

# Counterfactual thinking: The temporal order effect

CLARE R. WALSH

*Brown University, Providence, Rhode Island*

and

RUTH M. J. BYRNE

*Dublin University, Trinity College, Dublin, Ireland*

People often think about how things might have happened differently. Their counterfactual thoughts tend to mentally undo the most recent event in an independent sequence. Consider a game in which two players must each pick the same color card, both red or both black. The first picks black and the second picks red and so they lose. People think, "If only the second player had picked black." Our study tested the idea that the ways in which the players could have won provide counterfactual alternatives to the facts. In three experiments, the same set of facts (both players picked black cards), and the same winning conditions (to win in this new game they must pick different color cards) were presented, but the *description* of the winning conditions varied (e.g., "if one or the other but not both picks a red card" vs. "if one or the other but not both picks a black card"). The results showed that the temporal order effect can be produced or reversed by different descriptions. The descriptions make accessible different elements of the winning possibilities. A theory of the mental representations and cognitive processes underlying counterfactual thinking in the temporal order effect is described.

## Counterfactual Thinking

When people reflect on past events, they tend to think not only about the events that actually happened but also about how those events might have happened differently. For example, if your car breaks down and you are late, you might think that you would have been on time if you had had the car serviced or if you had taken the train. Such counterfactual thinking is pervasive (e.g., Kahneman & Tversky, 1982), and counterfactuals have been studied in philosophy (e.g., Lewis, 1973; Stalnaker, 1968), psychology (e.g., Kahneman & Miller, 1986; Roesch & Olson, 1995), and artificial intelligence (e.g., Costello & McCarthy, 1999; Ginsberg, 1986). Little is known of the mental representations and cognitive processes that underlie the generation of counterfactuals (for a review, see Byrne, 2002), and our goal is to provide such an account and to test it in three experiments.

Counterfactual thinking can help us to learn from past mistakes and to develop intentions for the future (Mark-

man, Gavanski, Sherman, & McMullen, 1993; Roesch, 1994; Sanna, Schwarz, & Stocker, 2002). The sorts of counterfactuals that are useful to people may also be useful to learning algorithms in artificial intelligence systems (Costello & McCarthy, 1999). Counterfactual thinking has also been shown to have an impact on a range of emotions and social judgments, including regret, guilt, and blame, in both laboratory settings (e.g., Landman, 1987; Miller & Turnbull, 1990; Niedenthal, Tangney, & Gavanski, 1994) and naturalistic studies (Davis, Lehman, Wortman, Silver, & Thompson, 1995; Zeelenberg, van der Pligt, & Manstead, 1998).

Psychological studies indicate that there are considerable regularities in the sorts of counterfactuals that people generate most readily, despite the infinite number of ways that past events could have happened differently (e.g., Kahneman & Miller, 1986). People are more likely to undo exceptional than routine events (Kahneman & Tversky, 1982), actions than inactions (e.g., Byrne & McEleney, 2000; Gilovich & Medvec, 1995; N'gbala & Branscombe, 1995), controllable than uncontrollable events (Giroto, Legrenzi, & Rizzo, 1991; Mandel & Lehman, 1996; McCloy & Byrne, 2000), and the first event in a causal chain (Segura, Fernandez-Berrocá, & Byrne, 2002; Wells, Taylor, & Turtle, 1987). In this article, we will focus on one important factor that influences the mutability of an event: its temporal order in relation to other events.

**The temporal order effect.** Counterfactuals that undo historical events such as the rise of the West tend to focus on the "last chance" juncture (Tetlock, in press). Greatest weight is given to a team's most recent performance in a basketball league (Sherman & McConnell, 1996). An individual is judged to be lucky when a good outcome is

---

We thank Phil Johnson-Laird and Mark Keane for their comments on a previous draft, and Orlando Espino, David Mandel, Rachel McCloy, and Alice McEleney for their comments on the experiments. The research has been supported by Enterprise Ireland, the Irish Research Council for the Humanities and Social Sciences, and Dublin University. Some of the results were presented at the International Conference on Thinking in Durham, 2000; the Workshop on Mental Models in Brussels, 2001; the EAESP small-group meeting on Counterfactual Thinking in Aix-en-Provence, France, 2001; the 23rd Conference of the Cognitive Science Society, Edinburgh, 2001; and the 12th Irish Conference on Artificial Intelligence and Cognitive Science Society, Ireland, 2001. Correspondence should be addressed to C. R. Walsh, Cognitive and Linguistic Sciences, Brown University, Box 1978, Providence, RI 02912 (e-mail: clare\_walsh@brown.edu).

described after a bad outcome: A second jump in a ski competition is well rated after a first one that was poorly rated (Teigen, Evensen, & Samoilow, 1999). These effects of temporal order can be clearly observed in the following scenario (from Byrne, Segura, Culhane, Tasso, & Berrocal, 2000, p. 280–281):

Imagine two individuals (John and Michael) who are offered the following very attractive proposition. Each individual is given a shuffled deck of cards, and each one picks a card from their own deck. If the two cards they pick are of the same color (i.e., both red or both black), each individual wins £1,000. Otherwise, neither individual wins anything. John goes first and picks a red card from his deck; Michael goes next and picks a black card from his deck. Thus the outcome is that neither individual wins anything.

When asked to imagine that one of the card selections came out differently, so that they won, participants tend to undo the second event: If only Michael had picked red too, and this finding has been termed the temporal order effect (Miller & Gunasegaram, 1990). Several subsequent studies have confirmed that when a series of events are independent of each other, people tend to mutate the most recent event (Byrne et al., 2000; Spellman, 1997). The effect occurs for sequences of more than two events (Segura et al., 2002). In addition, the second player, Michael, is usually expected to experience more guilt and to be blamed more by John. The mutability of the last event may enhance its perceived responsibility, and this tendency may also underlie everyday preferences—for example, for coaches to place the best player last in a relay race (Miller & Gunasegaram, 1990).

People may generate a counterfactual by selecting a salient fact and mentally undoing it or removing it from the scenario (Kahneman & Tversky, 1982; Legrenzi, Girotto, & Johnson-Laird, 1993; Seelau, Seelau, Wells, & Windschitl, 1995). But in addition, we suggest that people keep in mind not just the facts, but also the counterfactual alternatives (such as the ways in which players could win a game). We report the results of three experiments that provide the first empirical demonstration that people think counterfactually about the same set of facts in different ways depending on the alternatives that have been made available. Previous studies have compared the mutability of different facts to show that more mutable facts have more available alternatives (Kahneman & Miller, 1986); our study shows that a single set of facts differs in mutability depending on the accessibility of alternatives.

**Mental possibilities.** Our explanation of the temporal order effect rests on a small set of simple tenets. The first two assumptions are that people keep in mind true possibilities, and only a few possibilities (Johnson-Laird & Byrne, 2002):

1. The first principle is that people understand the card scenario by keeping in mind the true possibilities:

Facts: John picks RED and Michael picks BLACK and they LOSE

(Byrne et al., 2000), where the crucial elements are in capital letters in the diagram.

2. The second principle is that people do not keep in mind the full set of counterfactual possibilities; that is, the full set of situations that once were possible but now, given the facts, are so no longer, which are as follows:

Facts: John picks RED and Michael picks BLACK and they LOSE  
 Counterfactual: John picks RED and Michael picks RED and they WIN  
 John picks BLACK and Michael picks RED and they WIN  
 John picks BLACK and Michael picks BLACK and they LOSE

People can keep in mind several possibilities to understand a counterfactual conditional (Byrne & Tasso, 1999; Quelhas & Byrne, 2003; Thompson & Byrne, 2002). But the temporal order effect indicates that people think about just a subset of the possible counterfactual models:

Facts: John picks RED and Michael picks BLACK and they LOSE  
 Counterfactual: John picks RED and Michael picks RED and they WIN

3. The third principle is that the subset of counterfactual possibilities that people keep in mind is guided by the winning conditions—that is, the possibilities in which the players would have won. People do not keep in mind the situation in which the players could have lost (the last possibility in the full set above), because it is not an effective counterfactual; that is, it does not undo the outcome (Byrne, 1997). The winning conditions for the game are as follows: “If the two cards they pick are of the same color (i.e., both red or both black), each individual wins £1,000. Otherwise, neither individual wins anything”:

John picks RED and Michael picks RED and they WIN  
 John picks BLACK and Michael picks BLACK and they WIN

The first three principles are representational assumptions. The next two concern the strategies deployed to manipulate these representations:

4. The fourth principle is that people mutate the crucial elements of the facts:

Facts: John picks RED and Michael picks RED and they WIN

to be like the winning conditions:

John picks RED and Michael picks RED and they WIN  
 John picks BLACK and Michael picks BLACK and they WIN

People mutate an event more often if they have an alternative to it in mind—for example, exceptional events bring to mind their corresponding norms (Kahneman & Miller, 1986), and actions bring to mind the way things were before the action occurred (Byrne & McEleney, 2000).

5. The fifth principle is that the first element in the facts is an “anchor” that is presupposed and remains relatively immutable (Byrne, et al., 2000). We have developed a computational model of the temporal order effect that generates a model of the facts and a set of models of

the winning conditions and computes the matches between them. To illustrate the fifth principle, we will give a brief synopsis of how the program works (for a full description and a listing, see Walsh, 2001). The program selects the first element in the model of the facts—that is, John picks RED. It finds a match for it in the first possibility of the winning conditions:

John picks RED and Michael picks RED and they WIN

Because it readily finds a match for the first element of the facts, it mutates the second element of the facts (Michael picks BLACK) to be like the second element of this winning condition, and it concludes that if Michael had picked red, they would have won. The program illustrates the fifth principle: because John is the first player mentioned in the facts, his selection is held constant (Byrne et al., 2000). Accordingly, the program alters the second player's selection to fit with the winning conditions (Walsh & Byrne, 2001). The mutability of an event, such as the first player's selection, is reduced if it is presupposed (Miller & Gunasegaram, 1990), acting as a background against which later events are perceived (Sherman & McConnell, 1996), and playing an important contextualizing role (Byrne et al., 2000), as well as when there are prior events known to have caused it (Giroto et al., 1991; Wells et al., 1987).

This set of simple principles explains why people mutate the second event in the standard card selection scenario. We test this account in three experiments.

**Temporal order effect reversal.** To test our account we consider what people will mutate when they cannot find a match for the first element in the facts (i.e., John picks RED) in the winning conditions—for example, when the winning conditions contain only possibilities in which John picks BLACK. Our account makes a novel prediction in this case: The temporal order effect should be reversed. If no match is found for the first element of the facts, then people should mutate this first element to match the winning possibility; that is, they should say, if only John picked BLACK. We tested this prediction in the three experiments we report.

Our experimental manipulations rest on a final representational assumption (Johnson-Laird & Byrne, 1991, 2002):

6. People think about some elements of the true possibilities explicitly—the elements mentioned in the assertion—and they leave other elements implicit. For example, “John or Michael but not both pick black cards” corresponds to two true possibilities:

John picks BLACK and Michael picks RED  
John picks RED and Michael picks BLACK

People think explicitly about only some of the elements of these possibilities:

John picks BLACK  
Michael picks BLACK

In the first possibility, John picks BLACK is represented explicitly; Michael does not pick black, he picks red, but

that information remains implicit (see Johnson-Laird & Byrne, 2002).

Consider the same disjunction, but described somewhat differently: “John or Michael but not both picks red cards.” It is consistent with the same full set of possibilities as the “black” disjunction:

John picks BLACK and Michael picks RED  
John picks RED and Michael picks BLACK

but people will think about it differently because they represent just some information explicitly—the information corresponding to what is mentioned in the assertion:

Michael picks RED  
John picks RED

The exact same winning conditions, when they are described differently, lead people to keep in mind different elements of the possibilities.

This difference in the representation of the winning conditions in turn leads to a difference in the generation of a counterfactual, as our computer program illustrates. We gave it the facts:

Facts: John picks BLACK and Michael picks  
BLACK and they LOSE

and the exact same winning conditions, but this time described in terms of red cards: “If John or Michael but not both picks a red card, they each win £1000.” It constructs the initial possibilities:

John picks RED  
Michael picks RED and they WIN  
and they WIN

Again it selects the first fact, “John picked BLACK,” and searches for a match in the winning conditions, but this time it does not find a match. Instead, it must turn to its next tactic and look for a match to the alternative to the fact: John picks RED. It finds a match, and fleshes out the possibility to be more explicit:

John picks RED and Michael picks BLACK and they WIN

and concludes, if only John had picked red. The program produces a reversal of the temporal order effect.

The description of the winning conditions determines how accessible certain alternatives are. We rely on this difference in our three experiments. Our account predicts that people given the “black” disjunction should exhibit the standard temporal order effect, and think “if only Michael had picked black”; and it predicts that given the “red” disjunction, they should exhibit a reversal of the temporal order effect, and think “if only John had picked red.” A reversal of the temporal order effect has never previously been observed, and our aim was to test this novel prediction.

Our experiments were carried out using the color card scenario. In each of the experiments, the facts of the players' selections remained the same: John goes first and selects a black card, Michael goes second and also picks a black card, and the outcome is that both players lose. The winning conditions were also the same in each

of the three experiments: both players must pick different cards to win (in each of the four conditions in Experiments 1 and 2), or they must pick different cards or both must pick red to win (in the two conditions in Experiment 3).<sup>1</sup> We held constant the facts and the winning conditions but we varied the *description* of winning conditions in order to vary their representation and hence their accessibility from the facts. These descriptions are outlined in Table 1, together with the sorts of possibilities that we propose people think about to understand them.

**EXPERIMENT 1  
Temporal Order Effect Reversal**

Our aim in the first experiment was to demonstrate that the temporal order effect can be reversed when the facts do not readily match the winning conditions. We gave participants these winning conditions: “If one or the other but not both picks a card from a red suit, each individual wins £1,000.” We expect they will understand the winning conditions by thinking about red cards:

Michael picks RED                      and they WIN  
John picks RED                                and they WIN

We gave them the facts, “John picked black and Michael picked black and they both lost.”

The information about the first element of the facts—“John picked black”—does not readily match their initial understanding of the winning conditions, and so we expect they will change the first element and say, “If only John had picked red . . .”

We gave the experimental group of participants the red disjunction description of the winning conditions. We gave the control group a conjunction: “If the two cards they pick are of a different color (i.e., one from a black suit and one from a red suit), each individual wins £1,000.” The conjunction refers to exactly the same winning conditions as the disjunction:

John picks BLACK and Michael picks RED and they WIN  
John picks RED and Michael picks BLACK and they WIN

If reasoners keep both these models in mind, then they will exhibit the standard temporal order effect. Some reasoners may keep the gist of the conjunction in mind by thinking about just the first possibility or just the second possibility. In that case, the temporal order effect will be eliminated: Reasoners who think about the first possibility only will mutate the second event, and vice versa. Our primary interest is in the experimental group; we expect a reversal of the temporal order effect for the red disjunction; that is, participants will undo the first event.

**Method**

**Materials and Design.** We constructed a scenario based on the color card scenario (from Byrne et al., 2000). In our scenarios, each player won £1,000 if the players picked different rather than the same color cards. We compared a conjunctive description: “If the two cards they pick are of a different color (i.e., one from a black suit and one from a red suit), each individual wins £1,000” with a disjunctive description: “If one or the other but not both picks a card from a red suit, each individual wins £1,000.” The two conditionals describe the same state of affairs: The players could win if John picked black and Michael red, or vice versa. The facts of the outcome were the same in both scenarios: each player picked a black card and so they did not win the £1,000 (the full set of scenarios used in the three experiments is reported in the Appendix).

Participants completed three tasks—a counterfactual mutation task and judgments of guilt and blame, as follows:

1. Please complete the following sentence. John and Michael could each have won £1,000 if only one of them had picked a different card, for instance if . . .
2. Who would you predict would experience more guilt: John or Michael?
3. Who will blame the other more: John or Michael?

They completed the tasks in the fixed order given above, on the answer sheet provided. Participants were assigned to one of the two conditions, in a between-participants design.

**Participants and Procedure.** The participants were 148 undergraduate students from different departments of the University of Dublin, Trinity College, who took part in the experiment voluntarily. There were 70 women and 77 men; 1 participant did not state his/her gender. Their ages ranged from 16 to 37 years with a mean age of 18. Two participants were eliminated from the conjunction condition and 3 from the disjunction condition because they failed to follow the instructions or they failed to complete all of the questions. The remaining participants were assigned randomly to the

**Table 1**  
**Different Descriptions of the Winning Conditions Used in the Experiments, and the Initial Set of Possibilities That Represent Them**

1. Red disjunction: Predict temporal order effect reversed (Experiment 1)		
If one or the other but not both pick a card from a red suit, each individual wins £1,000.		
John RED		WIN
	Michael RED	WIN
2. Black disjunction: Predict temporal order effect observed (Experiment 2)		
If one or the other but not both picks a card from a black suit, each individual wins £1,000.		
John BLACK		WIN
	Michael BLACK	WIN
3. Inclusive red disjunction: Predict temporal order effect reversed (Experiment 3)		
If one or the other or both pick a card from a red suit, then each individual wins £1,000.		
John RED		WIN
	Michael RED	WIN
John RED	Michael RED	WIN

Note—In each of the three experiments, the facts were the same: John picked black and Michael picked black and so they lost.

disjunction (one but not both red) condition ( $n = 96$ ) or to the conjunction (one black and one red) condition ( $n = 47$ ). The greater number of participants in the disjunction condition reflects our primary interest in it.

We tested participants in several large groups. They were given a three-page booklet. The first page contained the instructions in which participants were asked to read the scenario carefully and to complete the questions in the order presented, and they were asked not to change an answer once they had written it. The second page contained one of the two versions of the scenario and the three questions, and the final page contained a debriefing paragraph.

**Results and Discussion**

The red disjunction reversed the temporal order effect. The results show that participants who mutated a single event (64%) exhibited the reverse of the standard temporal order effect; that is, more participants mutated the first event (40%) than the second event (24%), and this difference is reliable (binomial  $n = 61, z = 1.79, p < .04$ ). Both events were mutated by 25% of participants.

The conjunction of different cards eliminated the temporal order effect. Given the conjunction (one black and one red), as many participants mutated the first event (32%) as the second event (36%, binomial  $n = 32, z = .18, p = .86$ ). As Table 2 shows, the conjunction and disjunction conditions did not differ reliably for those participants who mutated the first or second event only [ $\chi^2(1, N = 93) = 2.04, p = .15$ ].

**Guilt and blame.** The standard temporal order effect occurred in both conditions for judgments of guilt and blame. Of those participants who judged that one of the individuals would experience more guilt, more participants expected the second player to experience guilt than the first player when they were given a disjunction (70% vs. 10%, binomial  $n = 77, z = 6.38, p < .0001$ ) and a conjunction (72% vs. 0%, binomial  $n = 34, z = 5.66, p < .0001$ ). Participants were no more likely to expect the second player to experience guilt in the conjunction than in the disjunction (72% vs. 70%) condition, and they expected the first player to experience guilt more in the disjunction than in the conjunction (10% vs. 0%) [ $\chi^2(1, N = 111) = 4.85, p < .03$ ] condition.

A similar pattern emerged for judgments of blame. Those participants who expected that one individual would blame the other more tended to believe that the first would blame the second more given a disjunction (62% vs. 10%, binomial  $n = 69, z = 5.78, p < .0001$ ) and a conjunction (66% vs. 6%, binomial  $n = 34, z = 4.63, p < .0001$ ). The conjunction and disjunction conditions did not differ reliably [ $\chi^2(1, N = 103) = .66, p = .42$ ].

The experiment provides the first demonstration that the typical temporal order effect can be reversed; that is, participants mutated the first event in the sequence, rather than the second event. The reversal does not depend on the factual plays of the contestants: In both scenarios, the players picked black cards. Nor does it depend on the nature of the winning conditions: In both conditions, the players would have won if the first had picked red and the second black, or vice versa. The reversal depends on the *description* of the winning conditions.

**Table 2**  
**Percentages of Mutations and Judgments of Guilt and Blame in the Three Experiments**

	Experiment 1		Experiment 2		Experiment 3	
	Disj (r) $n = 96$	Conj (b & r) $n = 47$	Disj (b) $n = 97$	Conj (r & b) $n = 50$	Disj (r) $n = 85$	Conj (b & r) $n = 69$
<b>Mutations</b>						
First only	40	32	25	34	33	33
Second only	24	36	38	38	17	41
Both	25	17	21	20	27	16
Neither	11	15	16	8	23	10
<b>Guilt</b>						
First	10	0	4	4	17	12
Second	70	72	76	76	55	77
Neither	20	28	20	20	28	12
<b>Blame</b>						
First	62	66	68	74	59	68
Second	10	6	6	6	14	10
Neither	28	28	26	20	27	22

Note—r, red; b, black.

The results corroborate our suggestion that people generate a counterfactual by keeping in mind not only the facts but also the winning conditions. They also corroborate our suggestion that people understand the winning conditions by keeping just some elements of the possibilities in mind, in this case, the individuals winning by picking red cards. When they are told that the players will win if one or the other but not both picks a card from a red suit, they think about the possibility that Michael picks a red card and they win, and the possibility that John picks a red card and they win.

The experiment shows a dissociation between mental mutations and judgments of guilt and blame: Regardless of the mutability of the first or second event, people judge that the second individual will experience greater guilt and that he will be blamed more. Dissociations between judgments of emotions and social ascriptions on the one hand and mutations on the other have been observed increasingly in recent research (e.g., Byrne et al., 2000; Roese & Olson, 1995). On the basis of these results, we may conjecture that judgments of guilt and blame appear to be affected by the factual outcome and the conditions under which the players can win, rather than by the nature of the description of the conditions under which they can win.

However, there is one crucial difference between our scenario and previous studies of the temporal order effect. In our scenario the players had to pick *different* color cards to win, whereas in previous studies, the focus was on something that was the same: pick the same color cards, toss the same face coins, pick the same examination questions, and perform to the same standard throughout several baseball games (Byrne et al., 2000; Miller & Gunasegaram, 1990; Segura et al., 2002; Sherman & McConnell, 1996; Spellman, 1997). It is possible that our results show simply that the temporal order effect does not occur when the players must pick different cards. The original temporal order effect may be an artifact of the

constraint that both players must choose the same card. In our next experiment, we rule out this possibility.

## EXPERIMENT 2 Temporal Order Effect Observed

Our aim in this experiment was to test whether the temporal order effect can be observed for situations in which the players must pick different cards. We gave the experimental group of participants a disjunction similar to that used in the first experiment: "If one or the other but not both picks a card from a black suit, each individual wins £1,000." In this case, the disjunction refers to the black suit (unlike the earlier disjunction, which referred to the red suit). We expect they will understand the winning conditions by thinking about black cards:

John picks BLACK and they WIN  
Michael picks BLACK and they WIN

We gave them the facts, "John picked black and Michael picked black and they both lost."

The information about the first aspect of the facts—"John picked black"—matches readily to their initial understanding of the winning conditions, and so we expect they will change the second aspect and say, "If only Michael had picked red . . ." The temporal order effect should be observed. We gave the control group a conjunction similar to that used in the first experiment: "If the two cards they pick are of a different color (i.e., one from a red suit and one from a black suit), each individual wins £1,000," but the order of reference to red and black suits was different from that in the first experiment, to control for any unforeseen confounding by that order.

### Method

**Materials and Design.** We used the same scenario as described in the previous experiment, with the same factual outcomes, except that we changed the conditionals. We compared a conditional that contained a disjunction: "If one or the other but not both picks a card from a black suit, each individual wins £1,000" with one that contained a conjunction: "If the two cards they pick are of a different color (i.e., one from a red suit and one from a black suit), each individual wins £1,000." The two conditionals describe the same states of affairs. The facts of the outcome were the same in both scenarios: Each player picked a black card and so the players did not win the £1,000. Participants completed the same three tasks. They were assigned to one of the two conditions, in a between-participants design.

**Participants and Procedure.** The participants were 152 undergraduate students from different departments in the University of Dublin, Trinity College, and their participation was voluntary. There were 82 women and 70 men, and their ages ranged from 17 to 38 years, with a mean age of 19 years. Five participants were eliminated from the disjunction condition prior to analysis because they failed to comply with the task. The remaining participants were assigned randomly to the disjunction (one but not both black) condition ( $n = 97$ ), or the conjunction (one red and one black) condition ( $n = 50$ ). The procedure was the same as in the previous experiment.

### Results and Discussion

The black disjunction produces the typical temporal order effect. The results show that for participants who

mutated a single event (63%), they exhibited the standard temporal order effect when they were given the disjunction; that is, more participants mutated the second event (38%) than the first event (25%), although the difference is somewhat marginal (binomial  $n = 61$ ,  $z = 1.54$ , one-tailed  $p = .06$ ). Twenty-one percent of the participants mutated both. We replicated the finding of the first experiment that the temporal order effect is eliminated when participants were given the conjunction (one red and one black) (38% vs. 34%, binomial  $n = 36$ ,  $z = .17$ ,  $p = .87$ ). As Table 2 shows, the conjunction and disjunction conditions did not differ reliably for those participants who mutated the first or second event only [ $\chi^2(1, N = 97) = .58$ ,  $p = .45$ ].

**Guilt and blame.** Once again, the standard temporal order effect occurred in both conditions for judgments of guilt and blame. Those participants who judged that one of the individuals would experience more guilt expected the second player to experience more guilt than the first when they were given a disjunction (76% vs. 4%, binomial  $n = 78$ ,  $z = 7.81$ ,  $p < .0001$ ) and a conjunction (76% vs. 4%, binomial  $n = 40$ ,  $z = 5.53$ ,  $p < .0001$ ). The conjunction and disjunction conditions did not differ reliably [ $\chi^2(1, N = 118) = .0009$ ,  $p = .98$ ].

A similar pattern emerged for judgments of blame. Those participants who expected that one individual would blame the other more tended to believe that the first would blame the second more given a disjunction (68% vs. 6%, binomial  $n = 72$ ,  $z = 6.95$ ,  $p < .0001$ ) and a conjunction (74% vs. 6%, binomial  $n = 40$ ,  $z = 5.22$ ,  $p < .0001$ ). The conjunction and disjunction conditions did not differ reliably [ $\chi^2(1, N = 112) = .024$ ,  $p = .88$ ].

The results show that the temporal order effect is observed for scenarios in which players must pick different color cards. The results corroborate our suggestion once again that people generate a counterfactual by keeping in mind not only the facts but also the winning conditions. They corroborate our suggestion that people understand the winning conditions by keeping just some elements of the possibilities in mind, in this case, the individuals winning by picking black cards.

In the first experiment, the red disjunction reversed the temporal order effect: Reasoners said, "If only John had picked red." In this second experiment, the black disjunction produced the temporal order effect: Reasoners said, "If only Michael had picked red." The facts were the same in both experiments: Both players picked black. The winning conditions were also the same in both experiments: The players would have won if the first had picked black and the second red, or vice versa:

John picks RED Michael picks BLACK and they WIN  
John picks BLACK Michael picks RED and they WIN

The logical form of the description was the same; it was an exclusive disjunction. The only difference was in the reference to the color of the suit, black or red. This small difference in wording created a large difference in mutation patterns—mutations of the first event versus mutations of the second event. In our third and final experiment,

we replicate and extend the reversal of the temporal order effect to descriptions consistent with more than two possibilities.

### EXPERIMENT 3 Temporal Order Reversals Again

Our aim in the final experiment was to replicate the reversal of the temporal order effect and extend the results to a set of winning conditions that contained more than two possibilities. We gave the experimental group of participants the winning conditions described in the following disjunction: "If one or the other *or both* pick a card from a red suit." This inclusive disjunction, unlike the exclusive disjunctions of the previous experiments, is consistent with three alternative possibilities:

John picks RED	Michael picks BLACK	and they WIN
John picks BLACK	Michael picks RED	and they WIN
John picks RED	Michael picks RED	and they WIN

Once again people are likely to think explicitly only about the red cards, for this red disjunction:

John picks RED		and they WIN
	Michael picks RED	and they WIN
John picks RED	Michael picks RED	and they WIN

The facts remained the same as in the previous experiment: John picked black and Michael picked black and so they both lost. We expect that the temporal order effect will be reversed given the inclusive disjunction just as it was for the exclusive disjunction, in Experiment 1.

We gave the control group of participants a conjunction, consistent with three possibilities, "If both pick a card from a red suit or if the two cards they pick are of different colors (i.e., one from a black suit and one from a red suit) . . ." The description is similar to the conjunctive description used in the previous experiments, except for the additional condition that both can pick a card from a red suit.

### Method

**Materials and Design.** We used the same scenario used in the previous experiments, except that we changed the conditional used. In one version, reasoners were given the inclusive disjunction "If one or the other or both pick a card from a red suit, then each individual wins £1,000," and in the other version they were told, "If both pick a card from a red suit or if the two cards they pick are of different colors (i.e., one from a black suit and one from a red suit), each individual wins £1,000." Once again in both versions, the players both selected a black card. Participants completed the same sentence completion task and questions regarding guilt and blame as in the previous experiments.

**Participants and Procedure.** The participants were 155 undergraduate students from different departments in the University of Dublin, Trinity College, who took part in the experiment voluntarily. There were 95 women and 60 men, and their ages ranged from 17 to 53 years with a mean age of 20. Prior to analysis, 1 participant was eliminated from the disjunctive condition because he failed to complete all three questions. The remaining participants were assigned randomly to the disjunctive (one or both red) condition ( $n = 85$ ) or the conjunctive (both red or one black and one red) condition ( $n = 69$ ). The procedure was the same as in the previous experiments.

### Results and Discussion

The inclusive disjunction reverses the temporal order effect. The results show that participants who mutated a single event (50%) exhibited the reverse of the standard temporal order effect when they were given the disjunction description (one or both red); that is, more participants mutated the first event (33%) than the second event (17%), and this difference is reliable (binomial  $n = 42$ ,  $z = 2.01$ , one-tailed  $p < .03$ ). Twenty-seven percent of the participants mutated both. The temporal order effect was eliminated when participants were given the conjunction (one black and one red or both red); that is, the percentage of participants who mutated the first event (33%) and the second event (41%) did not differ reliably (binomial  $n = 51$ ,  $z = .56$ ,  $p = .58$ ). As Table 2 shows, participants who mutated the first or second event only mutated the second event significantly less often in the disjunction than in the conjunction condition (17% vs. 41%) [ $\chi^2(1, N = 93) = 4.33$ ,  $p < .04$ ].

**Guilt and blame.** Once again, the standard temporal order effect occurred in both conditions for judgments of guilt and blame. Of those participants who judged that one of the individuals would experience more guilt, more participants expected the second player to experience guilt than the first when they were given a disjunction (55% vs. 17%, binomial  $n = 61$ ,  $z = 4.1$ ,  $p < .0001$ ) and a conjunction (77% vs. 12%, binomial  $n = 61$ ,  $z = 5.63$ ,  $p < .0001$ ). The conjunction and disjunction conditions did not differ reliably [ $\chi^2(1, N = 122) = 2.0$ ,  $p = .16$ ].

A similar pattern emerged for judgments of blame. Those participants who expected that one individual would blame the other more tended to believe that the first would blame the second more given a disjunction (59% vs. 14%, binomial  $n = 62$ ,  $z = 4.7$ ,  $p < .0001$ ) and a conjunction (68% vs. 10%, binomial  $n = 54$ ,  $z = 5.31$ ,  $p < .0001$ ). The conjunction and disjunction conditions did not differ reliably [ $\chi^2(1, N = 116) = .86$ ,  $p = .35$ ].

The experiment replicates the reversal of the typical temporal order effect; that is, people mutate the first event rather than the second when the description of the winning conditions refers to one color card (red) and the facts refer to the other (black). It extends this reversal to descriptions based on inclusive disjunctions. The experiment also replicates the elimination of any temporal order effect when the description of the same winning conditions is described in a conjunction. The experiment shows that the reversal of the temporal order effect is a robust phenomenon and occurs not only for winning conditions consisting of two possibilities but also for those consisting of three possibilities.

### GENERAL DISCUSSION

Our account of the temporal order effect rests on the possibilities that people keep in mind. We suggest that six simple tenets underlie counterfactual thinking in the temporal order effect:

1. People understand the scenario by keeping in mind the true possibilities.

2. They do not keep in mind the full set of counterfactual possibilities.

3. The subset of counterfactual possibilities that they keep in mind is guided by the winning conditions—that is, the possibilities in which the players would have won.

4. People mutate aspects of the facts to be like the counterfactual possibilities of the winning conditions.

5. The first element of the facts is an anchor. It is matched to the winning possibilities: If there is a match, the first element is held constant and the second element is changed to match the winning possibility; if there is no match, the first element is changed.

6. People think about some elements of the true possibilities explicitly—for example, the elements mentioned in the assertion, and they leave other elements implicit.

The three experiments corroborated the predictions made by this account. In all three experiments, the participants were given the same facts: Both players picked black cards. They were also given the same conditions under which the players could win or lose; the players could win if the first picked black and the second red, or vice versa. However, we varied the way we *described* the winning conditions. We gave participants a disjunction that referred to red cards: “If one or the other but not both picks a card from a red suit, each individual wins £1,000,” or one that referred to black cards: “If one or the other but not both picks a card from a black suit, each individual wins £1,000.” In the first experiment, the red disjunction reversed the temporal order effect; that is, participants undid the first event most often. In the second experiment, the black disjunction produced the standard temporal order effect; that is, participants undid the second event most often. The third experiment replicated the reversal and generalized it to an inclusive red disjunction that corresponded to three winning possibilities. The three experiments show that the mental representation of the conditions under which the players can win influences the temporal order effect.

The possibilities that people keep in mind when they think about situations other than games may also be influenced by the mental representations they construct of the conditions under which the outcome could have been better. People think “if only . . .” in many everyday situations, particularly after a bad outcome such as a car crash. Suppose you are told about an accident in which a dog ran onto the road as two individuals, John and Bob, were driving in opposite directions toward it. John swerved to avoid the dog, and Bob swerved to avoid it, and as a result their cars collided. You know the facts are that they both swerved, and you are told that they could each have escaped injury if one or the other but not both of them had swerved. The temporal order effect may lead you to wish “if only Bob hadn’t swerved . . .” But suppose instead that you know the facts are that they both swerved but you are told they could each have escaped injury if one or the other of them had continued driving straight on. The temporal order effect may be reversed and you may wish “if only John had driven straight on . . .” Our account suggests that the way in which a scenario is described

can have a dramatic effect on the counterfactual thoughts that people generate (somewhat akin to the effects of “framing” an option as a loss or a gain, on people’s preferences for risk seeking and risk aversion; see Baron, 2000, for a review). The description of the ways in which an event could have turned out differently may guide the sorts of counterfactual thoughts people generate in many everyday situations.

Theories that rely on the representation of the facts cannot account for our results because we kept the facts constant in each of our conditions. Our account of the temporal order effect provides an alternative to the view that people calculate the probability of an outcome before and after each event (Spellman, 1997). The change in probability determines the relative contribution of each event to the outcome, and events that result in a large change in probability are assigned a greater causal role. However, when people are given an explicit alternative to the first event, they mutate it as often as the second event, even when the explicit alternative does not alter the probability calculations (Byrne et al., 2000). The first event is immutable because it is presupposed (Byrne et al., 2000; Miller & Gunasegaram, 1990), but we have shown that this presupposition occurs only when there is an explicit match for the first fact in the representation of alternatives to the fact.

In all of our experiments, a substantial minority of participants did not counterfactually undo one or other of the events, but instead focused on both. It is possible that these participants kept both possibilities in mind explicitly. All three experiments show a dissociation between mental mutations and judgments of guilt and blame: Regardless of the mutability of the first or second event, people judge greater guilt for the second individual and they judge that he will be blamed more. The judgments of guilt and blame appear to be affected by the representation of the facts rather than by the representation of the hypothetical alternatives to the facts. Counterfactual thoughts often amplify emotions such as guilt and regret (e.g., Kahneman & Tversky, 1982; Mandel, 2003; Nienenthal et al., 1994) and social ascriptions such as blame and fault (e.g., Wells & Gavanski, 1989). However, recent studies have shown that there can be considerable dissociation between counterfactual thoughts and emotional and social judgments (Byrne et al., 2000; N’gbala & Branscombe, 1997). Our results suggest that judgments of guilt and blame in the temporal order effect may not always result from the same processes that give rise to counterfactual thoughts.

The present experiments provide the first demonstration that the description of the winning conditions can influence the mutability of facts. Selecting an alternative from several potential counterfactual candidates may depend on general cognitive economy: Once the first fact is known, it is used to reduce the number of possibilities held in mind, attenuating the load on working memory. These experiments show that people can recruit counterfactual alternatives not only from sources such as actual past experiences but also from imagined hypothetical situations

(Roese, Sanna, & Galinsky, in press; Tetlock & Lebow, 2001). The conditions under which a game can be won provide one source of imagined possibilities, and everyday counterfactual thoughts may mine many such seams.

## REFERENCES

- BARON, J. (2000). *Thinking and deciding* (3rd ed.). Cambridge: Cambridge University Press.
- BYRNE, R. M. J. (1997). Cognitive processes in counterfactual thinking about what might have been. In D. Medin (Ed.), *The psychology of learning and motivation, advances in research and theory* (Vol. 37, pp. 105-154). San Diego: Academic Press.
- BYRNE, R. M. J. (2002). Mental models and counterfactual thoughts about what might have been. *Trends in Cognitive Sciences*, **6**, 426-431.
- BYRNE, R. M. J., & McELENY, A. (2000). Counterfactual thinking about actions and failures to act. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 1318-1331.
- 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 & Cognition*, **28**, 264-281.
- BYRNE, R. M. J., & TASSO, A. (1999). Deductive reasoning with factual, possible, and counterfactual conditionals. *Memory & Cognition*, **27**, 726-740.
- COSTELLO, T., & MCCARTHY, J. (1999). Useful counterfactuals. *Electronic Transactions on the Web*, **3**, 51-76.
- DAVIS, C. J., LEHMAN, D. R., WORTMAN, C. B., SILVER, R. C., & THOMPSON, S. C. (1995). The undoing of traumatic life events. *Personality & Social Psychology Bulletin*, **21**, 109-124.
- GILOVICH, T., & MEDVEC, V. H. (1995). The experience of regret: What, when, and why. *Psychological Review*, **102**, 379-395.
- GINSBERG, M. L. (1986). Counterfactuals. *Artificial Intelligence*, **30**, 35-79.
- GIROTTO, V., LEGRENZI, P., & RIZZO, A. (1991). Event controllability in counterfactual thinking. *Acta Psychologica*, **78**, 111-133.
- JOHNSON-LAIRD, P. N., & BYRNE, R. M. J. (1991). *Deduction*. Hove, U.K.: Erlbaum.
- JOHNSON-LAIRD, P. N., & BYRNE, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review*, **109**, 646-678.
- KAHNEMAN, D., & MILLER, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review*, **93**, 136-153.
- KAHNEMAN, D., & TVERSKY, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgement under uncertainty: Heuristics and biases* (pp. 201-208). New York: Cambridge University Press.
- LANDMAN, J. (1987). Regret and elation following action and inaction: Affective responses to positive versus negative outcomes. *Personality & Social Psychology Bulletin*, **13**, 524-536.
- LEGRENZI, P., GIROTTO, V., & JOHNSON-LAIRD, P. N. (1993). Focussing in reasoning and decision-making. *Cognition*, **49**, 37-66.
- LEWIS, D. (1973). *Counterfactuals*. Oxford: Blackwell.
- MANDEL, D. R. (2003). Counterfactuals, emotions, and context. *Cognition & Emotion*, **17**, 139-159.
- MANDEL, D. R., & LEHMAN, D. R. (1996). Counterfactual thinking and ascriptions of cause and preventability. *Journal of Personality & Social Psychology*, **70**, 450-463.
- MARKMAN, K. D., GAVANSKI, I., SHERMAN, S. J., & McMULLEN, M. N. (1993). The mental simulation of better and worse possible worlds. *Journal of Experimental Social Psychology*, **29**, 87-109.
- McCLOY, R., & BYRNE, R. M. J. (2000). Counterfactual thinking about controllable events. *Memory & Cognition*, **28**, 1071-1078.
- MILLER, D. T., & GUNASEGARAM, S. (1990). Temporal order and the perceived mutability of events: Implications for blame assignment. *Journal of Personality & Social Psychology*, **59**, 1111-1118.
- 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.
- N'GBALA, A., & BRANSCOMBE, N. R. (1995). Mental simulation and causal attribution: When simulating an event does not affect fault assignment. *Journal of Experimental Social Psychology*, **31**, 139-162.
- N'GBALA, A., & BRANSCOMBE, N. R. (1997). When does action elicit more regret than inaction and is counterfactual mutation the mediator of this effect? *Journal of Experimental Social Psychology*, **33**, 324-343.
- NIEDENTHAL, P. M., TANGNEY, J. P., & GAVANSKI, I. (1994). "If only I weren't" versus "If only I hadn't": Distinguishing shame and guilt in counterfactual thinking. *Journal of Personality & Social Psychology*, **67**, 585-595.
- QUELHAS, A. C., & BYRNE, R. M. J. (2003). Reasoning with deontic and counterfactual conditionals. *Thinking & Reasoning*, **9**, 43-65.
- ROESE, N. J. (1994). The functional basis of counterfactual thinking. *Journal of Personality & Social Psychology*, **66**, 805-818.
- ROESE, N. J., & OLSON, J. M. (Eds.) (1995). *What might have been: The social psychology of counterfactual thinking*. Mahwah, NJ: Erlbaum.
- ROESE, N. J., SANNA, L. J., & GALINSKY, A. D. (in press). The mechanics of imagination: Automaticity and counterfactual thinking. In R. R. Hassin, J. S. Uleman, & J. A. Bargh (Eds.), *The new unconscious*. New York: Oxford University Press.
- SANNA, L. J., SCHWARZ, N., & STOCKER, S. L. (2002). When debiasing backfires: Accessible content and accessibility experiences in debiasing hindsight. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 497-502.
- SEELAU, E. P., SEELAU, S. M., WELLS, G. L., & WINDSCHITL, P. D. (1995). Counterfactual constraints. In N. J. Roese & J. M. Olson (Eds.), *What might have been: The social psychology of counterfactual thinking* (pp. 57-79). Mahwah, NJ: Erlbaum.
- SEGURA, S., FERNANDEZ-BERROCAL, P., & BYRNE, R. M. J. (2002). Temporal and causal order effects in thinking about what might have been. *Quarterly Journal of Experimental Psychology*, **55A**, 1295-1305.
- SHERMAN, S. J., & MCCONNELL, A. R. (1996). The role of counterfactual thinking in reasoning. *Applied Cognitive Psychology*, **10**, S113-S124.
- SPELLMAN, B. A. (1997). Crediting causality. *Journal of Experimental Psychology: General*, **126**, 323-348.
- STALNAKER, R. C. (1968). A theory of conditionals. In N. Rescher (Ed.), *Studies in logical theory* (pp. 98-112). Oxford: Blackwell.
- TEIGEN K. H., EVENSEN, P. C., & SAMOILOW, D. K. (1999). Good luck and bad luck: How to tell the difference. *European Journal of Social Psychology*, **29**, 981-1010.
- TETLOCK, P. E. (in press). The logic and psycho-logic of counterfactual thought experiments in the rise of the West debate. In P. E. Tetlock et al. (Eds.), *Unmaking the West: Exploring alternative histories of counterfactual worlds*. Cambridge: Cambridge University Press.
- TETLOCK, P. E., & LEBOW, R. N. (2001). Poking counterfactual holes in covering laws: Cognitive styles and historical reasoning. *American Political Science Review*, **95**, 829-843.
- THOMPSON, V. A., & BYRNE, R. M. J. (2002). Reasoning counterfactually: Making inferences about things that didn't happen. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 1154-1170.
- WALSH, C. R. (2001). *The role of context in counterfactual thinking*. Unpublished doctoral dissertation, University of Dublin, Trinity College.
- WALSH, C. R., & BYRNE, R. M. J. (2001). A computational model of counterfactual thinking: The temporal order effect. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the 23rd Annual Conference of the Cognitive Science Society* (pp. 1078-1083). Mahwah, NJ: Erlbaum.
- WELLS, G. L., & GAVANSKI, I. (1989). Mental simulation of causality. *Journal of Personality & Social Psychology*, **56**, 161-169.
- WELLS, G. L., TAYLOR, B. R., & TURTLE, J. W. (1987). The undoing of scenarios. *Journal of Personality & Social Psychology*, **53**, 421-430.
- ZEELENBERG, M., VAN DER PLIGT, J., & MANSTEAD, A. S. R. (1998). Undoing regret on Dutch television: Apologizing for interpersonal regrets involving actions or inactions. *Personality & Social Psychology Bulletin*, **24**, 1113-1119.

## NOTE

1. The requirement that the players must pick different color cards gave us greater flexibility in varying the winning conditions than the requirement that they must pick the same color cards.

**APPENDIX****The Scenarios Used in the Three Experiments**

---

First paragraph (common to all experiments):

Imagine two individuals (John and Michael) who are offered the following very attractive proposition. Each individual is given a shuffled deck of cards, and each one picks a card from his own deck.

Winning conditions:

**Experiment 1**

Conjunctive (one black and one red)

If the two cards they pick are of different colors (i.e., one from a black suit and one from a red suit), each individual wins £1,000.

Disjunctive (one but not both red)

If one or the other but not both picks a card from a red suit, each individual wins £1,000.

**Experiment 2**

Conjunctive (both red or one black and one red)

If both pick a card from a red suit or if the two cards they pick are of different colors (i.e., one from a black suit and one from a red suit), each individual wins £1,000.

Disjunctive (one or both red)

If one or the other or both pick a card from a red suit, each individual wins £1,000.

**Experiment 3**

Conjunctive (one red and one black)

If the two cards they pick are of different colors (i.e., one from a red suit and one from a black suit), each individual wins £1,000.

Disjunctive (one but not both black)

If one or the other but not both picks a card from a black suit, each individual wins £1,000.

**Last paragraph** (common to all experiments)

Otherwise, neither individual wins anything. John goes first and picks a black card from his deck; Michael goes next and also picks a black card from his deck. Thus the outcome is that neither individual wins anything.

---

(Manuscript received March 12, 2003;  
revision accepted for publication September 28, 2003.)