of what people imagine can we fully understand imagination. In short, following through with Byrne's analogy, when fault lines in reality fissure, metacognitive processes may provide the seismic waves that ripple through the imagination to give meaning to the whole experience.

ACKNOWLEDGMENT

I thank the Imagination, Goals, and Affect (IGoA, or ego) laboratory group members at the University of North Carolina at Chapel Hill for their comments.

NOTE

1. Because people normally generate only a few counterfactuals when asked in experiments, there can be a similar natural confound between counterfactuals and ease of generation.

Imagination is only as rational as the purpose to which it is put

doi: 10.1017/S0140525X07002713

Andrew Shtulman

Department of Psychology, Occidental College, Los Angeles, CA 90041.
shtulman@oxy.edu **http://faculty.oxy.edu/shtulman/**

Abstract: Byrne's criteria for considering imagination rational do not accord with standard notions of rationality. A different criterion – that is, the correspondence between an inference strategy and its domain of application – is offered and illustrated with recent work on possibility judgment. This analysis suggests that, although imagination can be put to rational purposes, imagination itself should not be considered rational.

Byrne's (2005) book, despite its clarity of analysis and clarity of exposition, advances a claim that is far from clear: Imagination is rational. This claim is unclear because the word *rational* typically denotes conformity to some normative standard, yet, as Byrne herself notes, "There is no normative standard against which to judge whether an imaginative thought is best" (p. 209). Even if one adopts Byrne's view of rationality as the *capacity* to draw normatively valid conclusions (as opposed to the disposition to do so), one is still left with the problem of deciding what constitutes such a capacity in the domain of counterfactuals.

Byrne acknowledges this problem, but rather than confront it directly by explaining why the *content* of counterfactual inferences should be considered rational, she circumvents it by explaining why the *process* of counterfactual reasoning should be considered rational. Her argument proceeds as follows: (1) the process by which most individuals make deductive inferences (i.e., searching for counterexamples to an argument's conclusion among a set of possibilities consistent with the argument's premises) is capable of yielding normatively valid conclusions; (2) counterfactual reasoning shares many similarities with deductive reasoning; (3) therefore, if deductive reasoning is considered rational, then counterfactual reasoning should be considered rational as well.

The problem with this argument is that, without any independent measure of what constitutes a valid counterfactual inference, we have no reason to believe that the analogy between deductive reasoning and counterfactual reasoning is, itself, valid. After all, the literature on strategy development (e.g., Siegler 1996) has documented many instances in which failures of reasoning are attributable to the misapplication of domain-specific strategies. For instance, when children are first introduced to decimal notation, they often compare decimals on the basis of digit length rather than digit location, judging a decimal like .125 to be larger than a decimal like .25 because the former contains more digits than the latter (Moss & Case 1999; Smith et al. 2005). Although this strategy is reliably correct when applied to integers, it is not reliably correct when applied to decimals.

Is the application of deductive-reasoning strategies to counterfactual-reasoning problems as inappropriate as the application of integer-comparison strategies to decimal-comparison problems? Perhaps not, but Byrne provides no reason for us to believe otherwise. By focusing on processing similarities between deductive reasoning and counterfactual reasoning, Byrne overlooks potential dissimilarities in their application. One such dissimilarity is the nature of the space of possibilities over which each type of inference is drawn. That is, when reasoning about factual conditionals of the form "if A, then B," individuals are limited to a small, well-defined space of possibilities (i.e., A and B, A and not-B, not-A and B, not-A and not-B), but when reasoning about counterfactual conditionals, they are confronted with the space of all possible worlds (Lewis 1973; Stalnaker 2003). Thus, the absence of a counterexample specifies a normatively valid conclusion in the former space of possibilities but not the latter. Indeed, to conclude that reality is immutable because no changes to reality are conceivable is, in Dennett's words (1993), to "mistake a failure of imagination for an insight into necessity" (p. 48).

Consistent with this idea, most adults recognize, at least implicitly, that failures of imagination do not count as evidence of necessity (Shtulman & Carey 2007). That is, when asked to judge the possibility of events that violate physical laws, like walking through a wall or walking on water, most adults not only deny the possibility of such events, but also justify their judgments with *positive evidence* of the events' impossibility (e.g., "both walls and people are solid," "water doesn't have enough surface tension"). In other words, rather than appeal to the perceived absence of a counterexample (e.g., "there's no way a person could walk on water"), adults tend to provide principled reasons for why no such counterexamples exist.

In contrast to adults, preschool-aged children do not tend to provide principled reasons for their judgments. Instead, they appeal to their own failures of imagination, either explicitly (e.g., "it just doesn't seem possible") or implicitly, via the comparison of a seemingly impossible event to a possible one (e.g., "you can't walk across water but you could swim across"). Such appeals suggest that preschoolers reason about physical possibility similarly to how adults are purported to reason about counterfactuals: by searching for counterexamples to the status quo. If they can identify such a counterexample, they judge the event possible; if they cannot, they judge the event impossible. Although this strategy does, in fact, lead children to deny the possibility of events that violate physical laws, it also leads them to deny the possibility of events that, although difficult to imagine occurring, do not violate any physical laws, like making pickle-flavored ice cream or finding an alligator under the bed.

The point of this illustration is not to suggest that the process of searching for a counterexample is irrational but to suggest that this process is rational in some contexts (i.e., small, well-defined domains) and not in others (i.e., large, ill-defined domains), and that the appreciation of this fact is a normal developmental achievement. Moreover, by considering whether the *application* of an inference strategy is rational – as opposed to the strategy itself – one can better appreciate what constitutes a valid conclusion in the domain at hand and what does not. Admittedly, the aforementioned findings come from studies of

hypothetical reasoning, not counterfactual reasoning; yet they pertain to Byrne's claims in so far as reasoning about the mutability of particular events in the past is structurally similar to reasoning about the mutability of generic events, past or present. At the very least, this comparison points to the need for additional research on how individuals *justify* their counterfactual inferences, for such data are likely to shed light on how those inferences were made.

In sum, imagination can be put to rational purposes but it should not be considered inherently rational. Although Byrne's careful analysis of the similarity between counterfactual reasoning and deductive reasoning provides evidence of imagination's systematicity, it does not provide evidence of its rationality.

On the relation between counterfactual and causal reasoning

doi: 10.1017/S0140525X07002725

Barbara A. Spellman and Dieynaba G. Ndiaye

Department of Psychology, University of Virginia, Charlottesville, VA 22904-4400.

Spellman@virginia.edu Dgn2 h@virginia.edu
http://people.virginia.edu/~bas6g/

Abstract: We critique the distinction Byrne makes between strong causes and enabling conditions, and its implications, on both theoretical and empirical grounds. First, we believe that the difference is psychological, not logical. Second, we disagree that there is a strict "dichotomy between the focus of counterfactual and causal thoughts." Third, we disagree that it is easier for people to generate causes than counterfactuals.

Psychologists studying the relation between counterfactual and causal reasoning have long asked: Why, despite their similarity, do people give different answers to counterfactual versus causal questions? (See Spellman & Mandel [1999] for history.) For example, when completing "if only . . ." statements about Mr. Jones who was hit by a drunk driver while taking an unusual route home, most people focus on the unusual route, yet they identify the drunk driver as the cause of the accident (Mandel & Lehman 1996).

In the chapter "Causal Relations and Counterfactuals," Byrne (2005) argues that people provide different answers because they focus on different things: in counterfactual reasoning they focus on "enabling" conditions, whereas in causal reasoning they focus on "strong causes." Imagine a dry forest floor and then a lightning strike resulting in a huge forest fire. People are likely to say, "if only there were not so many dry leaves," and "the lightning caused the fire," but not "the dry leaves caused the fire." Byrne argues that strong causes (lightning) are consistent with two possibilities: (1) lightning and fire, and (2) no lightning and no fire – however, people mentally represent only the first possibility. Enabling conditions (dry leaves) are consistent with three possibilities: (1) dry leaves and fire, (2) no dry leaves and no fire, and (3) dry leaves and no fire – however, people mentally represent two possibilities (or only the first, but the second comes "readily"). People, Byrne argues, use those representations to distinguish causes from enablers and, as a result, answer counterfactual questions with enablers and causal questions with strong causes.

We have trouble with some of the assumptions and assumed consequences of that characterization on both theoretical and empirical grounds. First, we believe that the difference between enablers and causes is psychological, not logical. Second, we do not believe that there is a strict "dichotomy between the focus of counterfactual and causal thoughts" (Byrne 2005, p. 100). Third, Byrne argues that as a result of the difference in representation, it is easier for people to generate causes than counterfactuals; we disagree.

Enablers versus causes. At first the dried-leaves-and-lightning example seems obvious: of course dried leaves constitute an enabler, whereas lightning is a cause. But on deeper reflection the logic is not so clear. Dried leaves would not lead to a conflagration without lightning; however, neither would lightning without dried leaves. Their logical status is equivalent: each is necessary but neither is sufficient.

Similarly, consider a lightning-torn stretch of wetlands. Despite countless lightning strikes, there was never a fire until the year's masses of dry leaves blew in. Now it seems natural to argue that leaves caused the fire, whereas lightning was an enabler. Again, calling one a cause and one an enabler is a psychological, not a logical, judgment, and to explain differences in counterfactual and causal judgments by saying that people represent causes and enablers differently is to finesse the importance of various factors (e.g., context) that get people to treat logically equivalent events as psychologically different. (See Einhorn & Hogarth 1986 and McGill 1989, for other context effects.) It is unclear how the mental representation of possibilities accounts for such context effects and informs people about which is the cause and which is the enabler; it seems that people must already know which is which based on the context before they represent the events. Byrne does mention alternative information sources (covariation, mechanisms, abnormality), but her argument implies that the mental representation of possibilities provides a better account of how people distinguish strong causes from enablers.

Not quite a "dichotomy". Second, it is inaccurate to characterize people's answers to causal and counterfactual questions as a strict "dichotomy." In some studies, the most prevalent answers are the same (e.g., Wells & Gavanski 1989, Experiment 1). Plus, differences in how counterfactual and causal reasoning are measured may contribute to belief in the dichotomy. Our participants read about a woman driving home from work. She stops at a red light and fiddles with the radio so that when the light turns green she hesitates before accelerating, delaying the cars behind her. Last in line is a school bus, which enters the intersection just as an irate man drives through the red light from the other direction hitting the bus and injuring many children.

Participants who listed counterfactuals focused on the hesitating woman; participants who rated causes focused on the irate man. These results replicate the "dichotomy." However, there is a confound: researchers usually measure counterfactuals with listings but causes with ratings. What if both are measured with ratings? Other participants saw 12 story events previously listed by earlier participants and rated each on either whether they agreed the event was an "undoing counterfactual" or whether it was causal. The irate man was rated as both most causal and most changeable (Spellman & Ndiaye 2007).

Thus, counterfactual and causal judgments are far from dichotomous; rather, depending on how questions are asked and answers are measured, they may focus on the same events.

Generating causes and counterfactuals. Byrne argues that because strong causes are represented by one possibility and enablers by two, and because "it is easier to think about one possibility than about several" (Byrne 2005, p. 119), it should be easier for people to generate causes than counterfactuals. McEleney and Byrne (2000) had participants imagine they had moved to a new town to start a new job and read about various events that happened to them. When asked what they would have written in their diaries, participants spontaneously generated *more* causal than counterfactual thoughts. In contrast, our participants read about a man who had been abused by his father, joined the army, learned to use explosives, then blew up his fathers' company's warehouse. Participants listed *fewer* causes ($M = 5.7$) than counterfactuals ($M = 7.7$) (Spellman & Ndiaye 2007). We have no problem

distinguishing the studies – Byrne's answers were spontaneous, whereas ours were evoked; Byrne's story was about the participants themselves, whereas ours was about someone else – yet Byrne's models approach cannot account for the difference in results.

In summary, we believe that the present explanation of the differences between causal and counterfactual judgments suffers on both theoretical and empirical grounds. We prefer to think that both the similarities and differences between those judgments can be explained by the idea that counterfactuals provide input into causal judgments (Spellman et al. 2005). But that argument is best left for another day.

How rational is the imagination?

doi: 10.1017/S0140525X07002737

Robert J. Sternberg

Office of the Dean of the School of Arts and Sciences, Tufts University, Medford, MA 02114.

Robert.sternberg@tufts.ed

Abstract: Byrne has written a terrific book that is, nevertheless, based on a mistaken assumption – that imagination is largely rational. I argue in this commentary that her book follows very well, if one accepts her assumption of rationality, but that the bulk of the evidence available to us contradicts this assumption.

In psychological science, the assumptions underlying one's work are often more important than the work itself. For example, the voluminous experiments conducted by B. F. Skinner and many of his colleagues (see Skinner 1965) in the name of behaviorism make perfect sense, if one accepts that the mind is an impenetrable black box. If one does not accept this assumption, one may wonder why one would have conducted any of these experiments in the first place, or reach the conclusion that Skinner did, even to the end of his life, that we understand the mind when we understand environmental contingencies. Cognitive psychology of the kind proposed by Miller et al. (1960) and Newell and Simon (1972), which has largely (although not entirely) replaced behaviorism, has argued that we can indeed penetrate the black box, and that, when we do, we find that an astonishing variety of behavior can be understood in terms of the rational thoughts underlying it. Economists have long accepted this view, although Kahneman and Tversky (1972) challenged it, only themselves to be challenged by Gigerenzer (2007) and others. Today, affective scientists, among others, are showing the extent to which cognition taken alone does not account for behavior we once thought was purely or even largely cognitive (see, e.g., Davidson et al. 2002).

The July 14, 2007, *New York Times* featured a story discussing mass murderers in the Muslim world and exploring how such acts could be committed by "people who have supposedly dedicated their lives to scientific rationalism and to helping others" (Fattah 2007). The story quotes a Jordanian researcher, Hassan Abu Hanieh, as stating that the most radical among the Muslims are those with the most scientific tendencies. Osama Bin Laden and Ayman Al-Zawahri (leaders of the terrorist group Al Qaeda), and George Habash (terrorist and former leader of the Popular Front for the Liberation of Palestine), are given as several of a number of examples. The focus of the story is on the doctors who unsuccessfully plotted the recent London bombings and then the smashing of their car into the Glasgow Airport. One could argue, of course, that the suicide bombings and mass murder, and the imaginations used to spawn them, are rational, at least for some people, but if one argues that, what meaning is left in the term "rational"?

The story in the July 13, 2007, *New York Times* does not necessarily bode much better for the rational imagination hypothesis. Louisiana Senator David Vitter spoke quite rationally, he thought, in arguing for family values and a whole host of right-wing causes. It is not clear that his imagination was so rational in imagining and acting on his desires to patronize an escort service, leading him to confess to "serious sin" (Associated Press 2007). He and other "reasonable" thinkers – Bill Clinton and Newt Gingrich among them – seem to have been anything but rational in their imaginings about how to satisfy their needs. Their thinking seems to have approached the kind of "miracle world" (Byrne 2005, p. 10) counterfactual thinking that Byrne seems to think is so rare – a miracle world in which they, as extremely public figures, can say one thing, do the opposite, and hope that no one notices the contradictions. At this point, Bush's cheery, optimistic thinking regarding the Iraq War perhaps comes even closer to miracle world thinking, as does that of the roughly 30% of the people in the United States who are satisfied with his job performance. The Bush camp once imagined that the battle would be a cakewalk and that the mission was already accomplished, but there was little rationality in their imaginings.

Note that the argument here is not about error (Byrne 2005, p. 17). When it comes to imagination, it is not even clear what constitutes error. Rather, it is about the kind of largely irrational imagining people do much of the time. Some of that thinking leads nowhere (e.g., "It would be nice if I were President of the United States"). Some of it leads somewhere (e.g., Munch imagining *The Scream*, and it is not clear what is "rational" about that painting). And some of it leads to disaster (e.g., Clinton imagining his encounter with Monica Lewinsky the day she wore the dress that later would provide DNA evidence against him).

Most of Byrne's examples throughout her book are hypothetical and, like laboratory studies, more susceptible to rational thought. When she gets into real-world examples, I believe, her notion of rational imagination breaks down. One such example is Lee Harvey Oswald's murder of John F. Kennedy. What was rational about Oswald's imagining the results? Perhaps he was deluded. But today, countless people still believe, against all evidence, that he was part of a larger plot. They might be right. But the evidence is not there. Millions of people believe that the 9/11 attacks, another real-world example used by Byrne, were a Jewish plot. Again, one cannot prove that this attack was not a Jewish (or Buddhist or Hindu) plot, but the evidence is not there. Are their imaginations rational, or in the service of what their emotions wish them to believe? In the news on July 13, 2007, Bush was reported as still drawing a link between Al Qaeda and 9/11, years after any such link was shown to be false (Gordon & Rutenberg 2007). There is not much rationality in that imagination. Byrne's examples work better when they are hypothetical laboratory fictions than in the few cases they are drawn from the real world. Byrne's section on "Why people focus on forbidden fruit" (Byrne 2005, p. 87) just doesn't explain why people focus on forbidden fruit.

One even might argue that the emphasis on rationality and imagination is not only incorrect, but can be harmful in the practical domain. This is largely the argument of Westen (2007), whose main thesis is that the Democrats repeatedly lose U.S. and other presidential elections because they severely misunderstand the minds of the electorate. They focus on the rational aspect of the imagination, whereas in fact people's votes are controlled by their emotions, an understanding the Republicans reached long ago. Stanovich (2002) has even coined the term "dysrationalia" to characterize the very irrational thinking and imagining of intelligent people.

To conclude, Byrne's *The Rational Imagination* is a brilliant book that deserves great commendation. But its fundamental assumption regarding the rationality of imagination reflects a wishful view of cognitive psychology that people's behavior, for the most part, can be understood in rational terms. As we look at the current messes in our own country (the United States), as well as in others, we may think – if only it were so.

The development of the counterfactual imagination

doi: 10.1017/S0140525X07002749

Jennifer Van Reet, Ashley M. Pinkham, and Angeline S. Lillard

Department of Psychology, University of Virginia, Charlottesville, VA 22904.
vanreet@virginia.edu **apinkham@virginia.edu**
lillard@virginia.edu **http://people.virginia.edu/~jlv2t/**
http://people.virginia.edu/~asl2h/

Abstract: How the rational imagination develops remains an open question. The ability to imagine emerges early in childhood, well before the ability to reason counterfactually, and this suggests that imaginative thought may facilitate later counterfactual ability. In addition, developmental data indicate that inhibitory control may also play a role in the ability to reason counterfactually.

Byrne (2005) argues that imagination is crucial to adult cognition. To support this claim, she demonstrates that the ability to envision alternatives to reality is regularly employed by adults when they reason about how the world may be or may have been. Because imagination is especially salient in childhood, one might also ask whether the imagination plays a similar role in children's reasoning. Does imagination allow children to reason counterfactually? The standard view is that, in infancy, we are tied to reality, unable to consider alternatives (Perner 1991). Considering this, how do we develop into rational imaginers? Although Byrne argues that rational principles underlie imaginative thought, we suggest that the opposite is true: that imagination underlies rationality. Here, we consider how developmental research on pretense, imagination, and inhibitory control demonstrates that early imaginative thought may provide the foundation for counterfactual reasoning.

By regarding pretense as a type of counterfactual thought (p. 2), Byrne seems to suggest that imaginative and rational thought are governed by the same processes. It is certainly true that pretense and counterfactual reasoning share a common skill: namely, the ability to mentally create an alternative to reality. In pretense, when a child pretends that a block is a cookie, he or she must mentally represent an alternative world in which he or she is acting with a cookie, not a block. However, reality cannot be completely ignored. A real representation of the world must be maintained simultaneously with the pretend representation because, if reality did not limit pretense behavior, the child may act irrationally, such as attempting to actually eat the cookie (Leslie 1987). Thus, to pretend successfully, an individual must concurrently create an alternative reality while maintaining an accurate representation of reality.

Counterfactual reasoning also requires the ability to mentally represent alternative realities, because the initial step in addressing a counterfactual situation is mentally changing an aspect of reality. To reason about the question "Would the paper have blown away if I had closed the window?" one's mental representation of reality (i.e., an open window) must first be changed to the alternative state (i.e., a closed window). As Byrne discusses, if one were unable to complete this first step and imagine the alternative reality, she or he would also be unable to rationally answer the question about how the present might have been different.

Although the initial step of creating an alternative representation is the same in both pretense and counterfactual reasoning, these two processes are actually quite different. Consider the steps following the creation of the alternative in each case. In pretense, an individual enters the alternative world and acts online from within the pretend framework. Little thought about the real world is required because the pretend world is not entirely constrained by reality. Although a representation of reality must be maintained and kept separate from the alternative to avoid confusion, anything can happen in the pretend world. Pretending a block is a cookie is just as permissible as pretending to have tea with the Queen of England. Furthermore, there is no overt goal to pretend play. It exists solely in the present and does not require any explicit comparison between the real and alternative state. The two states coexist without conflict; it is not problematic to mentally represent the block as both a block and a cookie.

However, counterfactual reasoning does have an explicit goal, namely, to determine how the present reality might be different given an alternative premise. This situation necessitates a much closer relationship between real and alternative representations than in pretense. According to Byrne, following the creation of an alternative state in counterfactual reasoning, the individual manipulates that alternative based upon what actually exists or actually has happened in the real world. When the alternative temporally catches up to the present reality, one must explicitly compare the two to determine whether the change made in the alternative would have affected the current reality. This clearly differs from pretense because counterfactual reasoning requires not only representing the real and the alternative, but also comparing the two.

How do children develop this latter, more complex ability? We hypothesize that the development of inhibitory control facilitates counterfactual reasoning because inhibitory control is necessary to manage two conflicting representations successfully. Consider a simple counterfactual problem, such as reasoning from a false premise. In the example, "All fish live in trees. Where does Bob the fish live?" the alternative (i.e., all fish live in trees) directly conflicts with reality (i.e., all fish live in water). To answer this question correctly, one would have to inhibit the real world in order to reason based on the alternative. Preschool-age children reliably make "realist" errors when asked this question, answering in a manner that conforms to what they know about the real world. Given that children of this age have no trouble creating alternative states in pretense or making deductions that do not conflict with reality (Hawkins et al. 1984), we argue that children make these realist errors because they do not yet have the inhibitory control required to disregard the real state of the world. In other words, they cannot inhibit their knowledge that fish live in water in order to answer based upon the alternative premise (Lillard 2007; Ma 2007).

If the inhibitory demands are removed from counterfactual reasoning tasks, children's performance should improve. This is exactly what happens. When children are explicitly instructed to use their imaginations in counterfactual deductive reasoning tasks such as the aforementioned Bob-the-fish example, they are more successful than children who are not given an imagination warm-up (Dias & Harris 1988; 1990; Richards & Sanderson 1999). By instructing children to imagine, the inhibitory load is reduced by turning a counterfactual problem into a pretense situation. When children use their imaginations, they set up an alternative that does not require comparison to reality. Thus, the inhibitory demand is markedly reduced because children do not have to shift repeatedly between the alternative and real representations.

Overall, developmental research concurs with Byrne's thesis that imaginative thought is required for counterfactual thought. However, we argue that such research also clarifies the relationship between the two by showing how children's early imaginative thought, in conjunction with inhibitory control, may facilitate the ability to reason counterfactually.

Is considering true possibilities a truly explanatory principle for imaginative thought?

doi: 10.1017/S0140525X07002750

Thomas B. Ward

Department of Psychology, University of Alabama, Tuscaloosa, AL 35487.
tward@bama.ua.edu **http://bama.ua.edu/~tward**

Abstract: Byrne (2005) demonstrates that reasoning and imagination are logical and governed by the same processing principles. In extending those principles to other forms of imaginative functioning, however, problems arise. The meaning of "true possibility" is stretched, and the causal role of the principles is not well established. Nevertheless, consideration of the extent to which ordinary cognitive processes govern creative functioning is valuable.

The greatest strength of Byrne's (2005) book is that it explores a particular aspect of a broader question: namely, the extent to which creative forms of cognition spring from the same fundamental processes as "ordinary," non-creative cognition. The issue is critical because so much of human cognition is creative and yet, historically, comparatively little theoretical and empirical effort has been expended to examine creative functioning (Finke et al. 1992; Sternberg & Lubart 1996; 1999; Ward et al. 1999).

Byrne has marshaled considerable evidence for the ideas that human reasoning is rational, that imagination is rational, and, most importantly, that both types of cognitive functioning may be governed by the same small set of principles. In extending the analysis to other aspects of creative functioning, including conceptual expansion (e.g., see Ward 1994) and conceptual combination (e.g., see Wisniewski 1997), however, some questions arise. More specifically, Byrne speculates that when people must generate new instances of already existing categories (e.g., novel science fiction creatures), they consider true possibilities and few possibilities. That is, their thinking is governed by the same principles that guide logical and counterfactual reasoning in the types of problems considered elsewhere in the book. The principles seem to be proposed as playing a causal role in conceptual expansion just as they do in counterfactual reasoning. To illustrate with one of Byrne's examples, in generating novel birds, people might retrieve sparrows rather than penguins because the former are "true" instances that possess wings and feathers, whereas the latter are not "true" in that they lack those attributes.

It has been shown that the majority of people assigned the task of generating new instances rely on highly representative category exemplars (e.g., hammers as a basis for devising new tools) (Ward et al. 2002). To the extent that those more representative exemplars are more "true" (i.e., possess more of the category's characteristic attributes), the data could be interpreted as supportive of the "true possibilities" principle. However, there are two problems with this interpretation. First, sharing more characteristic properties is a very different sense of being a true possibility than, for example, the true states of affairs posed in standard logical and counterfactual problems (i.e., the true states of affairs implied by if–then statements). The danger is that by stretching the meaning of "truth" it becomes a slippery construct that is no longer of tight explanatory value.

The second problem is that, even if shared characteristic attributes can be defended as being the same type of trueness as in logical implication, the connection between it and the use of an exemplar in creative generation may be entirely correlational. The proposed tendency to consider true possibilities may not be what causes people to retrieve and rely on representative exemplars. Rather, the most accessible instances in a category may only coincidentally happen to possess more characteristic attributes, but their accessibility may be determined by any number of factors other than possessing those attributes. To extend this point, once variability due to accessibility is taken into account, typicality (presumably linked to shared attributes) is not positively related to the tendency for people to retrieve and rely on specific category exemplars in creative generation (Ward et al. 2002). Hence, there is reason to doubt trueness, in the sense of possessing certain characteristics of the category, as a causal factor in conceptual expansion.

In a related way, Byrne seeks to account for interpretations of conceptual combinations by way of the same types of principles. Consider that a "cactus fish" might be interpreted as a fish with spikes or prickles. According to Byrne's account, this may be due to people only representing the single (presumably true) possibility of a cactus' spines. Again, there is a question of whether "true" possibility in the sense of distinctive properties is really the same as true possibility as considered in most logic problems. And, again there is a question of whether "considering true possibilities" serves as a casual mechanism, or only happens to be correlated with the myriad properties of concepts that affect the interpretation of conceptual combinations (Bock & Clifton 2000; Estes 2003; Gagné 2000; Gagné et al. 2005; Murphy & Wisniewski 2006; Wilkenfeld & Ward 2001; Wisniewski 1997; Wisniewski & Love 1998). That is, it is the structure of the concepts and the effect of that structure on retrieval and combination that drives interpretation, rather than the principles of representing true and few possibilities.

Explaining the mechanisms that govern conceptual combination is critical to an understanding of creative functioning, because combinations are often mentioned as a source of creative discovery and invention (e.g., see Costello & Keane 2000; 2001; Rothenberg 1979; Scott et al. 2005; Thagard 1984). Similarly, a good deal of creative behavior involves conceptual expansion. As a purely descriptive account, it is appropriate to note the types of connections that Byrne has pointed out that may help to characterize these phenomena. However, offering the principles as general explanatory ones for multiple forms of creative cognition is more problematic.

Finally, Byrne is appropriately cautious in distinguishing between counterfactual imagination and more extraordinary forms of imagination and creativity, such as writing novels and creating symphonies. It is essential not to overextend the principles being proposed to domains of creative functioning for which they are plainly not applicable. However, it should also be noted that many extraordinary forms of creativity arise from continued, recursive, extensive application of basic processes. An example described by the noted fantasy author, Stephen Donaldson (1991), is that he got the idea for his award-winning novels about Thomas Covenant: The Unbeliever by combining the concepts of unbelief and leprosy. The combination sparked his thinking, but he then spent months on developing characters, scenes, maps, and so on (Ward 2001). Presumably, the bulk of that continued creative effort which led to his extraordinary novels was underpinned by ordinary cognitive processes of analogy, mental models, imagery, and so forth. Therefore, although restraint is in order, considering just how far a basic process account can be extended in service of explaining extraordinary creativity is definitely a useful exercise.

Emotional consequences of alternatives to reality: Feeling is for doing

doi: 10.1017/S0140525X07002762

Marcel Zeelenberg

Department of Social Psychology, Tilburg University, 5000 LE Tilburg, The Netherlands.

Marcel@uvt.nl

Abstract: When creating alternatives to reality, people often feel emotions in response to these imaginary worlds. I argue that these emotions serve an important purpose. They signal how the world could have been better and prioritize actions to bring this better world about.

Addressing the question of *how* people create alternatives to reality calls also for the question of *why* they should do so. Byrne (2005) provides an excellent overview of the counterfactual thinking literature that is primarily focused on the first question. Admittedly, she goes quite some way into answering the second question, as well. However, in my view, the choice to

emphasize mostly the mechanisms by which counterfactuals are created, causes us to miss out on the perhaps most relevant consequence of counterfactual generation. Counterfactuals are often a source of emotion, and it is this emotional reaction that drives many of the implications of counterfactual thinking. For example, counterfactuals serve a clear learning function, as Byrne notes on several occasions. Being aware of an upward counterfactual and the way in which it could have been obtained can clearly help one to learn from mistakes and to do better next time. However, I propose that the awareness of the counterfactual is motivating precisely because we become emotionally aroused by it.

Some of the emotional correlates of counterfactual thinking are discussed in the *The Rational Imagination* (Byrne 2005). But, the function of these emotional reactions, the role they serve in instigating behavior, have fallen beyond the scope of the book. In my review, I build upon our own work on the causes and consequences of what I would call the prototypical counterfactual emotion, *regret* (for a review, see Zeelenberg & Pieters 2007), and on our recent attempt to formulate a pragmatic approach to the role of emotions in behavior (Zeelenberg & Pieters 2006). The most important point I want to make is that to fully understand counterfactual generation, one not only needs to know how counterfactuals are constructed, but also how they elicit emotions and subsequently how these counterfactual emotions drive our future choices. The mechanisms reviewed by Byrne, which are captured in a set of seven principles (see p. 200), address counterfactual generation, but remain relatively mute with respect to the behavioral consequences.

We have been studying the relation between counterfactual thinking and the specific emotions *regret* and *disappointment* (Zeelenberg et al. 1998b). Although both stem from a comparison between "what is" and "what might have been," regret originates from comparisons between the factual decision outcome and a counterfactual outcome that might have been, *had one chosen differently*. Disappointment originates from a comparison between the factual decision outcome and a counterfactual outcome that might have been, *had another state of the world occurred*. Building on this, we realized that one should be able to shape the emotional response to a negative outcome by giving direction to the counterfactuals that are generated. In one study, participants were asked to imagine themselves being in a situation that resulted in a bad outcome that was produced by various elements in the situation, including the protagonist's own choices and also uncontrollable aspects of the situation. Thus, the situation could elicit either regret or disappointment (or a combination of the two). Next, participants were instructed to mentally undo the event either by mutating aspects of their behavior, or by mutating aspects of the situation. Next, they indicated the regret and disappointment they would feel in that situation. Interestingly, participants indicated feeling significantly more regret than disappointment when they mutated their behavior, and the opposite occurred when they mutated situational aspects. The results of this study clearly indicate the malleability of counterfactual generation and the consequences of it. Further studies (reviewed in Zeelenberg & Pieters 2007) have shown that the type of counterfactual emotion experienced strongly influences how people deal with negative events.

The non-mechanistic relation between counterfactuals and emotions is also apparent from another series of experiments. Eric van Dijk and I (Van Dijk & Zeelenberg 2005) found that the presence of counterfactuals (i.e., feedback on the outcomes of unchosen alternatives) does not always produce equally strong emotions. We asked participants to imagine being at a small fair, at which they took part in an instant scratch-card lottery. They could choose one of the only two scratch cards that were left. Participants then learned that they had won a €15 liquor-store token or a €15 book-store token. After this, participants were informed that someone else, who bought the last remaining scratch card, won either a €50 book-store token or a €50 liquor-store token. We found that complex comparisons of outcomes that were in different product categories (i.e., you win a book-store token but miss out on a liquor-store token, or vice versa) resulted in less regret than did relatively easy comparisons. Put differently, reduced comparability of counterfactual and obtained outcome attenuates the emotional response that is evoked by the counterfactual. Interestingly, we also found that for people who have a natural tendency to engage in comparison processes, the emotional response to the counterfactual outcome was also intensified when it was complex (i.e., between categories). Thus, depending on the individual characteristics of the decision maker, the same counterfactual comparison can result in mild or severe emotional reactions.

This work discussed here is of course highly selective, idiosyncratic, and incomplete, and many other relevant articles have been published. What these two articles (Van Dijk & Zeelenberg 2005; Zeelenberg et al. 1998b) show, however, is that the specific content of the counterfactual thoughts that people may generate is crucial in determining the emotional response that will follow from these thought processes. Importantly, emotions typically arise when one evaluates an event or outcome as relevant for one's concerns or preferences (Frijda 1986). Emotions do not only have a backward-looking informational function (signaling how we are doing with respect to our goals), but also a *future-oriented* motivational function (telling us what to do next). Emotions are motivational processes that prioritize certain goals and thereby mobilize and give direction to behavior. Different specific emotions thus motivate and facilitate different behavioral responses to the event that was initially responsible for the emotion. This could explain how and why different counterfactual thoughts, focusing on one's own behavior or on elements in the situation, ultimately, but via the experience of either regret or disappointment, result in completely different behaviors. In this way, the consequences of counterfactual thoughts provide input for understanding their origins. After all, the elicitation of emotions serves the purpose of behavioral regulation. Put differently, *feeling is for doing*.

Author's Response

The rational imagination and other possibilities

doi: 10.1017/S0140525X07002774

Ruth M. J. Byrne

School of Psychology and Institute of Neuroscience, Trinity College Dublin, University of Dublin, Dublin 2, Ireland.

rmbyrne@tcd.ie

http://www.tcd.ie/Psychology/Ruth_Byrne/

Abstract: In this response I discuss some of the key issues raised by the commentators on *The Rational Imagination*. I consider whether the imaginative creation of alternatives to reality is rational or irrational, and what happens in childhood cognition to enable a rational imagination to develop. I outline how thoughts about causality, counterfactuality, and controllability are intertwined and why some sorts of possibilities are more readily imagined than others. I conclude with a consideration of what the counterfactual imagination is for.

R1. Introduction

In his book *Touching the Void* (1997), Joe Simpson describes his struggle to survive following an accident when he and his friend Simon Yates made the first ascent on the West Face of Siula Grande in 1985. Their ascent to the top of the 20,813 foot mountain took three

days and by the time they reached the summit, they had exhausted most of their food supplies as well as their canisters of gas for melting snow to drink. The descent was even harder and on its second day, climbing in difficult snow conditions, Simpson fell and shattered his leg. In the remote Peruvian Andes, rescue was out of the question and Yates attempted the impossible task of a one-man recovery operation. He helped his friend down the mountain by lowering him on a rope 300 feet at a time, climbing down to him, and lowering him again. But in darkness and a blinding storm, Simpson was lowered over a cliff edge. As he hung in mid-air out of sight and hailing distance from his friend, his weight began to pull Yates, little-by-little, off the mountain. Yates made the decision to cut the rope and his friend fell into a seemingly bottomless crevasse. Yates continued the difficult climb back down to their base camp alone. He was unaware that his friend was still alive. An ice bridge had broken Simpson's fall and eventually he abseiled from it to a thin ice roof, and from there crawled back out onto the glacier. Badly injured, without food or water, and at times delirious, it took him three long days to crawl back to the base camp, and he reached it just hours before Yates was about to depart from the region entirely.

Twelve years later Simpson (1997, p. 204–205) wrote:

Analysing after a climb what you did correctly or incorrectly is as important as being fit or talented. So it was natural that for several years I too mulled over what had happened and tried to work out where we had gone wrong and what vital mistakes we had made. At first I was convinced we had done nothing wrong. I would still have back-climbed the ice cliff the way I did, although perhaps with a little more care about the quality of the ice. We would still have climbed Alpine-style, used snow-caves instead of tents and carried the same equipment and food. It was Simon who eventually pointed out to me where we had made our fatal mistake and it happened before we left base camp.

Gas.

We hadn't allowed ourselves enough gas to keep us adequately hydrated. One small canister between the two of us per day was simply insufficient. So as to save weight we had pared everything down to the minimun. It left us no room to manoeuvre when things started to go so drastically wrong. When Simon lowered me to near the Santa Rosa col, and before we committed to descending the West Face in a gathering storm and imminent darkness, we had considered digging a snow-cave and sitting the storm out. If we had done that we could have made the lowers on a bright and sunny day. We would have seen and avoided the ice cliff, and remained in control.

Instead as the storm cloud gathered over the col, we were painfully aware that we had run out of food and gas the previous night. Already dangerously dehydrated, we couldn't risk the possibility of being trapped by a prolonged storm with no way of producing fluid. I was already suffering the dehydrating and weakening effect of a traumatic fracture of a major bone and the consequent internal bleeding. We had no choice. For the sake of a cannister of gas to melt ice and snow for warm drinks we had to carry on. And so we lost control, and nearly our lives.

Simpson's reflections on what might have been are a potent illustration of the nature of counterfactual thoughts. The commentaries on *The Rational Imagination* (Byrne 2005) go to the heart of some of the key issues in understanding this most intriguing of cognitive propensities: Is the imaginative creation of alternatives to reality irrational? What happens in childhood cognition to enable a rational imagination to develop? How are thoughts about causality, counterfactuality, and controllability intertwined? Why are some sorts of possibilities more readily imagined than others? And what is the counterfactual imagination for?

R2. Is the human imagination irrational?

Is it wishful thinking to maintain that people are capable of rational imagination – indeed, capable of any sort of rational thought – in the face of so many striking demonstrations of irrational judgments, bizarre actions, inconsistent beliefs, and poorly thought-out consequences, as illustrated so eloquently in **Sternberg**'s commentary? Is the undoubted influence that emotions, personality characteristics, desires, motivations, opinions, and prejudices have upon thinking better conceived as an input, among several, into a rational process, or as a process in itself that obviates any rational analysis? Do bleak instances of individuals at their cognitive worst amount to evidence that we are irrational? Perhaps, not. As the excerpt from Joe Simpson's account of his accident in the mountains of South America highlights, the cognitive processes that people rely on to create counterfactual alternatives are capable of yielding rational imaginative outputs, just as the cognitive processes that people rely on to reason and make deductions are capable of yielding rational outputs. To understand the counterfactual imagination fully, we need to be able to explain not only what people get wrong, but also what they get right. Emotional responses may play an important role in modulating understanding and reasoning about counterfactual alternatives to reality, but they do not constitute an explanation of the process.

The observation that people sometimes make irrational judgments and sometimes make rational judgments poses a difficulty for anyone attempting to understand whether people are rational or irrational: which behavior reflects their true nature? *The Rational Imagination* advances the idea that people are rational in principle, even if they err in practice. People have the competence to be rational, but their performance may be constrained by other factors, such as their limited working memories. But, just as a glass can be viewed as half full or half empty, **Church** suggests that the opposite perspective is equally valid: the counterfactual imagination is irrational in practice, even if rational in principle. According to Church, people may not even have the competence to be rational, if their competence depends on a set of principles which limit the alternatives they consider to just a few possibilities.

The book outlines a fragment of the principles that guide the possibilities that people envisage – principles which constitute a competence to be rational (e.g., Ch. 2; pp. 16–29). The following example illustrates that reasoners have the competence to make valid deductions even though they tend to think initially about a limited number of possibilities. Consider the assertion "if the climbers dig a snow cave, they continue to the lowers in good weather." The four possibilities corresponding to the occurrence and non-occurrence of the conditional's antecedent and consequent are: (1) they dig a snow cave and continue in good weather; (2) they do not dig a snow

cave and they do not continue in good weather; (3) they do not dig a snow cave but they do continue in good weather; and (4) they dig a snow cave but they do not continue in good weather. A principle of truth guides people to think about true possibilities, they do not tend to think about what is false. A conditional interpretation of "if" rules out as false the last possibility – that the climbers dig a snow cave but they do not continue in good weather (Johnson-Laird & Byrne 2002). And a principle of parsimony guides people to think of few possibilities, hence they think initially about just a single one, in this case the first – that the climbers dig a snow cave and they continue in good weather (Johnson-Laird & Byrne 1991). People may be aware that there are alternatives to such a possibility but they have not thought them through at this initial point. They may not be able to envisage readily multiple possibilities because of the constraints of their limited working memories.

Suppose it turns out that the climbers continued in good weather. What, if anything, follows? Individuals may conclude that the climbers dug a snow cave, based on the initial possibility they have thought about. The conclusion is an error, if they have made a conditional interpretation. But individuals have the competence to go beyond their initial representation of the possibilities to consider alternatives, such as that the climbers did not dig a snow cave but they continued in good weather (perhaps because the weather changed). As a result, people discover a counterexample to the initial conclusion. The example shows that the principles which guide the possibilities that people think about, ensure that people have the competence to be rational, although they may make mistakes in practice.

There are more possibilities to think of when people reason about counterfactual alternatives compared to when they reason about facts. *The Rational Imagination* describes evidence that individuals mentally represent a counterfactual conditional such as "if the climbers had dug a snow cave, they would have continued in good weather" by envisaging more possibilities than they mentally represent for its indicative counterpart, "if the climbers dug a snow cave, they continued in good weather." When people understand the counterfactual, they think about the conjecture "they dug a snow cave and continued in good weather" and they also think about the known (or presupposed) facts "they did not dig a snow cave and they did not continue in good weather." They keep track of the epistemic status of these possibilities as real or imagined.

Shtulman proposes that the space of counterfactual possibilities may be so large that it may be mistaken for people to apply strategies they have developed for thinking about facts when they think about counterfactual alternatives. He suggests that the search for a counterexample to an inference based on a counterfactual may fail because of the difficulty of considering many counterfactual possibilities, rather than because of the absence of a counterexample, and so the search will not guarantee a valid deduction. Nonetheless, the data show that people appear to be able to identify counterexamples to counterfactual inferences and to reject conclusions as invalid from counterfactual premises (Thompson & Byrne 2002). For example, given the counterfactual conditional, "if the car had been out of petrol, it would have stalled" people reject the denial of the antecedent inference from the premise "the car was not out of petrol" to the conclusion "therefore it did not stall." They reject the inference because they can think of the possibility, "the car was not out of petrol and it stalled," which provides a counterexample to the conclusion. They reject the inference more readily from a counterfactual conditional than from a counterfactual biconditional, such as "if the water had been heated to 100 degrees, it would have boiled" for which they cannot think of a counterexample (Byrne 2005, Ch. 6, pp. 150–52). The result shows that people are able to construct and process counterexamples when they think about counterfactual alternatives.

When people create counterfactual alternatives, there are striking regularities in what they focus on: for example, actions, controllable events, or the last event in a temporal sequence. These regularities occur even in very different domains – ranging from historical analysis to fantasy – and the mutations made to the mental representation of the facts in each domain may reflect the constraints of the domain, as **Legrenzi** observes. The tendencies to focus on actions, controllable events, and so on are exhibited not only by individuals in their everyday lives, but also by experts such as scientists and engineers who may generate novel solutions to domain-specific problems by constructing counterfactual scenarios, as **Chandrasekharan & Nersessian** note. Is the observation that even experts exhibit tendencies to focus on actions or controllable events, and that they must develop ways to overcome these tendencies, an indictment of the idea of a rational imagination? To view these tendencies as "biases" misses their point: The regularities that people exhibit reflect a mechanism for computing minimal changes, based on a rational exploitation of the fault-lines that occur in the mental representation of reality (Ch. 8, pp. 189–90).

R3. How does the rational imagination develop?

If reasoning and imagination depend on the same processes, it makes sense to ask, as **Van Reet, Pinkham, & Lillard** (**Van Reet et al.**) do, whether these processes develop in children initially in the service of reasoning or in the service of imagination. *The Rational Imagination* examines the evidence that the counterfactual imagination relies on the same sorts of principles as reasoning does. The key claim is that just as reasoning has been found to rely on imagination (e.g., Johnson-Laird & Byrne 1991), so too imagination relies on reasoning. Increasingly, evidence from the development of reasoning in children indicates that counterfactual thought is the nexus for the development of key processes in *both* reasoning and imagination.

A major development in childhood is the development of working memory, and consequently, the ability to envisage alternative possibilities. Children must also learn to keep track of the epistemic status of alternatives, as corresponding to reality or conjecture (e.g., Ch. 5, pp. 122–23). As **Markovits** observes, not only do quantitative changes occur in the number of possibilities that children can consider, but also qualitative changes in the nature of the possibilities, from experientially based to more abstract possibilities (an observation also made by **Agassi**). Markovits suggests that the imagination of inconsistent possibilities may be a key mechanism by which the nature of the processes for rational thought change. Evidence from

https://doi.org/10.1017/S0140525X07002658 Published online by Cambridge University Press

adults suggests that resolving contradictions can indeed bring about belief change; and the tendency to revise beliefs by altering either the belief in the empirical evidence or the belief in the theoretical explanation may reflect the development of different strategies to deal with inconsistency (e.g., Ch. 8, pp. 181–89). Adults appear to rely on one or other of these strategies to resolve contradictions of different sorts. But whether or not these observed changes in the content of beliefs reflect developmental changes in the very processes by which beliefs are formed and maintained, as Markovits implies, is as yet unknown.

Children begin to appreciate counterfactuals as early as 2 and 3 years of age, but their understanding may not approximate adult comprehension until the later years of childhood when they have developed the ability to suppress temporarily their belief in the presupposed facts, to suppose temporarily the counterfactual possibility to be true, and to compute a comparison between the two, as **Riggs & Beck** note (a point also made by **Van Reet et al.**). Indeed, recent evidence shows that in adulthood the ability to switch attention away from one possibility to focus on a different possibility remains a crucial skill for counterfactual reasoning. Switching attention from one task to another or from one object to another carries a cost (in terms of both accuracy and latency): It is harder to do a task when the task before it was a different one compared to when it was the same task (for a review, see Pereda et al., submitted). Attention switching costs vary depending on the difficulty of the task, and a long-standing but somewhat counterintuitive finding has been that there is a greater cost to switching attention from an easy task into a hard task, compared to vice versa.

In a recent experiment, we observed such asymmetrical attention switching costs in adults when they were reasoning from indicative and counterfactual conditionals (Pereda et al., submitted). *Modus ponens* is an easy inference because it requires a single possibility to be envisaged and *modus tollens* is hard because it requires two possibilities to be fleshed out (Ch. 2, pp. 22–28). Over several hundred trials, adults continued to take more time to carry out the hard *modus tollens* inference (if A then B, not-B therefore not-A) when they carried it out after the easy *modus ponens* one (if A then B, A therefore B), compared to when they carried it out after a similar *modus tollens* one. However, they found it just as easy to carry out the easy *modus ponens* inference whether they carried it out after the hard *modus tollens* inference or a similar *modus ponens* one (Pereda et al., submitted). The discovery of an asymmetrical attention-switching cost for inferences from conditionals advances the view that switching attention effectively between two possibilities is a crucial skill required for good reasoning. The development of this attention switching skill may be important in childhood for the creation and comprehension of counterfactual alternatives.

R4. Counterfactual thought, causal thought, and controllability

Joe Simpson's thoughts about his catastrophic experiences in the mountains identify the root cause of the disaster as the decision about how much gas to carry:

> *It was Simon who eventually pointed out to me where we had made our fatal mistake and it happened before we left base camp. Gas. We hadn't allowed ourselves enough gas to keep us adequately hydrated.* (Simpson 1997, p. 204)

But his thoughts about how things could have turned out differently focus instead on the decision to continue climbing rather than sit the storm out:

> *. . . we had considered digging a snow-cave and sitting the storm out. If we had done that we could have made the lowers on a bright and sunny day. We would have seen and avoided the ice cliff, and remained in control.* (Simpson 1997, p. 204–205)

The divergence in the focus of causal and counterfactual thoughts is curious, as discussed in *The Rational Imagination* (Ch. 5). **Spellman & Ndiaye** are disinclined to believe that such a divergence reflects anything other than methodological conventions of measurement in laboratory studies. But its ready occurrence in spontaneous causal and counterfactual thoughts undermines their claim (e.g., Ch. 5, pp. 100–101). Causes and counterfactual alternatives are intricately related, as illustrated in Simpson's subsequent intertwining of the two:

> *Instead . . . we were painfully aware that we had run out of food and gas the previous night . . . we couldn't risk the possibility of being trapped by a prolonged storm with no way of producing fluid . . . For the sake of a cannister of gas to melt ice and snow for warm drinks we had to carry on.* (Simpson 1997, p. 205)

The Rational Imagination proposes that the close interconnectedness of causal and counterfactual thoughts, and the frequent divergence in their focus, arises because counterfactual thoughts tend to focus on enabling conditions and causal thoughts tend to focus on strong causes (e.g., Ch. 5).

A strong cause, for example, "the lack of gas caused the climbers' descent in poor conditions" is consistent with two alternative possibilities: (1) a lack of gas and a poor descent, (2) a supply of gas and a good descent. The principles that guide the possibilities that people think about ensure that people tend initially to mentally represent just the first possibility. By contrast, an enabling condition, for example, "sitting the storm out allowed the climbers to descend in good conditions" is consistent with three possibilities: (1) sitting the storm out and a good descent, (2) not sitting the storm out and a poor descent, (3) sitting the storm out and a poor descent (Johnson-Laird & Byrne 2002). People may tend initially to think about the first possibility, and they are usually able to think readily about the second possibility. **Spellman & Ndiaye** claim that strong causes and enablers do not differ in their logical meaning, noting that people may interpret an event as a cause in one context and as an enabler in another. But the key distinction occurs after people have reached an interpretation of an event as one or the other – a strong cause and an enabler are consistent with different possibilities, and people mentally represent them in different ways. Likewise, **Mandel**'s contention that strong causes are by definition also enablers and therefore counterfactual thoughts should focus on them, too, overlooks the difference in their mental representation.

The account in *The Rational Imagination* led to the prediction that people should create more causal thoughts (which require them to think about a single possibility), rather than counterfactual thoughts (which require them

to think about two possibilities). The evidence corroborates the prediction: people spontaneously produce more causal thoughts than counterfactual thoughts (Ch. 5, pp. 123–25; McEleney & Byrne 2006). The observation that people can be provoked to produce more counterfactual thoughts than causal ones in specific situations, as described by **Spellman & Ndiaye**, is explained by prompts, such as questions or scenarios leading people to flesh out more possibilities.

The frequent focus of counterfactual thoughts on actions that could prevent an outcome leads **Mandel** to propose that counterfactuals focus on disablers rather than enablers. Of course, disablers and enablers share much in common; for example, the enabler "dry leaves enable fire" and the disabler "wet leaves prevent fire" are consistent with the same three possibilities – the assertions describe the same situations, albeit in different ways (Ch. 5, p. 117).

An enabler, for example, "dry leaves enable fire," is consistent with the three possibilities: "dry leaves and fire," "no dry leaves and no fire," and "dry leaves and no fire." A disabler (or a missing enabler), "no dry leaves," can be expressed by implicit negation as "wet leaves" (on the assumption of a binary situation in which there are leaves which are either dry or wet), and the disabling relation can be expressed by "prevent," as in "wet leaves prevent fire." The disabler is consistent with the same three possibilities: "wet leaves and no fire," "no wet leaves and fire," and "no wet leaves and no fire," as Table R1 shows. But the different descriptions bring different possibilities most readily to mind: the enabler is understood initially by envisaging "dry leaves and fire," with ready access to "no dry leaves and no fire"; the disabler is understood initially by envisaging "wet leaves and no fire," with ready access to "no wet leaves and fire" (Ch. 5, pp. 117–18). The first possibility envisaged for the enabler, "dry leaves and fire," corresponds to the presupposed facts for its corresponding counterfactual, "if there had been no dry leaves there would have been no fire"; the first possibility envisaged for the disabler "wet leaves and no fire" corresponds to the imagined conjecture in its corresponding counterfactual "if there had been wet leaves there would have been no fire."

These nuances cannot be captured within **Mandel**'s account that counterfactuals are understood by envisaging the facts only. The facts are known for some counterfactuals (e.g., "if Simpson had not climbed Siula Grande, he would not have fallen into a crevasse") and they are unknown for others, such as "if Simpson had brought another gas canister, he would not have fallen into a crevasse." But the evidence shows that people think about two possibilities even for counterfactuals with unknown facts (Ch. 2, pp. 30–34; Ch. 3, pp. 48–52). If people thought only about the facts when they understood a counterfactual, they would be unable to make a *modus ponens* inference (if A then B, A therefore B). In fact, when asked to consider a counterfactual and a supposition, such as, "suppose Simpson did not climb Siula Grande," most people can conclude readily, "in that case he did not fall into a crevasse."

Counterfactual thoughts tend to focus on enablers rather than strong causes, and people tend to focus on events they can control rather than on uncontrollable ones. The two tendencies often coincide but the theory does not confound them, contrary to **Chang & Herrmann**'s charge – enablers can be outside an individual's control, as illustrated in the example from Chapter 5, "dry leaves enabled the forest fire." The principles outlined in the book explain how people think about counterfactual possibilities that may once have been possible but are no longer possible. Chang & Herrmann question whether the set of principles constitute an explanation for how people create counterfactual alternatives to reality or merely a description. The query is belied by their acknowledgement that the principles lead to testable empirical predictions. The predictions are corroborated by the data (e.g., Ch. 5, pp. 119–26). The central importance of controllability in counterfactual and causal thoughts is brought into sharp relief in Simpson's conclusion from his analysis of the events leading up to his accident – "and so we lost control, and nearly our lives" (Simpson 1997, p. 205).

Table R1. *The relations between enablers and disablers and their corresponding counterfactuals*

Causal relation	Counterfactual
1. Dry leaves enable fire	*If there had been no dry leaves there would have been no fire*
dry leaves and fire	dry leaves and fire (facts)
no dry leaves and no fire	no dry leaves and no fire (imagined)
dry leaves and no fire	dry leaves and no fire
2. Wet leaves prevent fire	*If there had been wet leaves there would have been no fire*
wet leaves and no fire	wet leaves and no fire (imagined)
no wet leaves and fire	no wet leaves and fire (facts)
no wet leaves and no fire	no wet leaves and no fire

Note: People tend to think about two possibilities when they understand a counterfactual, the presupposed facts and the conjecture referred to explicitly in the conditional.

R5. Possibilities for "if"

People understand a conditional, such as, "if there were puppies, there were kittens" by thinking about possibilities. They think about true possibilities – there were puppies and there were kittens, there were no puppies and there were no kittens, there were no puppies and there were kittens. They think initially about few possibilities – there were puppies and there were kittens. They do not think about false possibilities – there were puppies and there were no kittens (Johnson-Laird & Byrne 2002). But **Handley & Feeney** propose instead that people think about the situations in which the antecedent occurs, in which there were puppies; that is, they think about the true situation "there were puppies and there were kittens" and also the false situation, "there were puppies and there were no kittens." Contrary to Handley & Feeney's proposal, recent evidence shows that people do not think about false possibilities (Espino et al.,

submitted). The Espino et al. experiment compared conditionals and "biconditionals," such as "if and only if there were puppies, there were kittens." The two sorts of assertion differ in relation to the possibility in which there were no puppies but there were kittens: it is a true possibility for the conditional but it is a false possibility for the biconditional. The distinction leads to an important prediction. The view advanced in *The Rational Imagination*, that people think only about true possibilities, predicts that they should think about the possibility "there were no puppies and there were kittens" when they understand a conditional – for which it is true, but they should not think about it when they understand a biconditional – for which it is false. In the experiment, participants read the conditionals or biconditionals and the various conjunctions within a story (see Espino et al., submitted), such as the following:

> Joey went shopping to the pet-shop. When he looked at the poster he saw written on it, "if there are puppies, there are kittens." When he looked at the cages he saw that there were puppies and there were kittens. He checked his list of purchases.

Participants took a very short time, under 2 milliseconds, to read the conjunction "there were no puppies and there were kittens" when it was "primed" by a conditional for which it is a true possibility; they took reliably longer to read the same conjunction when it was primed by the biconditional, for which it is a false possibility. And they took the same amount of time to read the possibility "there were puppies and there were kittens" when it was primed by either the conditional or the biconditional – it is a true possibility for both assertions. These data corroborate the view that people think about true possibilities and not false possibilities (pace **Handley & Feeney**).

For counterfactual conditionals, people think about the presupposed facts, and they also think about the counterfactual conjecture, a false possibility temporarily supposed to be true. Some people may tend to think only about the presupposed facts, and some people may find their attention is focused on the counterfactual conjecture (Ch. 8, pp. 182–89; Byrne & Tasso 1999; Thompson & Byrne 2002). **Markman & McMullen** provide an important reminder of the tendency people exhibit to suspend sometimes any comparison of the counterfactual possibility with their representation of reality, in favor of vividly imagining the counterfactual possibility "as if" it were the case. It may be unlikely that a counterfactual possibility is ever entirely unfettered by comparison with reality: The initial construction of the counterfactual depends on mutations to the representation of reality, and its epistemic status needs to be maintained to ensure an ongoing appreciation that the counterfactual conjecture does not correspond to reality. The crucial distinction may depend on whether the counterfactual possibility is maintained within the primary focus of attention, or on what the current reality is, and what the cognitive and affective causes and consequences of focusing on one possibility or the other may be.

A key mechanism by which people create counterfactual possibilities is the mutation of those aspects of reality that are explicitly represented in their mental representation of reality (e.g., Ch. 2, pp. 34–38; Ch. 7, pp. 167–80). The book contains a brief speculation that a similar mechanism may underlie the creation of novel possibilities when people invent new ideas by expanding existing categories or combining existing concepts (Ch. 8, pp. 190–94). The nature of the possibilities that people entertain when they create new ideas may be influenced by the structure of their ideas, as **Ward** suggests. Their mental representation of an assertion corresponds to the structure of the way the world would be if the assertion were true, rather than to the language used to describe the world (Johnson-Laird & Byrne 1991); in this important respect, the possibilities that people think about to represent the structure of the world are guided by such principles as truth and parsimony.

How people perceive and represent reality affects how they create counterfactual alternatives to it. It is not just what people imagine but how easy or hard they find it to imagine that determines their judgment, as **Šanna** has shown. Recent evidence highlights that what individuals find easy to imagine differs depending on whether they imagine alternatives to something that happened to them, or imagine alternatives to something that happened to someone they have merely read about (Girotto et al. 2007). Consider the scenario:

> Anna, an undergraduate at your university, was asked to participate in a game by a research assistant who told her: "In order to win two chocolates, you have to mentally multiply either two one-digit numbers or two two-digit numbers, in 30 seconds. If you fail, you do not receive the chocolates. The two multiplication problems are contained in two sealed envelopes. Let us call them envelope A and envelope B. Of course, we do not know which envelope contains the one-digit multiplication problem and which one contains the two-digit multiplication problem." Anna agreed to participate. She chose envelope A. It contained the two-digit multiplication problem. She failed. Things would have been better for Anna, if . . .

Most readers tend to construct counterfactuals that change Anne's choice, e.g., "if only she had chosen the other envelope." But when individuals carry out the task as actors who take part in the game just as Anna did, rather than as readers, their counterfactuals focus on various problem features, such as "if I had had more time" or "if I had had a pen" (Girotto et al. 2007).

Following the discovery of an actor–reader difference in how people create counterfactual alternatives, we examined whether observers witnessing a situation think about it in the same way as readers or as actors (Pighin et al., submitted). Do you expect an observer to imagine an alternative in the same way as the actor imagines it, for example, by focusing on the time or the aids available, or to imagine an alternative in the same way as a reader imagines it, for example, by focusing on the choice of envelope? In a new experiment, observers were experimental participants who were present during the sequence of events that led an actor (a confederate) to experience the negative outcome. The data show that observers behaved just like actors: they tended to focus on problem features such as "if only there were more time." Like actors, observers can gather information about the events that lead to the outcome, including not only the actors' choice but also the features of the problem-solving phase, such as the perceived shortness of the time available for the task, or the lack of a pen and paper (Pighin et al., submitted). The discovery that actors and observers create counterfactuals that are different from the ones readers create has far-reaching consequences for the methods used in the study of counterfactual thoughts.

R6. Counterfactual thoughts and the future

The commentaries on *The Rational Imagination* raise many wide-ranging and challenging questions about how we can reach a proper understanding of the way the human mind creates alternatives to reality, and my response has focused on the general and common themes. What is undisputed is that people engage in counterfactual thoughts frequently and it is a lynchpin of their mental, emotional, and social lives. The rationality of the counterfactual imagination may be crucial in its impact on decisions about a future goal that an individual envisages, such as climbing a mountain safely. **Epstude & Roese** propose that a key role of counterfactual thoughts is in goal-oriented decisions. **Bonnefon** notes that the influence of a rationale for action on the experience of regret is mediated through thoughts about future consequences (see also Walsh & Byrne 2007). And **Zeelenberg** distinguishes between the influence of regret and of disappointment on future decisions for action. An individual may identify the means to achieve a goal by carrying out an accurate analysis of errors made in previous experiences, for instance, potential errors in the style of climbing, the equipment chosen, or the supplies carried. The principles that guide the possibilities that people construct to create a counterfactual alternative provide a basis for a mechanism that transforms the representation of past events and outcomes into the representation of future plans and goals. An important function of counterfactual thoughts may be to learn to recover from failure. Counterfactual thoughts may help us in our efforts to try again, as Samuel Beckett (1983) put it: "Try again. Fail again. Fail better."

ACKNOWLEDGMENT

I am grateful to Vittorio Girotto for giving me Joe Simpson's book, *Touching the Void*.

References

Anderson, R. A. (1984) Did I do it or did I only imagine doing it? *Journal of Experimental Psychology: General* 113:594–613. [KDM]

Ansburg, P. I. & Dominowski, R. L. (2000) Promoting insightful problem-solving. *Journal of Creative Behaviour* 34(1):30–60. [aRMJB]

Associated Press (2007) His own words likely to confront Vitter. Available at: http://www.foxnews.com/story/0,2933,288740,00.html/ [Retrieved September 12, 2007.] [RJS]

Athanasiadou, A. & Dirven, R., eds. (1997) *On conditionals again*. John Benjamins. [aRMJB]

Atran, S. (2002) *In gods we trust*. Oxford University Press. [PL]

Baron, J. (1988/2000) *Thinking and deciding*, 3rd edition. Cambridge University Press. (Original work published 1988.) [J-FB]

Barrouillet, P. & Lecas, J.-F. (1999) Mental models in conditional reasoning and working memory. *Thinking and Reasoning* 5:289–302. [aRMJB]

Beck, S. R., Riggs, K. J. & Gorniak, S. L. (submitted) Are developments in preschoolers' counterfactual thinking driven by developments in executive functions? *Child Development*. [KJR]

Beck, S. R., Robinson, E. J., Carroll, D. J. & Apperly, I. A. (2006) Children's thinking about counterfactuals and future hypotheticals as possibilities. *Child Development* 77(2):413–26. [KJR]

Beckett, S. (1983) *Worstward Ho!* Grove/Atlantic. [rRMJB]

Bock, J. & Clifton, C. (2000) The role of salience in conceptual combination. *Memory and Cognition* 28:1378–86. [TBW]

Bodenhausen, G. V. & Wyer, R. S. (1987) Social cognition and social reality: Information acquisition and use in the laboratory and the real world. In: *Social information processing and survey methodology*, ed. H. J. Hippler, N. Schwarz & S. Sudman, pp. 6–41. Springer-Verlag. [LJS]

Boninger, D. S., Gleicher, F. & Strathman, A. (1994) Counterfactual thinking: From what might have been to what may be. *Journal of Personality and Social Psychology* 67:297–307. [aRMJB]

Bonnefon, J. F. & Hilton, D. J. (2004) Consequential conditionals: Invited and suppressed inferences from valued outcomes. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30:28–37. [J-FB]

Bonnefon, J. F., Zhang, J. & Deng, C. (2007) Is the effect of justifications on regret direct or indirect? *Revue Internationale de Psychologie Sociale – International Review of Social Psychology* 20:131–45. [J-FB]

Braine, M. D. S. & O'Brien, D. (1998) *Mental logic*. Erlbaum. [aRMJB]

Brandstätter, V., Lengfelder, A. & Gollwitzer, P. M. (2001) Implementation intentions and efficient action initiation. *Journal of Personality and Social Psychology* 81:946–60. [KE]

Branscombe, N. R., N'gbala, A., Kobrynowicz, D. & Wann, D. L. (1997) Self and group protection concerns influence attributions but they are not determinants of counterfactual mutation focus. *British Journal of Social Psychology* 36:387–404. [aRMJB]

Branscombe, N. R., Owen, S., Garstka, T.A. & Coleman, J. (1996) Rape and accident counterfactuals: Who might have done otherwise and would it have changed the outcome? *Journal of Applied Social Psychology* 26:1042–67. [aRMJB]

Bucciarelli, M. & Johnson-Laird, P. N. (2005) Naïve deontics: A theory of meaning, representation, and reasoning. *Cognitive Psychology* 50:159–93. [aRMJB]

Byrne, R. M. J. (1997) Cognitive processes in counterfactual thinking about what might have been. In: *The psychology of learning and motivation: Advances in research and theory, vol. 37*, ed. D. Medin, pp. 105–54. Academic Press. [aRMJB]

(2005) *The rational imagination: How people create alternatives to reality*. MIT Press. [JA, J-FB, arRMJB, SC, WC, JC, KE, SJH, PL, DRM, KDM, HM, KJR, LJS, AS, BAS, RJS, JVR, TBW, MZ]

Byrne, R. M. J., Espino, O. & Santamaria, C. (1999) Counterexamples and the suppression of inferences. *Journal of Memory and Language* 40:347–73. [aRMJB]

Byrne, R. M. J. & McEleney, A. (2000) Counterfactual thinking about actions and failures to act. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26:1318–31. [aRMJB]

Byrne, R. M. J., Segura, S., Culhane, R., Tasso, A. & Berrocal, P. (2000) The temporality effect in counterfactual thinking about what might have been. *Memory and Cognition* 28:264–81. [aRMJB]

Byrne, R. M. J. & Tasso, A. (1999) Deductive reasoning with factual, possible, and counterfactual conditionals. *Memory and Cognition* 27:726–40. [arRMJB]

Carver, C. S. & Scheier, M. F. (1998) *On the self-regulation of behavior*. Cambridge University Press. [KE]

Catellani, P. & Milesi, P. (2001) Counterfactuals and roles: Mock victims' and perpetrators' accounts of judicial cases. *European Journal of Social Psychology* 31:247–64. [aRMJB]

Chandrasekharan, S. (2005) Epistemic structure: An inquiry into how agents change the world for cognitive congeniality. Unpublished doctoral dissertation, Carleton University, Ottawa, Canada. Available as Carleton University Cognitive Science Technical Report, at: http://www.carleton.ca/iis/TechReports/files/2005-02.pdf. [SC]

Cheng, P. N. & Holyoak, K. J. (1985) Pragmatic reasoning schemas. *Cognitive Psychology* 17:391–416. [aRMJB]

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

Chisholm, R. M. (1946) The contrary-to-fact conditional. *Mind* 2005:289–307. [aRMJB]

Cohen, L. J. (1986) *The dialogue of reason*. Clarendon Press. [JC]

Collingwood, R. G. (1940) *An essay on metaphysics*. Clarendon. [DRM]

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* 8:1255–62. [KE]

Cosmides, L. (1989) The logic of social exchange. *Cognition* 31:187–276. [aRMJB]

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

(2001) Testing two theories of conceptual combination: Alignment and diagnosticity in the comprehension and production of combined concepts. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27:255–71. [aRMJB, TBW]

Davidson, R. J., Scherer, K. R. & Goldsmith, H. H., eds. (2002) *Handbook of affective sciences*. Oxford University Press. [RJS]

Davis, C. G., Lehman, D. R., Silver, R. C., Wortman, C. M. & Ellard, J. H. (1996) Self-blame following a traumatic event: The role of perceived avoidability. *Personality and Social Psychology Bulletin* 22:557–67. [aRMJB]

Davis, C. G., Lehman, D. R., Wortman, C. M., Silver, R. C. & Thompson, S. C. (1995) The undoing of traumatic life events. *Personality and Social Psychology Bulletin* 21:109–24. [aRMJB]

Dennett, D. C. (1993) *Consciousness explained*. Penguin Books. [AS]

Dias, M. G. & Harris, P. L. (1988) The effect of make-believe play on deductive reasoning. *British Journal of Developmental Psychology* 6:207–21. [JVR]
(1990) The influence of the imagination on reasoning by young children. *British Journal of Developmental Psychology* 8:305–18. [JVR]
Donaldson, S. R. (1991) *The gap into conflict: The real story*. Bantam. [TBW]
Dunning, D. & Parpal, M. (1989) Mental addition versus subtraction in counterfactual reasoning: On assessing the impact of personal actions and life events. *Journal of Personality and Social Psychology* 57:5–15. [aRMJB]
Einhorn, H. J. & Hogarth, R. M. (1986) Judging probable cause. *Psychological Bulletin* 99:3–19. [DRM, BAS]
Epstude, K. & Roese, N. J. (submitted) The functional theory of counterfactual thinking. *Personality and Social Psychology Review*. [KE]
Espino, O., Santamaria, C. & Byrne, R. M. J. (submitted) Conditionals and biconditionals prime true possibilities, not false possibilities. [rRMJB]
Estes, Z. (2003) A tale of two similarities: Comparison and integration in conceptual combination. *Cognitive Science: A Multidisciplinary Journal* 27:911–21. [TBW]
Evans, J. St. B. T. (1977) Linguistic factors in reasoning. *Quarterly Journal of Experimental Psychology* 29:297–306. [aRMJB]
(1989) *Bias in reasoning*. Erlbaum. [aRMJB]
Evans, J. St. B. T., Handley, S. J. & Over, D. E. (2003) Conditionals and conditional probability. *Journal of Experimental Psychology: Learning, Memory and Cognition* 29:321–55. [SJH]
Evans, J. St. B. T. & Over, D. E. (1996) *Rationality and reasoning*. Psychology Press. [J-FB]
Evans, J. St. B. T., Over, D. E. & Handley, S. J. (2005) Supposition, extensionality and conditionals: A critique of the mental model theory of Johnson-Laird and Byrne (2002). *Psychological Review* 112:1042–52. [SJH]
Fattah, H. M. (2007) Radicalism among Muslim professionals worries many. *New York Times*, Saturday, July 14, 2007 [156(54005)], p. A3. [RJS]
Feeney, A. & Handley, S. J. (2006) Mental models, comparisons, and the action effect in judgements of regret. *Memory and Cognition* 34:1422–30. [SJH]
Fiddick, L., Cosmides, L. & Tooby, J. (2000) No interpretation without representation: The role of domain-specific representations and inferences in the Wason selection task. *Cognition* 77:1–79. [aRMJB]
Finke, R. A., Ward, T. B. & Smith, S. M. (1992) *Creative cognition: Theory, research and applications*. MIT Press. [TBW]
Fischhoff, B. (1975) Hindsight ≠ foresight: The effect of outcome knowledge on judgments under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance* 1:288–99. [LJS]
Fishbach, A. & Ferguson, M. J. (2007) The goal construct in social psychology. In: *Social psychology: Handbook of basic principles*, 2nd edition, ed. A. W. Kruglanski & E. T. Higgins, pp. 490–515. Guilford. [KE]
Friedrich, J. (1993) Primary error detection and minimization (PEDMIN) strategies in social cognition: A reinterpretation of confirmation bias phenomena. *Psychological Review* 100:298–319. [DRM]
Frijda, N. H. (1986) *The emotions*. Cambridge University Press. [MZ]
Gagné, C. (2000) Relation-based combinations versus property-based combinations: A test of the CARIN theory and the dual-process theory of conceptual combination. *Journal of Memory and Language* 42:365–89. [TBW]
Gagné, C., Spalding, T. & Ji, H. (2005) Re-examining evidence for the use of independent relational representations during conceptual combination. *Journal of Memory and Language* 53:445–55. [TBW]
Galinsky, A. D. & Moskowitz, G. B. (2000) Counterfactuals as behavioural primes: Priming the simulation of heuristics and consideration of alternatives. *Journal of Experimental Social Psychology* 36:384–409. [aRMJB]
Galinsky, A. D., Moskowitz, G. B. & Skurnik, I. W. (2000) Counterfactuals as self-generated primes: The effect of prior counterfactual activation on person perception judgments. *Social Cognition* 18:252–80. [KE]
Gigerenzer, G. (2007) *Gut feelings: The intelligence of the unconscious*. Viking. [RJS]
Gigerenzer, G. & Hug, K. (1992) Domain specific reasoning: Social contracts, cheating, and perspective change. *Cognition* 43:127–71. [aRMJB]
Gigerenzer, G. & Selten, R., eds. (2001) *Bounded rationality: The adaptive toolbox*. MIT Press. [aRMJB]
Gilbert, D. T., Krull, D. S. & Malone, P. S. (1990) Unbelieving the unbelievable: Some problems in the rejection of false information. *Journal of Personality and Social Psychology* 59:601–13. [KDM]
Gilovich, T. & Medvec, V. H. (1995) The experience of regret: What, when, and why. *Psychological Review* 102:379–95. [aRMJB]
Ginsberg, M. L. (1986) Counterfactuals. *Artificial Intelligence* 30:35–79. [aRMJB]
Girotto, V., Ferrante, D., Pighin, S. & Gonzalez, M. (2007) Post-decisional counterfactual thinking by actors and readers. *Psychological Science* 18:510–15. [rRMJB]
Girotto, V., Legrenzi, P. & Rizzo, A. (1991) Event controllability in counterfactual thinking. *Acta Psychologia* 78:111–33. [aRMJB]
Goldvarg, E. & Johnson-Laird, P. N. (2001) Naive causality: A mental model theory of causal meaning and reasoning. *Cognitive Science* 25(4):565–610. [aRMJB]
Gordon, M. R. & Rutenberg, J. (2007) Bush distorts Qaeda links, critics assert. *New York Times*, Friday, July 13, 2007 [156(54004)], pp. A1, A6. [RJS]
Guilbault, R. L., Bryant, F. B., Posavac, E. J. & Brockway, J. H. (2004) A meta-analysis of research on hindsight bias. *Basic and Applied Social Psychology* 26:103–17. [LJS]
Guttentag, R. E. & Ferrell, J. M. (2004) Reality compared with its alternatives: Age differences in judgments of regret and relief. *Developmental Psychology* 40(5):764–75. [KJR]
Handley, S. J., Evans, J. St. B. T. & Thompson, V. (2006) The negated conditional: Litmus test for the suppositional conditional? *Journal of Experimental Psychology: Learning Memory and Cognition* 13:378–95. [SJH]
Handley, S. J. & Feeney, A. (2004) Reasoning and pragmatics: The case of EVEN IF. In: *Towards experimental pragmatics*, ed. I. Noveck & D. Sperber, pp. 228–53. Palgrave MacMillan. [SJH]
(2007) Representation, pragmatics and process in model-based reasoning. In: *The mental models theory of reasoning: Refinement and extensions*, ed. W. Schaeken, A. Vandierendonck, W. Schroyens & G. d'Ydewalle, pp. 25–52. Erlbaum. [SJH]
Harley, E. M., Carlsen, K. A. & Loftus, G. R. (2004) The "saw-it-all-along" effect: Demonstrations of visual hindsight bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30:960–68. [LJS]
Harris, P. (2000) *The work of the imagination*. Blackwell. [aRMJB]
Harris, P. L., German, T. & Mills, P. (1996) Children's use of counterfactual thinking in causal reasoning. *Cognition* 61:233–59. [aRMJB]
Hart, H. L. & Honore, A. M. (1959) *Causation and the law*. Clarendon Press. [aRMJB]
Hawkins, J., Pea, R., Glick, J. & Scribner, S. (1984) Merds that laugh don't like mushrooms: Evidence for deductive reasoning in preschoolers. *Developmental Psychology* 20:584–94. [JVR]
Hawkins, S. A. & Hastie, R. (1990) Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin* 107:311–27. [LJS]
Higgins, E. T. (2006) Value from hedonic experience and engagement. *Psychological Review* 113:439–60. [KE]
Hofstadter, D. R. (1985) *Metamagical themas: Questing for the essence of mind and pattern*. Penguin. [aRMJB]
Holyoak, K. J. & Cheng, P. W. (1995) Pragmatic reasoning with a point of view. *Thinking and Reasoning* 1:289–313. [aRMJB]
Hume, D. (1739/2000) *A treatise of human nature*. Reprint edition, ed. D. F. Norton & M. J. Norton. Oxford University Press. [aRMJB]
Jackson, F. (1987) *Conditionals*. Blackwell. [SJH]
Janveau-Brennan, G. & Markovits, H. (1999) Reasoning with causal conditionals: Developmental and individual differences. *Developmental Psychology* 35(4):904–11. [HM]
Jeffrey, R. (1981) *Formal logic: Its scope and limits*, 2nd edition. McGraw-Hill. [aRMJB]
Johnson, J. T. (1986) The knowledge of what might have been: Affective and attributional consequences of near outcomes. *Personality and Social Psychology Bulletin* 12:51–62. [aRMJB]
Johnson, M. K. & Raye, C. L. (1981) Reality monitoring. *Psychological Review* 88:67–85. [KDM]
Johnson-Laird, P. N. (1983) *Mental models*. Cambridge University Press. [aRMJB, KE]
(2001) Mental models and deduction. *Trends in Cognitive Sciences* 5:434–42. [aRMJB]
Johnson-Laird, P. N. & Byrne, R. M. J. (1989) *Only* reasoning. *Journal of Memory and Language* 28:313–30. [aRMJB]
(1991) *Deduction*. Erlbaum. [arRMJB]
(2002) Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review* 109:646–78. [aRMJB, DRM]
Kahneman, D. & Miller, D. (1986) Norm theory: Comparing reality to its alternatives. *Psychological Review* 93:136–53. [aRMJB, SC, KE, LJS]
Kahneman, D. & Tversky, A. (1972) Subjective probability: A judgment of representativeness. *Cognitive Psychology* 3:430–54. [RJS]
(1982) The simulation heuristic. In: *Judgment under uncertainty: heuristics and biases*, ed. D. Kahneman, P. Slovic & A. Tversky, pp. 201–208. Cambridge University Press. [aRMJB, SC, KE, LJS]
Keane, M. T. (1997) What makes an analogy difficult? The effects of order and causal structure in analogical mapping. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 123:946–67. [aRMJB]
King, M. L (1968) "I've been to the mountaintop" speech delivered on 3rd April, 1968, in Memphis, Tennessee. Available online at: http://www.americanrhetoric.com/speeches/mlkivebeentothemountaintop.htm [aRMJB]
Klauer, K. C., Jacobsen, T. & Migulla, G. (1995) Counterfactual processing: Test of a hierarchical correspondence model. *European Journal of Social Psychology* 25:577–95. [aRMJB]
Klayman, J. & Ha, Y.-W. (1987) Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review* 94:211–28. [DRM]

Knight, R. T. & Grabowecky, M. (1995) Escape from linear time: Prefrontal cortex and conscious experience. In: *The cognitive neurosciences*, ed. M. S. Gazzaniga, pp. 1357–71. MIT Press. [aRMJB]

Kocovski, N. L., Endler, N. S., Rector, N. A. & Flett, G. L. (2005) Ruminative coping and post-event processing in social anxiety. *Behaviour Research and Therapy* 43:971–84. [KE]

Konig, E. (1986) Conditionals, concessive conditionals and concessives: Areas of contrast, overlap and neutralization. In: *On conditionals*, ed. E. C. Traugott, A. T. Meulen, J. S. Reilly & C. A. Ferguson. Cambridge University Press. [SJH]

Landman, J. (1987) Regret and elation following action and inaction: Affective responses to positive versus negative outcomes. *Personality and Social Psychology Bulletin* 13:524–36. [aRMJB]

Leslie, A. M. (1987) Pretense and representation: The origins of "theory of mind." *Psychological Review* 94:412–26. [JVR]

Lewin, K. (1935) *A dynamic theory of personality*. McGraw-Hill. [KE]

Lewis, D. (1973) *Counterfactuals*. Blackwell. [aRMJB, AS]

Lillard, A. S. (2007) Pretend play in toddlers. In: *Socioemotional Development in the Toddler Years: Transitions and Transformations*, ed. C. A. Brownell & C. B. Kopp, pp. 149–76. Guilford. [JVR]

Ma, L. (2007) Reasoning with make-believe: Young children's object categorization based on pretend identities. Unpublished doctoral dissertation, University of Virginia, Charlottesville. [JVR]

Mandel, D. R. (2003a) Counterfactuals, emotions, and context. *Cognition and Emotion* 17:139–59. [aRMJB]

(2003b) Effect of counterfactual and factual thinking on causal judgments. *Thinking and Reasoning* 9:245–65. [DRM]

(2003c) Judgment dissociation theory: An analysis of differences in causal, counterfactual, and covariational reasoning. *Journal of Experimental Psychology: General* 137:419–34. [DRM]

(2005) Counterfactual and causal explanation: From early theoretical views to new frontiers. In: *The psychology of counterfactual thinking*, ed. D. R. Mandel, D. J. Hilton & P. Catellani, pp. 11–23. Routledge. [DRM]

Mandel, D. R. & Lehman, D. R. (1996) Counterfactual thinking and ascriptions of cause and preventability. *Journal of Personality and Social Psychology* 71:450–63. [aRMJB, DRM, BAS]

(1998) Integration of contingency information in judgments of cause, covariation, and probability. *Journal of Experimental Psychology: General* 127:269–85. [DRM]

Marcus, S. & Rips, L. J. (1979) Conditional reasoning. *Journal of Verbal Learning and Verbal Behaviour* 18:199–223. [aRMJB]

Markman, K. D. & McMullen, M. N. (2003) A reflection and evaluation model of comparative thinking. *Personality and Social Psychology Review* 7:244–67. [KDM]

(2005) Reflective and evaluative modes of mental simulation. In: *The psychology of counterfactual thinking*, ed. D. R. Mandel, D. J. Hilton & P. Catellani, pp. 77–93. Routledge. [KDM]

Markman, K. D., McMullen, M. N., Elizaga, R. A. & Mizoguchi, N. (2006) Counterfactual thinking and regulatory fit. *Judgment and Decision Making* 1 98–107. [KE]

Markman, K. D. & Miller, A. K. (2006) Depression, control and counterfactual thinking: Functional for whom? *Journal of Social and Clinical Psychology* 25:210–27. [KE]

Markovits, H. & Vachon, R. (1990) Conditional reasoning, representation and level of abstraction. *Developmental Psychology* 26:942 –51. [HM]

McCloy, R. & Byrne, R. M. J. (2000) Counterfactual thinking about controllable events. *Memory and Cognition* 28:1071–78. [aRMJB]

(2002) Semifactual "Even if" thinking. *Thinking and Reasoning* 8:41–67. [aRMJB]

McEleney, A. & Byrne, R. (2000) Counterfactual thinking and causal explanation. In: *Mental models in reasoning*, ed. J. A. Garcia-Madruga, N. Carriedo & M. J. Gonzalez-Labra, pp. 301–14. UNED (Universidad de Nacional de Education a Distancia). [BAS]

(2006) Spontaneous counterfactual thoughts and causal explanations. *Thinking and Reasoning* 12(2):235–55. [arRMJB]

McGill, A. L. (1989) Context effects in judgments of causation. *Journal of Personality and Social Psychology* 57:189–200. [BAS]

McMullen, M. N. (1997) Affective contrast and assimilation in counterfactual thinking. *Journal of Experimental Social Psychology* 33:77–100. [KDM]

McMullen, M. N. & Markman, K. D. (2000) Downward counterfactuals and motivation: The wake-up call and the Pangloss effect. *Personality and Social Psychology Bulletin* 26:575–84. [aRMJB, KDM]

(2002) Affective impact of close counterfactuals: Implications of possible futures for possible pasts. *Journal of Experimental Social Psychology* 38:64–70. [aRMJB]

Michelson, A. A. & Morley, E. W. (1887) On the relative motion of the Earth and the luminiferous aether, *Philosophy Magazine* S.5, 24(151):449–63. Available at: http://en.wikipedia.org/wiki/Michelson-Morley_experiment) [HM]

Mill, J. S. (1872/1956) *A system of logic, ratiocinative and inductive*, 8th edition. Longmans, Green, & Reader. [aRMJB]

Miller, D. T. & Gunasegaram, S. (1990) Temporal order and the perceived mutability of events: Implications for blame assignment. *Journal of Personality and Social Psychology* 59:1111–18. [aRMJB]

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. [aRMJB]

Miller, D. T., Turnbull, W. & McFarland, C. (1990) Counterfactual thinking and social perception: Thinking about what might have been. In: *Advances in experimental social psychology, vol. 22*, ed. P. Zanna, pp. 305–31. Academic Press. [aRMJB]

Miller, G., Galanter, E. & Pribram, K. (1960) *Plans and the structure of behavior*. Holt, Rinehart, & Winston. [RJS]

Miller, G. & Johnson-Laird, P. N. (1976) *Language and perception*. Cambridge University Press. [aRMJB]

Moreno-Ríos, S., García-Madruga, J. A. & Byrne, R. M. J. (in press) Interferences from semi-factual "even if" conditionals. *Acta Psychologica*. [SJH]

Moss, J. & Case, R. (1999) Developing children's understanding of the rational numbers: A new model and an experimental curriculum. *Journal for Research in Mathematics Education* 30(2):122–47. [AS]

Murphy, G. & Wisniewski, E. (2006) Familiarity and plausibility in conceptual combination: Reply to Gagné and Spalding (2006) *Journal of Experimental Psychology: Learning, Memory, and Cognition* 32:1438–42. [TBW]

Murray, M. A. & Byrne, R. M. J. (2007) Single step and multiple step insight problems. Unpublished manuscript. [aRMJB]

Nasco, S. A. & Marsh, K. L. (1999) Gaining control through counterfactual thinking. *Personality and Social Psychology Bulletin* 25:556–68. [KE]

Nersessian, N. J. (2002) The cognitive basis of model-based reasoning in science. In: *The cognitive basis of science*, ed. P. Carruthers, S. Stich & M. Siegal. Cambridge University Press. [SC]

(in press) *Creating scientific concepts*. MIT Press. [SC]

Newell, A. & Simon, H. (1972) *Human problem solving*. Prentice-Hall. [RJS]

Oakhill, J. & Johnson-Laird, P. N. (1985) Rationality, memory, and the search for counterexamples. *Cognition* 20:79–84. [aRMJB]

Oettingen, G. (1996) Positive fantasy and motivation. In: *The psychology of action: Linking cognition and motivation to action*, ed. P. M. Gollwitzer & J. A. Bargh, pp. 236–59. Guilford. [KDM]

Ortony, A., Clore, G. L. & Collins, A. (1988) *The cognitive structure of emotions*. Cambridge University Press. [LJS]

Over, D. E., Hadjchristidis, C., Evans, J. St. B. T., Handley, S. E. & Sloman, S. (2007) The probability of causal conditionals. *Cognitive Psychology* 54:62–97. [SJH]

Pearl, J. (2000) *Causality: Models, reasoning, and inference*. Cambridge University Press. [WC]

Pereda, A., Garavan, H. & Byrne, R. M. J. (submitted) The costs of switching attention on conditional inferences. [rRMJB]

Perner, J. (1991) *Understanding the representational mind*. MIT Press. [JVR]

Piaget, J. (1981) *Le possible et le nécessaire I: l'Évolution du nécessaire chez l'enfant*. Presses Universitaires de France. [HM]

Pighin, S., Byrne, R. M. J., Gonzalez, M., Ferrante, D. & Girotto, V. (submitted) Role effects in counterfactual thinking. [rRMJB]

Pollock, J. L. (1986) *Subjunctive reasoning*. Reidel. [aRMJB]

Quelhas, A. C. & Byrne, R. M. J. (2003) Reasoning with deontic and counterfactual conditionals. *Thinking and Reasoning* 9:43–66. [aRMJB]

Quine, W. V. (1972) *Methods of logic*, 3rd edition. Holt, Rinehart & Winston. [aRMJB]

Resche, C. (2004) Investigating "Greenspanese": From hedging to "fuzzy transparency." *Discourse and Society* 15:723–44. [PL]

Richards, C. A. & Sanderson, J. A. (1999) The role of imagination in facilitating deductive reasoning in 2-, 3-, and 4-year-olds. *Cognition* 72:B1–B9. [JVR]

Riggs, K. J. & Peterson, D. M. (2000) Counterfactual thinking in preschool children: Mental state and causal inferences. In: *Children's reasoning and the mind*, ed. P. Mitchell & K. J. Riggs. Psychology Press. [aRMJB]

Riggs, K. J., Peterson, D. M., Robinson, E. J. & Mitchell, P. (1998) Are errors in false belief tasks symptomatic of a broader difficulty with counterfactuality? *Cognitive Development* 13:73–90. [aRMJB, KJR]

Rips, L. J. (1994) *The psychology of proof*. MIT Press. [aRMJB]

Ritov, I. & Baron, J. (1990) Reluctance to vaccinate: Omission bias and ambiguity. *Journal of Behavioral Decision Making* 3:263–77. [aRMJB]

(1992) Status-quo and omission biases. *Journal of Risk and Uncertainties* 5:49–61. [aRMJB]

Roese, N. J. (1994) The functional basis of counterfactual thinking. *Journal of Personality and Social Psychology* 66:805–18. [aRMJB, KE]

(1997) Counterfactual thinking. *Psychological Bulletin* 121:133–48. [aRMJB, KE]

(2005) *If only*. Broadway Books. [KE]

Roese, N. J. & Hur, T. (1997) Affective determinants in counterfactual thinking. *Social Cognition* 15:274–90. [KE]

Roese, N. J., Hur, T. & Pennington, G. L. (1999) Counterfactual thinking and regulatory focus: Implications for action versus inaction and sufficiency versus necessity. *Journal of Personality and Social Psychology* 77:1109–20. [KE]
Roese, N. J. & Olson, J. M. (1997) Counterfactual thinking: The intersection of affect and function. In: *Advances in experimental social psychology*, vol. 29, ed. M. P. Zanna, pp. 1–59. Academic Press. [KE]
Roese, N. J. & Olson, J. M., eds. (1995) *What might have been: The social psychology of counterfactual thinking*. Erlbaum. [aRMJB]
Roese, N. J., Park, S., Smallman, R. & Gibson, C. (in press) Schizophrenia involves impairment in the activation of intentions by counterfactual thinking. *Schizophrenia Research*. [KE]
Roese, N. J., Sanna, L. J. & Galinsky, A. D. (2005) The mechanics of imagination: Automaticity and counterfactual thinking. In: *The new unconscious*, ed. R. Hassin, J. Uleman & J. A. Bargh, pp. 138–70. Oxford University Press. [aRMJB, KDM]
Rothenberg, A. (1979) *The emerging goddess*. University of Chicago Press. [TBW]
Sanford, D. H. (1989) *If P then Q: Conditionals and the foundations of reasoning*. Routledge. [SJH]
Sanna, L. J. & Schwarz, N. (2006) Metacognitive experiences and human judgment: The case of hindsight bias and its debiasing. *Current Directions in Psychological Science* 15:172–76. [LJS]
(2007) Metacognitive experiences and hindsight bias: It's not just the thought (content) that counts! *Social Cognition* 25:185–202. [LJS]
Sanna, L. J., Schwarz, N. & Small, E. M. (2002a) Accessibility experiences and the hindsight bias: I knew it all along versus it could never have happened. *Memory and Cognition* 30:1288–96. [LJS]
Sanna, L. J., Schwarz, N. & Stocker, S. L. (2002b) When debiasing backfires: Accessible content and accessibility experiences in debiasing hindsight. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28:497–502. [aRMJB, LJS]
Sanna, L. J., Turley-Ames, K. J. & Meier, S. (1999) Mood, self-esteem, and simulated alternatives: Thought-provoking affective influences on counterfactual direction. *Journal of Personality and Social Psychology* 76:543–58. [LJS]
Santamaria, C. & Espino, O. (2002) Conditionals and directionality: On the meaning of if versus only if. *Quarterly Journal of Experimental Psychology* 55A:41–57. [aRMJB]
Santamaria, C., Espino, O. & Byrne, R. M. J. (2005) Counterfactual and semifactual conditionals prime alternative possibilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31:1149–54. [aRMJB]
Schwarz, N., Sanna, L. J., Skurnik, I. & Yoon, C. (2007) Metacognitive experiences and the intricacies of setting people straight: Implications for debiasing and public information campaigns. *Advances in Experimental Social Psychology* 39:127–61. [LJS]
Scott, G., Lonergan, D. & Mumford, M. (2005) Conceptual combination: Alternative knowledge structures, alternative heuristics. *Creativity Research Journal* 17:79–98. [TBW]
Seeleu, E. P., Seeleu S. M., Wells, G. L. & Windschitl, P. D. (1995) Counterfactual constraints. In: *What might have been: The social psychology of counterfactual thinking*, ed. N. J. Roese & J. M. Olson. Erlbaum. [aRMJB]
Sherman, S. J. & McConnell, A. R. (1996) Counterfactual thinking in reasoning. *Applied Cognitive Psychology* 10:113–24. [aRMJB]
Shtulman, A. & Carey, S. (2007) Improbable or impossible? How children reason about the possibility of extraordinary claims. *Child Development* 78(3):1015–32. [AS]
Siegler, R. S. (1996) *Emerging minds: The process of change in children's thinking*. Oxford University Press. [AS]
Simon, H. A. (1956) Rational choice and the structure of the environment. *Psychological Review* 63:129–38. [DRM]
Simpson, J. (1997) *Touching the void*. Vintage. [rRMJB]
Skinner, B. F. (1965) *Science and human behavior*. Free Press. [RJS]
Smith, C., Solomon, G. E. A. & Carey, S. (2005) Never getting to zero: Elementary school students' understanding of the infinite divisibility of number and matter. *Cognitive Psychology* 51:101–40. [AS]
Spellman, B. A., Kincannon, A. & Stose, S. (2005) The relation between counterfactual and causal reasoning. In: *The psychology of counterfactual thinking*, ed. D. R. Mandel, D. J. Hilton & P. Catellani, pp. 28–43. Routledge Research. [BAS]
Spellman, B. A. & Mandel, D. R. (1999) When possibility informs reality: Counterfactual thinking as a cue to causality. *Current Directions in Psychological Science* 8:120–23. [BAS]
Spellman, B. A. & Ndiaye, D. G. (2007) The (dis)similarity between counterfactual and causal judgments: The importance of underlying information, availability, and measurement. Unpublished manuscript. [BAS]
Sperber, D., Cara, F. & Girotto, V. (1995) Relevance theory explains the selection task. *Cognition* 52:3–39. [aRMJB]
Sperber, D. & Girotto, V. (2003) Does the selection task detect cheater detection? In: *New directions in evolutionary psychology*, ed. J. Fitness & K. Sterelny. *Macquarie Monographs in Cognitive Science*. Psychology Press. [aRMJB]
Stalnaker, R. C. (1968) A theory of conditionals. In: *Studies in logical theory*, ed. N. Rescher. Basil Blackwell. [aRMJB]
(2003) *Ways a world might be*. Oxford University Press. [AS]
Stanovich, K. E. (1999) *Who is rational?* Erlbaum. [aRMJB]
(2002) Rationality, intelligence, and levels of analysis in cognitive science: Is dysrationalia possible? In: *Why smart people can be so stupid*, ed. R. J. Sternberg, pp. 124–58. Yale University Press. [RJS]
Sternberg, R. J. (1997) *Thinking styles*. Cambridge University Press. [aRMJB]
Sternberg, R. J. & Lubart, T. I. (1996) Investing in creativity. *American Psychologist* 51:677–88. [TBW]
(1999) The concept of creativity: Prospects and paradigms. In: *Handbook of creativity*, ed. R. J. Sternberg, pp. 3–15. Cambridge University Press. [TBW]
Summerville, A. & Roese, N. J. (in press) Dare to compare: Fact based versus simulation based comparison in daily life. *Journal of Experimental Social Psychology*. [KE]
Surian, L., Caldi, S. & Sperber, D. (2007) Attribution of beliefs by 13-month-old infants. *Psychological Science* 18:580–86. [PL]
Tetlock, P. E. & Lebow, R. N. (2001) Poking counterfactual holes in covering laws: Cognitive styles and historical reasoning. *American Political Science Review* 95:829–43. [aRMJB]
Tetlock, P. E., Lebow, R. N. & Parker, G. (2005) *Unmaking the West: Counterfactuals, contingency, and causation*. University of Michigan Press. [PL]
Tetlock, P. E. & Parker, G. (2005) Counterfactual thought experiments: Why we cannot live without them and how we can learn to live with them. In: *Unmaking the West: Counterfactuals, contingency and causation*, ed. P. E. Tetlock, R. N. Lebow & G. Parker, pp. 3–35. University of Michigan Press. [aRMJB]
Thagard, P. (1984) Conceptual combination and scientific discovery. In: *Philosophy of Science Association (PSA), vol. 1*, ed. P. Asquith & P. Kitcher. Philosophy of Science Association. [TBW]
Thompson, V. A. & Byrne, R. M. J. (2002) Reasoning about things that didn't happen. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28:1154–70. [arRMJB, KDM]
Torrens, D., Thompson, V. A. & Cramer, K. M. (1999) Individual differences and the belief-bias effect: Mental models, logical necessity and abstract reasoning. *Thinking and Reasoning* 5:1–28. [aRMJB]
Ursu, S. & Carter, S. T. (2005) Outcome representations, counterfactual comparisons and the human orbitofrontal cortex: Implications for neuroimaging studies of decision-making. *Cognitive Brain Research* 34:51–60. [KE]
Van Dijk, E. & Zeelenberg, M. (2005) On the psychology of "if only": Regret and the comparison between factual and counterfactual outcomes. *Organizational Behavior and Human Decision Processes* 97:152–60. [MZ]
Venet, M. & Markovits, H. (2001) Understanding uncertainty with abstract conditional premises. *Merrill-Palmer Quarterly* 47(1):74–99. [HM]
Walsh, C. R. & Byrne, R. M. J. (2004) Counterfactual thinking: The temporal order effect. *Memory and Cognition* 32:369–78. [aRMJB]
(2007) The effects of reasons for acting on counterfactual thinking. *Thinking and Reasoning*, 13:461–83. [rRMJB]
Ward, T. B. (1994) Structured imagination: The role of category structure in exemplar generation. *Cognitive Psychology* 27:1–40. [TBW]
(2001) Creative cognition, conceptual combination, and the creative writing of Stephen R. Donaldson. *American Psychologist* 56:350–54. [TBW]
Ward, T. B., Patterson, M. J. & Sifonis, C. (2004) The role of specificity and abstraction in creative idea generation. *Creativity Research Journal* 16:1–9. [aRMJB]
Ward, T. B., Patterson, M. J., Sifonis, C. M., Dodds, R. A. & Saunders, K. N. (2002) The role of graded category structure in imaginative thought. *Memory & Cognition* 30:199–216. [TBW]
Ward, T. B., Smith, S. M. & Finke, R. A. (1999) Creative cognition. In: *Handbook of creativity*, ed. R. J. Sternberg. Cambridge University Press. [TBW]
Wells, G. L. & Gavanski, I. (1989) Mental simulation of causality. *Journal of Personality and Social Psychology* 56:161–69. [aRMJB, BAS]
Werth, L. & Strack, F. (2003) An inferential approach to the knew-it-all-along phenomenon. *Memory* 11:411–19. [LJS]
Westen, D. (2007) *The political brain: The role of emotion in deciding the fate of the nation*. Public Affairs. [RJS]
White, K. & Roese, N. J. (submitted) The social side of counterfactual thinking. *Personality and Social Psychology Bulletin*. [KE]
Wilde, O. (1889/1989), *Complete Works of Oscar Wilde. Essays and Letters. The Decay of Lying*. Harper & Row Publishers. [PL]

Wilkenfeld, M. & Ward, T. (2001) Similarity and emergence in conceptual combination. *Journal of Memory and Language* 45:21–38. [TBW]

Wisniewski, E. (1997) When concepts combine. *Psychonomic Bulletin and Review* 4:167–83. [TBW]

Wisniewski, E. & Love, B. (1998) Relations versus properties in conceptual combination. *Journal of Memory and Language* 38:177–202. [TBW]

Wisniewski, E. J. (1996) Construal and similarity in conceptual combination. *Journal of Memory and Language* 35:434–53. [aRMJB]

Wolff, P. (2007) Representing causation. *Journal of Experimental Psychology: General* 136:82–111. [DRM]

Zeelenberg, M. & Pieters, R. (2006) Feeling is for doing: A pragmatic approach to the study of emotions in economic behavior. In: *Social psychology and economics*, ed. D. de Cremer, M. Zeelenberg & J. K. Murnighan, pp. 117–37. Erlbaum. [MZ]

(2007) A theory of regret regulation 1.0. *Journal of Consumer Psychology* 17:3–18. [KE, MZ]

Zeelenberg, M., van der Bos, K., van Dijk, E. & Pieters, R. (2002) The inaction effect in the psychology of regret. *Journal of Personality and Social Psychology* 82:314–27. [aRMJB]

Zeelenberg, M., Van der Pligt, J. & Manstead, A. S. R. (1998a) Undoing regret on Dutch television. *Personality and Social Psychology Bulletin* 24:1113–19. [aRMJB]

Zeelenberg, M., van Dijk, W. W., Van der Pligt, J., Manstead, A. S. R., Van Empelen, P. & Reinderman, D. (1998b) Emotional reactions to the outcomes of decisions: The role of counterfactual thought in the experience of regret and disappointment. *Organizational Behavior and Human Decision Processes* 75:117–41. [MZ]

Zhang, J., Bonnefon, J.-F. & Deng, C. (2004) Regret and reasons to act or not to act. Unpublished manuscript. Sociology Institute, Shanghai Academy of Social Science. [aRMJB]