Spontaneous counterfactual thoughts and causal explanations

Alice McEleney

Northumbria University, Newcastle upon Tyne, UK

Ruth M. J. Byrne

Trinity College, Dublin University, Ireland

We report two experiments to compare counterfactual thoughts about how an outcome could have been different and causal explanations about why the outcome occurred. Experiment 1 showed that people generate counterfactual thoughts more often about controllable than uncontrollable events, whereas they generate causal explanations more often about unexpected than expected events. Counterfactual thoughts focus on specific factors, whereas causal explanations focus on both general and specific factors. Experiment 2 showed that in their spontaneous counterfactual thoughts, people focus on normal events just as often as exceptional events, unlike in directed counterfactual thoughts. The findings are consistent with the suggestion that counterfactual thoughts tend to focus on how a specific unwanted outcome could have been prevented, whereas causal explanations tend to provide more general causal information that enables future understanding, prediction, and intervention in a wide range of situations.

People often engage in counterfactual thinking about how the outcome of an event could have turned out differently. Suppose a person drives home by an unusual route and has a car accident. They may think, "If only I had gone home by my usual route, I wouldn't have been injured." The tendency to compare reality to what might have been is a pervasive and universal characteristic of human thought: it occurs naturally even in very young children (e.g., Harris, 2000) and it has been documented across cultures (e.g., Gilovich, Wang, Regan, & Nishina, 2003). It has been implicated in a range of important psychological processes, including reasoning (e.g., Byrne, 1997, 2005), creativity (e.g., Hofstadter, 1985), social judgements (e.g., Macrae & Milne, 1992), emotions such as guilt and regret (e.g., Byrne & McEleney, 2000), and reactions to traumatic events (e.g., Davis, Lehman, Wortman, Silver, & Thompson, 1995). Most intriguingly, counterfactual thoughts may be related to causal explanations, and our aim in this paper is to examine the relation between counterfactual thoughts about how a past outcome could have been different and causal explanations about why the outcome occurred.

Philosophers have suggested that the concept of causation depends on the contrast between reality and a counterfactual alternative (e.g., Hume, 1739/1978; Lewis, 1973; Mackie, 1974; Mill, 1843/1967) and social psychologists have suggested close links between counterfactual thoughts and causal attribution (e.g., Hilton, 1990; Lipe, 1991; McGill & Klein, 1993). When people think about a counterfactual conditional, such as "If Jim had got up early, then he would have caught the train", they tend to envisage not only the hypothesised counterfactual situation, in which Jim got up early and caught the train (A and B), but also the presupposed factual situation, in which Jim didn't get up early and didn't catch the train (not-A and not-B) (Byrne & Tasso, 1999). The contrast between the two possibilities may in turn suggest that getting up late *caused* Jim to miss the train (e.g., Roese & Olson, 1997). Counterfactual thoughts may be one part of a more complex causal judgement process (e.g., Einhorn & Hogarth, 1986; Hilton, 1990; McGill & Klein, 1993). They may be used as a heuristic for assigning causality (e.g., Kahneman & Tversky, 1982; Wells & Gavanski, 1989). People sometimes engage in counterfactual thinking in order to determine the causes of outcomes; for example, when they consider historical events (e.g., Tetlock & Lebow, 2001). Counterfactual and causal thoughts can even seem indistinguishable (e.g., Mandel, 2003; Roese, 1997).

Evidence for the link between counterfactual thinking and causal explanation comes from experiments that show that the availability of a counterfactual alternative to an outcome affects people's causal ratings of the events that preceded the outcome (Roese & Olson, 1996; McCloy & Byrne, 2002). Consider a couple who had to drive their own car to a

restaurant because they were refused a lift by a taxi driver (Wells & Gavanski, 1989). On the way, they drove off a collapsed bridge and were killed. Participants given a version of the story, in which the taxi made it across the bridge safely, could generate the counterfactual alternative, "If the taxi driver had given them a lift, they wouldn't have died." Participants given another version, in which the taxi also drove off the bridge, could not generate a counterfactual alternative about it. Participants given the counterfactual version rated the taxi driver's decision as more causal of the outcome than those given the second version.

But counterfactual and causal thoughts often focus on different sorts of events; for example, participants given the taxi scenario tend to generate counterfactual alternatives that focus on the taxi driver's decision, but they tend to judge that the cause of the accident was the collapsed bridge (N'gbala & Branscombe, 1995). Likewise, participants given a scenario about a car crash between a victim and a reckless driver tend to generate counterfactual thoughts that focus on the victim's choice of route, whereas they tend to generate causal explanations that focus on the reckless driver (Mandel & Lehman, 1996). One view is that counterfactual thoughts focus on necessary causes, whereas causal thoughts focus on sufficient causes (N'gbala & Branscombe, 1995). Another view is that counterfactual thoughts focus on controllable antecedents, whereas causal thoughts focus on antecedents that covary with an outcome (Mandel & Lehman, 1996). Of course there are different sorts of causal relations (Cheng & Novick, 1991; Miller & Johnson-Laird, 1976). A strong causal relation—for example, "Joe cutting his finger causes it to bleeds"—identifies the cause as necessary and sufficient for the outcome (e.g., Goldvarg & Johnson-Laird, 2001). It is consistent with the possibility "Joe's finger was cut and it bled" and a second possibility "his finger was not cut and it did not bleed". An enabling condition—for example, "turning the ignition key allows the car to start"— identifies the cause as necessary but not sufficient for the outcome. It is consistent with the possibility "the ignition key was turned and the car started" and the possibility "the ignition key was not turned and the car did not start". But it is also consistent with a third possibility, "the ignition key was turned but the car did not start", because the car starting depends on other factors too, for example the battery must not be flat. People may keep in mind different possibilities when they think about strong causes and enabling conditions. Their counterfactual thoughts may focus on enablers because people envisage the same sorts of possibilities for counterfactuals and enabling conditions (Byrne, 2005).

Our aim in this paper is to test the idea that counterfactual thoughts and causal explanations tend to serve different purposes: causal explanations provide general causal information that enables future understanding, prediction, and control in a wide range of situations, whereas counterfactual

thoughts focus on how a specific unwanted outcome could have been prevented.

Researchers in philosophy, psychology, and artificial intelligence have argued that the quality of a causal explanation depends on its usefulness for future situations (e.g., Gopnik, 2000; Keil & Wilson, 2000; Lewis, 1988; Pearl, 2001). The purpose of causal explanation may be to refine interpretive schemata to facilitate future understanding and prediction (Keil & Wilson, 2000). People may generate causal explanations in order to construct "abstract, coherent, defeasible representations of the causal structure of the world" to allow a wide range of predictions and interventions (Gopnik, 2000, p. 300). A causal explanation may need to be based on "durable relationships, transportable across situations" rather than "transitory relationships", in order to be useful in the future (Pearl, 2001, p. 26). A natural byproduct of the "transportability" (and hence familiarity) of the causal relationship used in an explanation is a "sense of "comprehensibility"" (Pearl, 2001, p. 26). Causal inferences are often based on prior knowledge of causal mechanisms (Ahn, Kalish, Medin, & Gelman, 1995). Recent research has shown that adults are more likely to accept a causal explanation when the causal relation seems general, that is, it conforms to a predictable pattern (Lombrozo & Carey, in press). We suggest that unlike causal explanations, counterfactual thoughts are not subject to the constraint of "generality", because they tend to have a different goal from causal explanations; that is, learning how a specific unwanted outcome could have been prevented.

We report two experiments that provide the first direct comparison of spontaneous counterfactual thoughts and causal explanations. Despite extensive examination of the link between counterfactual thoughts and causal explanations, only two sets of empirical studies have directly compared them (Mandel & Lehman, 1996; N'gbala & Branscombe, 1995). Participants generated different counterfactual and causal thoughts about antecedents that were conjectured to differ in their necessity and sufficiency for the outcome (N'gbala & Branscombe, 1995) or their controllability and covariation with the outcome (Mandel & Lehman, 1996). In these earlier experiments participants were *directed* to generate counterfactual thoughts or causal thoughts. Our aim instead was to examine *spontaneous* counterfactual and causal thoughts. Differences in directed counterfactual and causal thoughts may result from task demands: "Asking someone how something could have been different might cue implicit expectations for personal, controllable actions, whereas asking someone about the cause of an event might implicitly suggest a wider causal search, one that embraces a variety of causal candidates" (Roese & Olson, 1997, p. 37). No previous studies have compared *spontaneous* counterfactual and causal thoughts.

We also took care in our experiments to ensure that our scenarios were not confined to highly negative outcomes, such as the car accidents and plane crashes that featured in earlier studies, because they are dramatic and rare and may not reflect how people think about everyday situations. Moreover, they may reflect only the failure of prevention goals (avoiding a negative outcome) rather than promotion goals (bringing about a positive outcome) (Roese, Hur, & Pennington, 1999). In our experiments, participants imagined themselves in a scenario about a common situation concerned with promotion goals, moving to a new town, and attempting to make new friends.

EXPERIMENT 1

The aim of the experiment was to compare the spontaneous generation of counterfactual thoughts and causal explanations. We examined the effects of three variables on spontaneous counterfactual thoughts and causal explanations: controllable versus uncontrollable outcomes, expected versus unexpected outcomes, and good versus bad outcomes.

We expected that *controllable* outcomes would lead to more spontaneous counterfactual thoughts than uncontrollable ones, because only controllable events are preventable. Second, we expected that *unexpected* outcomes would lead to more spontaneous causal thoughts than expected ones (e.g., Kanazawa, 1992; Weiner, 1985). Unexpected outcomes indicate a violation of a causal relation, and that a failure of prediction has occurred (cf. Klauer & Migulla, 1995; Roese & Olson, 1997; Sanna & Turley, 1996). Third, we expected that *bad* outcomes would lead to more spontaneous counterfactual thoughts and causal explanations than good outcomes. Bad outcomes evoke counterfactual thoughts (e.g., Klauer & Migulla, 1995; Roese & Olson, 1997; Sanna & Turley, 1996) and causal thoughts (Bohner, Bless, Schwarz, & Strack, 1988; Weiner, 1985), perhaps because bad outcomes evoke greater cognitive processing in general than good outcomes (e.g., Taylor, 1991).

We also examined the *content* of spontaneous counterfactual thoughts and causal explanations. We distinguished between *general* or stable factors (e.g., laziness, or bad luck, to explain an exam failure) and *specific* or unstable ones (e.g., lack of study for a particular exam). General factors can be *internal* (e.g., an individual's own actions or characteristics), or *external* (e.g., other people's actions or other external events). Specific factors can be *inhibitory* antecedents that bring about a bad outcome or *facilitative* ones that bring about a good outcome. We expected that causal explanations would tend to focus on general factors because they enable future understanding, prediction, and control in a broad range of situations. We expected that when counterfactual thoughts focused on general factors, they would focus more on internal than external ones because they are most

controllable. We expected that counterfactual thoughts would focus on specific factors because they enable prevention of a specific outcome. We expected that counterfactual thoughts would focus more on inhibitory antecedents, because undoing an inhibitory antecedent would prevent a bad outcome.

Method

Participants. The 248 Dublin University undergraduates, who took part in the experiment voluntarily, included 190 women, 57 men, and 1 person who did not record gender. Their average age was 21 years, with a range from 17 to 55 years.

Design. We examined three between-participant variables: controllable versus uncontrollable outcome, unexpected versus expected outcome, and bad versus good outcome. The participants were randomly assigned to eight groups, controllable bad unexpected ($n = 32$), bad expected ($n = 30$), good unexpected ($n = 32$), and good expected ($n = 29$); uncontrollable bad unexpected ($n = 31$), bad expected ($n = 30$), good unexpected ($n = 31$), and good expected ($n = 33$).

Materials and procedure. We constructed eight versions of a scenario about moving to a new town and attempting to make new friends. The outcome was the individual's social situation six weeks after the move. The versions varied in whether the outcome was controllable or uncontrollable, unexpected or expected, and bad or good. Outcome controllability was manipulated by describing each of four antecedent events as the result of either a controllable decision or some uncontrollable factor. Two of the events would inhibit a good outcome (not going to a party, and not joining a club), and two would facilitate a good outcome (attending a staff dinner, and meeting an old friend). We included the four events to provide participants with a range of antecedents they could use either to generate causal explanations for the outcome or to construct counterfactual alternatives. The four antecedents were presented in one of two possible orders (the order in the scenario below and the reverse order), to control for possible order effects. Outcome expectancy was manipulated by describing the individual's expectations about settling in to the new town at the start of the scenario as either expecting it to be easy or hard, and by describing the person's reaction to the outcome at the end of the scenario as either very surprised or not surprised. Outcome valence was manipulated by describing the outcome as either bad (not having made friends and feeling lonely and isolated), or good (having made friends and feeling happy and at home).

An example of the scenario, the controllable, unexpected, bad outcome version, is as follows (the eight scenarios are in Appendix 1):

> You're moving house to start a new job in a different city. The night before you leave, you write down your thoughts about the move in your diary: . . . I've got mixed feelings about moving to a place where I know hardly anyone – my friends and social life are so important to me. But I'm sure it will be easy to settle in to the new town – I've never had any trouble making new friends . . . A lot happens in your first two weeks in the new town. During your first week at work, a staff dinner is held. You decide to go because you want to get to know your colleagues. You enjoy the evening and meet a lot of people. That weekend, your next-door neighbours invite you to a party. Most of the people who live on your road will be there. However, that evening you decide to go to the cinema instead. The next week, you decide to ring an old friend who lives in the town and ask him to show you around. You arrange to go out with him the following evening and he introduces you to a lot of his friends. A few days later, a colleague tells you there's a membership vacancy at her sports club. You think joining would be a good way to meet people, but then you decide to spend the money on a new stereo instead.

> Six weeks after the move, things have turned out nothing like you had expected. You haven't made any real friends in the new town and you feel very lonely and isolated. You are very upset, and very surprised.

The materials and instructions were presented in booklet form with another unrelated experiment. The first page contained the instructions: "The following story describes a real-life experience. Please read it slowly and carefully, and try to imagine yourself in the situation described. Try to experience the events as if they are really happening to you, and imagine what you would think." The second page contained one of the eight versions of the scenario. At the end of the page, participants were instructed, "Before turning the page, please take a moment to consider the events that have occurred and imagine the thoughts you might have about them." The third page contained the information, "That night, while thinking over the events of the past six weeks, you realise that you haven't written in your diary since the night before you moved. In the space below, jot down what you might write in your diary about your current situation and the events leading up to it." Participants could use the rest of the page to write their thoughts. Note that the instructions did *not* direct participants to explain the causes of the outcome or to generate counterfactual alternatives to it.

Coding.   The number and content of counterfactual thoughts and causal explanations about the outcome in each diary entry were recorded. A counterfactual thought was defined as any change to a scenario event that would change the outcome, e.g., "If only I had gone to that party, I would have made friends". A causal explanation was defined as any statement that

attempted to explain why the outcome occurred, e.g., "I haven't made friends because I didn't go to that party". Participants' counterfactual thoughts and causal explanations were also classified by content, based on whether they focused on general factors (either internal or external), or on any of the four specific antecedents in the scenario (the inhibitory ones: not going to a party, and not joining a club; or the facilitative ones: attending a staff dinner, and meeting an old friend). Both of the examples above focus on missing the neighbours' party; that is, a specific inhibitory antecedent. All responses were coded by the first author and by a second rater who was unaware of the hypotheses and experimental conditions. Initial inter-rater agreement for both classifications (number and content) was 90% for counterfactual thoughts and 79% for causal explanations. Discrepancies were resolved by discussion. An example of a counterfactual thought generated by one participant is "if I had chosen to go to the other functions I could have made more friends", and an example of a causal explanation is "I haven't made an effort to get to know people so obviously I can expect to feel isolated" (Appendix 2 provides examples for each of the categories).

Results and discussion

The average length of participants' diaries was 125 words. They contained a total of 197 counterfactual thoughts, a mean of 0.79 per participant, ranging from 0 to 4 per person ($SD = 1.05$), and a total of 402 causal explanations, a mean of 1.62 per participant, ranging from 0 to 6 per person $SD = 1.3$), as shown in Table 1. Overall, participants produced more causal explanations than counterfactual thoughts ($M = 1.62$ vs $M = 0.79$, Wilcoxon test, $z = 6.92$, $n = 248$, $p < .01$). As Table 1 shows, in six of the eight conditions, participants produced significantly more causal explanations than counterfactual thoughts, smallest $z = 2.31$, $p = .02$. The exceptions were the two conditions with outcomes that were bad and controllable: there were no reliable differences in the numbers of counterfactual thoughts and causal explanations, for the unexpected version, $z = 1.49$, $n = 32$, $p = .14$; or the expected one, $z = 0.05$, $n = 30$, $p = .96$.

As we predicted, participants produced more counterfactual thoughts for controllable than uncontrollable outcomes ($M = 1.1$ vs $M = 0.5$, Mann-Whitney $U = 5662.5$, $n = 248$, $p < .01$), whereas they produced an equal frequency of causal explanations for controllable and uncontrollable outcomes ($M = 1.72$ vs $M = 1.6$, $U = 7559$, $n = 248$, $p = .41$). The difference between controllable and uncontrollable outcomes was greater for counterfactual thoughts than causal explanations ($U = 6262.5$, $n = 248$, $p = .01$).

Also as expected, participants produced more causal explanations for unexpected than expected outcomes ($M = 1.79$ vs $M = 1.47$, Mann-Whitney

TABLE 1
Frequencies: Experiment 1

	Controllable	Uncontrollable	Overall
Counterfactual			
Bad Unexpected	56 (1.75)	26 (0.81)	82 (1.3)
Expected	49 (1.63)	19 (0.65)	68 (1.1)
Good Unexpected	13 (0.42)	9 (0.29)	22 (0.3)
Expected	17 (0.57)	8 (0.24)	25 (0.4)
Overall			197 (0.79)
Causal			
Bad Unexpected	70 (2.19)	58 (1.81)	128
Expected	51 (1.70)	56 (1.93)	107
Good Unexpected	45 (1.45)	48 (1.55)	93
Expected	36 (1.20)	38 (1.15)	74
Overall			402 (1.62)

Frequencies (means per participant in parentheses) of counterfactual thoughts and causal explanations for the eight conditions in Experiment 1.
The n in each cell ranges from 29 to 33, see Participants section of text.

$U = 6708.5$, $n = 248$, $p = .04$), whereas they produced an equal frequency of counterfactual thoughts for unexpected and expected outcomes, ($M = 0.76$ vs $M = 0.86$, $U = 7500$, $n = 248$, $p = .36$). However, the difference between unexpected and expected outcomes was not significantly greater for causal explanations than counterfactual thoughts ($U = 7120$, $n = 248$, $p = .31$).

Third, as expected, participants produced more counterfactual thoughts for bad than good outcomes ($M = 1.22$ vs $M = 0.38$, Mann-Whitney $U = 4384.5$, $n = 248$, $p < .01$), and more causal explanations for bad than good outcomes ($M = 1.91$ vs $M = 1.34$, $U = 5924$, $n = 248$, $p < .01$). The difference between bad and good outcomes was no greater for counterfactual thoughts than causal explanations ($U = 7068.5$, $n = 248$, $p = .26$).

As shown in Table 2, the content of participants' counterfactual thoughts and causal explanations focused on *specific inhibitory* antecedents, i.e., missing the neighbours' party, not joining the sports club (32%), *specific facilitative* antecedents, i.e., going to the staff dinner, meeting an old friend (23%), *general internal* factors such as their own characteristics, decisions, and effort (20%), or *general external* factors, such as luck, the availability of opportunities, and other people's actions (17%; see Appendix 2)—8% of counterfactual and causal thoughts could not be classified into any of these categories.

As predicted, causal explanations, more than counterfactual thoughts, focused on general factors (49% vs 14%) rather than specific ones (47% vs 71%), and this difference was reliable, $z = 4.44$, $n = 248$, $p < .01$. Causal explanations focused equally on specific and general factors (47% vs 49%,

TABLE 2
Percentages: Experiment 1

	Counterfactual	Causal	All thoughts
Specific			
Inhibitory	68	14	32
Facilitative	3	33	23
Total	*71*	*47*	*55*
General			
Internal	14	24	20
External	0	26	17
Total	*14*	*49*	*38*
Other	*15*	*4*	*8*

Percentages of counterfactual thoughts and causal explanations in each content category in Experiment 1.

binomial $z = 0.41$, $n = 386$, $p = .68$), but counterfactual thoughts focused more on specific than general factors (71% versus 4%, binomial $z = 8.59$, $n = 167$, $p < .01$).

As predicted, counterfactual thoughts more than causal explanations focused on internal (100% vs 48%) rather than external factors (0% vs 52%) and this difference was reliable, $z = 2.02$, $n = 248$, $p = .04$. Counterfactual thoughts focused more on internal than external factors (100% vs 0%, binomial $z = 4.91$, $n = 28$, $p < .01$), but causal explanations focused equally on internal and external factors (48% vs 52%, binomial $z = 0.64$, $n = 197$, $p = .52$).

Also as predicted, counterfactual thoughts more than causal explanations focused on inhibitory (96% vs 30%) rather than facilitative (4% vs 70%) antecedents, and this difference was reliable, $z = 8.58$, $n = 248$, $p < .01$. Counterfactual thoughts focused more on inhibitory than facilitative antecedents (96% vs 4%, binomial $z = 10.77$, $n = 139$, $p < .01$), whereas causal explanations focused more on facilitative than inhibitory antecedents (70% vs 30%, binomial $z = 5.46$, $n = 189$, $p < .01$).

Overall, the results of the experiment are consistent with the suggestion that causal explanations provide general information to enable future prediction, understanding, and intervention in a wide range of situations, whereas counterfactual thoughts provide specific information to enable future prevention of unwanted outcomes. Spontaneous counterfactual thoughts were evoked more by controllable than uncontrollable outcomes, and they focused on specific antecedents that could inhibit a bad outcome. In contrast, spontaneous causal explanations were evoked more by unexpected than expected outcomes, and they focused on both general and specific factors.

Our finding that participants produced as many counterfactual thoughts for unexpected outcomes as expected ones is surprising in light of earlier research on the exceptionality effect; that is, the tendency to mentally alter exceptional rather than normal events (e.g., Gavanski & Wells, 1989; Kahneman & Tversky, 1982). A possible explanation is that unexpected events evoked causal explanations, and this may have limited the number of counterfactual thoughts produced, as the two sorts of thoughts may sometimes compete in spontaneous generation. However, bad outcomes evoked both counterfactual thoughts and causal explanations, and so unexpected outcomes reasonably could be expected to do the same. We indicated that an outcome was unexpected by describing the protagonist's expectations at the outset: "I'm sure it will be easy to settle in to the new town..." and indicating that the outcome did not match up to these expectations "...things have turned out nothing like you had expected... You are ... very surprised". This manipulation of expectedness is akin to the standard manipulation of exceptionality; that is, with reference to intrapersonal norms. Hence, the unexpected outcomes could be considered to be exceptional and the expected outcomes might be viewed as normal. If so, our result shows that people produced as many spontaneous counterfactual thoughts for exceptional outcomes as normal ones. Our next experiment examines this result further.

EXPERIMENT 2

The aim of our second experiment was to compare spontaneous counterfactual thoughts and causal explanations about exceptional and normal antecedents. When people are *directed* to generate counterfactual thoughts, they focus on exceptional antecedents more than normal ones (e.g., Gavanski & Wells, 1989; Kahneman & Tversky, 1982). The exceptionality effect may arise because an exceptional antecedent makes available an alternative in the guise of its normal counterpart (Kahneman & Miller, 1986). But it remains unknown whether people focus more on exceptional antecedents when they *spontaneously* generate counterfactual thoughts (Roese & Olson, 1996, 1997). The results of our first experiment suggest that spontaneous counterfactual thinking may be unaffected by the exceptionality of events, as participants generated as many counterfactual thoughts about unexpected as expected outcomes.

Method

Participants and design.   The 84 Dublin University and University College Dublin undergraduates, who took part in the experiment voluntarily, included 55 women and 29 men. Their average age was 20

years, with a range from 17 to 45 years. They were randomly assigned to two groups: exceptional antecedents ($n = 42$) or normal antecedents ($n = 42$).

Materials and procedure. We constructed two versions of a scenario based on the ones used in Experiment 1. We based them on the bad, unexpected, uncontrollable outcome scenario because this outcome evoked approximately equal numbers of counterfactual thoughts and causal explanations in Experiment 1. The two versions differed in whether each of the four antecedent events was described as exceptional or normal, as shown in Appendix 1. We adapted the materials from Experiment 1 to emphasise that the events were normal or exceptional. The four antecedents were presented in one of two possible orders (the order shown in Appendix 1 and the reverse order) to control for possible order effects. The instructions and task were identical to those used in Experiment 1.

Coding. The numbers of counterfactual thoughts and causal explanations in each diary entry were recorded using the coding scheme from Experiment 1. All responses were coded by the first author and by a second rater who was unaware of the hypotheses or experimental conditions. Initial inter-rater agreement was 93% for counterfactual thoughts and 85% for causal explanations. Discrepancies were resolved by discussion. An example of a counterfactual thought produced by one participant is "If I'd gone to my neighbours' party and didn't have the migraine I could have made some real friends", and an example of a causal explanation is "Things have not turned out well at all, maybe it's just bad luck".

Results and discussion

As Table 3 shows, a total of 38 counterfactual thoughts were produced, a mean of 0.45 per participant, ranging from 0 to 3 per person. A total of 110 causal explanations were produced, a mean of 1.31 per participant, ranging from 0 to 5 per person. As in Experiment 1, participants produced more causal explanations than counterfactual thoughts overall ($M = 1.31$ vs

TABLE 3
Frequencies: Experiment 2

	Counterfactual	Causal
Unusual antecedents	15 (0.36)	68 (1.62)
Normal antecedents	23 (0.55)	42 (1.00)
Total	38 (0.45)	110 (1.31)

Frequencies (with means per participant in parentheses) of counterfactual thoughts and causal explanations in Experiment 2 ($n = 42$ in each condition).

$M = 0.45$, Wilcoxon test, $z = 4.87$, $n = 84$, $p < .01$), and they did so in both conditions (exceptional: $M = 1.62$ vs $M = 0.36$, $z = 4.38$, $n = 42$, $p < .01$; normal: $M = 1.00$ vs $M = 0.55$, $z = 2.26$, $n = 42$, $p = .02$).

Participants produced an equal frequency of counterfactual thoughts for exceptional and normal antecedents ($M = 0.55$ vs $M = 0.36$, $U = 798$, $n = 84$, $p = .36$), whereas they produced more causal explanations for exceptional than normal antecedents ($M = 1.62$ vs $M = 1.00$, $U = 671$, $n = 84$, $p = .03$).

The results of the experiment show that people generate as many spontaneous counterfactual thoughts about exceptional events as normal ones. Past research shows that people's directed counterfactual thoughts tend to focus more on exceptional antecedents, perhaps because exceptional events bring to mind their normal alternatives. But in spontaneous thought, the generation of counterfactuals may be guided by the goal of prevention, and exceptional events may be as preventable as normal ones. In contrast, spontaneous causal thoughts may be guided by the goal of future understanding, prediction, and intervention, and exceptional antecedents may be better predictors than normal ones.

GENERAL DISCUSSION

Spontaneous counterfactual thoughts and causal explanations have different emphases. People create more counterfactual thoughts about controllable than uncontrollable events, whereas they create causal explanations equally about both. People produce more spontaneous causal explanations about unexpected and exceptional events than expected or normal ones, whereas they produce counterfactual thoughts equally about both. This finding is true whether it is the outcome that is unexpected, as the first experiment showed, or the antecedents that are exceptional, as the second experiment showed. Counterfactual thoughts focus on specific antecedents that could inhibit a bad outcome, whereas causal explanations focus on both general and specific factors, as the first experiment showed. Spontaneous counterfactual and causal explanations do share some similarities: both are produced more about bad outcomes than good outcomes.

The results of the experiments are consistent with the suggestion that causal explanations focus on general causal relations, such as strong causes, that help their future understanding, prediction, and control in a wide range of situations, whereas counterfactual thoughts focus on specific causal relations, such as enabling conditions whose absence may help their future prevention of bad outcomes. Counterfactual thoughts and causal explanations were both evoked by bad rather than good outcomes, and both kinds of thoughts may helpful for avoiding bad outcomes.

Participants' spontaneous causal explanations focused on general factors such as effort and luck as well as specific antecedents, whereas their counterfactual thoughts focused mostly on specific antecedents. General factors may explain and predict future outcomes in a broad range of situations, whereas specific antecedents may be more easily converted into plans for preventing failures (Taylor & Pham, 1996). Causal explanations focused on both internal factors such as effort, and external factors such as luck, whereas counterfactual thoughts focused mostly on internal factors— internal and external factors are equally relevant to general causal understanding, but only internal factors are potentially controllable and therefore directly relevant to prevention. Causal thoughts focused on antecedents that would facilitate a good outcome, whereas counterfactual thoughts focused on antecedents that would inhibit a good outcome—a facilitative antecedent predicts a good outcome, whereas the undoing of an inhibitory antecedent prevents a bad outcome.

Previous studies have shown that people undo exceptional antecedents more than normal ones, at least when they are directed to generate counterfactual alternatives to controllable events (e.g., Kahneman & Miller, 1986; Kahneman & Tversky, 1982). But we found that they spontaneously generate as many counterfactual thoughts about outcomes with exceptional antecedents as normal ones, for uncontrollable events. The finding suggests some limits to the exceptionality effect and is consistent with our suggestion that spontaneous counterfactual thinking enables prevention (and outcomes of exceptional antecedents are no more preventable than outcomes of normal antecedents).

We suggest that counterfactual thoughts and causal explanations both depend on an understanding of the causal relations between events, but they focus on different sorts of causal relations. One view is that causal thoughts focus on relations in which the antecedent is both necessary and sufficient to produce the outcome, and counterfactual thoughts are used to test the necessity of hypothesised causes (e.g., Einhorn & Hogarth, 1986; Hilton, 1990; McGill & Klein, 1993). However, this proposal does not predict the pattern of similarities and differences in the antecedents and content of counterfactual and causal thoughts demonstrated in our experiments. In contrast, our results support the alternative view that causal explanations tend to focus on general causal relations, whereas counterfactual thoughts tend to focus on specific prevention relations. The distinction we have made between causal explanations and preventative counterfactual thoughts is consistent with findings from other research. For example, victims of traumatic events often spontaneously generate counterfactual thoughts about how they could have prevented the event without believing that they caused it (e.g., Davis et al., 1995). Pre-school children spontaneously refer to counterfactual alternatives in response to a question about how an outcome

could have been prevented more than to a question about why an outcome occurred, which suggests that even young children distinguish between counterfactual and causal thoughts (Harris, German, & Mills, 1996).

Our experiments extend the scope of past comparisons of counterfactual and causal thinking, but of course they do not provide a complete comparison of counterfactual and causal thoughts. We have examined just a few specific attributes of counterfactual thoughts and causal explanations produced in response to a particular kind of scenario. Further research is needed to determine whether our findings hold in a range of scenarios and real-life situations. In some situations, the distinction between counterfactual and causal thoughts may not be clear-cut. Moreover, it is clear that counterfactual and causal thoughts may serve functions other than understanding causal relations. For example, people sometimes generate "downward" counterfactual thoughts about how an outcome could have been worse (e.g., "If I hadn't remembered my umbrella, I would have got soaked") and the function of such thoughts seems to be to improve affect, rather than to learn anything about causation or prevention (Roese & Olson, 1997).

Our participants generated about twice as many causal explanations as counterfactual thoughts. One possible reason is that our instructions to write about "your current situation and the events leading up to it" may have implicitly prompted participants to generate causal explanations more than counterfactual thoughts. An alternative possibility is that people find it more difficult to generate counterfactual thoughts than causal explanations. We suggest that counterfactual and causal thoughts depend on the same sorts of underlying mental representations of causal relations. Different causal relations may be represented differently (Byrne, 2005; Goldvarg & Johnson-Laird, 2001; Miller & Johnson-Laird, 1976). A strong causal relation, in which an antecedent is necessary and sufficient for the outcome, such as "heating the butter causes it to melt", is consistent with two possibilities: the butter is heated and it melts, and the butter is not heated and it does not melt. The fully explicit mental representation of an assertion such as "heating the butter caused it to melt" includes a mental notation that the first possibility corresponds to the facts and the second possibility is counterfactual, and so the representation does not correspond simply to a conjunction (Johnson-Laird & Byrne, 1991). But people may construct an economical initial representation that makes explicit just the first possibility: the butter is heated and it melts, and the second possibility remains implicit as an unformed thought (Goldvarg & Johnson-Laird, 2001; Johnson-Laird & Byrne, 1991). The causal explanation, "heating the butter caused it to melt", can be reconstructed from this economical representation. The counterfactual thought "if the butter hadn't been heated, then it wouldn't have melted" may be more readily created when the mental representation is

fleshed out to be explicit, to keep in mind the alternative possibility that the butter is not heated and it does not melt. Thinking about alternative possibilities is difficult: people take longer and they make more errors when an inference requires them to do so (e.g., Johnson-Laird & Byrne, 1991). For this reason, it may be harder for people to generate counterfactual thoughts than causal explanations. Consistent with this idea, developmental evidence suggests that children can think about the causes of an event before they can generate explicit counterfactual alternatives (Robinson & Beck, 1999).

But counterfactual thoughts may render explicit previously implicit knowledge, and so people may learn more from generating counterfactual thoughts than from generating causal explanations (e.g., Kahneman, 1995; Roese, 1994; Tetlock & Belkin, 1996). When people think about controllable, unwanted outcomes, they may be prompted to think about more possibilities than their initial economical mental representation of a causal relation, to generate explicit counterfactual alternatives. We propose that what people typically gain from counterfactual thinking is not the general causal understanding that they gain from causal explanation, but more specific knowledge of how they can intervene to prevent a similar outcome in the future.

REFERENCES

Ahn, W., Kalish, C. W., Medin, D. L., & Gelman, S. A. (1995). The role of covariation versus mechanism information in causal attribution. *Cognition*, *54*(3), 299–352.

Bohner. G., Bless, N., Schwarz, N., & Strack, F. (1988). What triggers causal attributions? The impact of valence and subjective probability. *European Journal of Social Psychology*, *18*, 335–345.

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

Byrne, R. M. J. (2005). *The rational imagination: How people create alternatives to reality*. Cambridge, MA: MIT Press.

Byrne, R. M. J., & McEleney, A. (2000). Counterfactual thinking about actions and failures to act. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *26*(5), 1318–1331.

Byrne, R. M. J., & Tasso, A. (1999). Deductive reasoning from factual, possible and counterfactual conditionals. *Memory and Cognition*, *27*, 726–740.

Cheng, P. W., & Novick, L. R. (1991). Causes versus enabling conditions. *Cognition*, *40*, 83–120.

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

Einhorn, H. J., & Hogarth, R. M. (1986). Judging probable cause. *Psychological Bulletin*, *99*, 3–19.

Gavanski, I., & Wells, G. L. (1989). Counterfactual processing of normal and exceptional events. *Journal of Experimental Social Psychology*, *25*, 314–325.

Gilovich, T., Wang, R. F., Regan, D., & Nishina, S. (2003). Regrets of action and inaction across cultures. *Journal of Cross-Cultural Psychology*, *34*, 61–71.

Goldvarg, E., & Johnson-Laird, P. N. (2001). Naive causality: A mental model theory of causal meaning and reasoning. *Cognitive Science*, *25*(4), 565–610.

Gopnik, A. (2000). Explanation as orgasm and the drive for causal knowledge: The function, evolution, and phenomenology of the theory-formation system. In F. Keil & R. A. Wilson (Eds.), *Explanation and cognition* (pp. 299–324). Cambridge, MA: MIT Press.

Harris, P. L. (2000). *The work of the imagination*. Oxford: Blackwell.

Harris, P. L., German, T., & Mills, P. (1996). Children's use of counterfactual thinking in causal reasoning. *Cognition*, *61*, 233–259.

Hilton, D. J. (1990). Conversational processes and causal explanations. *Psychological Bulletin*, *107*, 65–81.

Hofstadter, D. R. (1985). *Metamagical themas*. New York: Basic Books.

Hume, D. (1978). *A treatise of human nature*. Oxford: Oxford University Press. [Original work published 1739]

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hove, UK: Lawrence Erlbaum Associates Ltd.

Kahneman, D. (1995). Varieties of counterfactual thinking. In N. J. Roese & J. M. Olson (Eds.), *What might have been: The social psychology of counterfactual thinking* (pp. 375–396). Mahwah, NJ: Lawrence Erlbaum Associates Inc.

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.), *Judgment under uncertainty: Heuristics and biases* (pp. 201–208). New York: Cambridge University Press.

Kanazawa, S. (1992). Outcome or expectancy? Antecedents of spontaneous causal attribution. *Personality and Social Psychology Bulletin*, *18*(6), 659–668.

Keil, F., & Wilson, R. A. (2000). Explaining explanation. In F. Keil & R. A. Wilson (Eds.), *Explanation and cognition* (pp. 1–18). Cambridge, MA: MIT Press.

Klauer, K. C., & Migulla, G. (1995). Spontaneous counterfactual processing. *Zeitschrift für Sozialpsychologie*, *26*, 34–42.

Lewis, C. (1988). Why and how to learn why: Analysis-based generalization of procedures. *Cognitive Science*, *12*, 211–256.

Lewis, D. (1973). Causation. *Journal of Philosophy*, *70*, 556–567.

Lipe, M. G. (1991). Counterfactual reasoning as a framework for attribution theories. *Psychological Bulletin*, *108*, 3–18.

Lombrozo, T., & Carey, S. (in press). Functional explanation and the function of explanation. *Cognition*.

Mackie, J. L. (1974). *The cement of the universe: A study of causation*. Oxford: Oxford University Press.

Macrae, C. N., & Milne, A. B. (1992). A curry for your thoughts: Empathic effects on counterfactual thinking. *Personality and Social Psychology Bulletin*, *18*(5), 625–630.

Mandel, D. R. (2003). Effect of counterfactual and factual thinking on causal judgments. *Thinking & Reasoning*, *9*, 245–265.

Mandel, D. R., & Lehman, D. R. (1996). Counterfactual thinking and ascriptions of cause and preventability. *Journal of Personality and Social Psychology*, *71*(3), 450–463.

McCloy, R., & Byrne, R. M. J. (2002). Semifactual ``even if'' thinking. *Thinking & Reasoning*, *8*, 41–67.

McGill, A. L., & Klein, J. G. (1993). Contrastive and counterfactual thinking in causal judgment. *Journal of Personality and Social Psychology*, *64*, 897–905.

Mill, J. S. (1967). *A system of logic raciocentive and inductive*. London: Longmans. [Original work published 1843]

Miller, G. A., & Johnson-Laird, P. N. (1976). *Language and perception*. Cambridge, MA: Harvard University Press.

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.

Pearl, J. (2001). *Causality: Models, reasoning, and inference*. Cambridge: Cambridge University Press.

Robinson, E. J., & Beck, S. (1999). What is difficult about counterfactual reasoning? In P. Mitchell & K. J. Riggs (Eds.), *Children's reasoning and the mind*. Hove, UK: Psychology Press.

Roese, N. J. (1994). The functional basis of counterfactual thinking. *Journal of Personality and Social Psychology*, *66*(5), 805–818.

Roese, N. J. (1997). Counterfactual thinking. *Psychological Bulletin*, *121*, 133–148.

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*, 1109–1120.

Roese, N. J., & Olson, J. M. (1996). Counterfactuals, causal attributions and the hindsight bias. *Journal of Experimental Social Psychology*, *32*, 197–227.

Roese, N. J., & Olson, J. M. (1997). Counterfactual thinking: The intersection of affect and function. In M. P. Zanna (Ed.), *Advances in experimental social psychology*. *Vol. 29* (pp. 1–59). San Diego, CA: Academic Press.

Sanna, L. J., & Turley, K. J. (1996). Antecedents to spontaneous counterfactual thinking: Effects of expectancy violation and outcome valence. *Personality and SocialPsychology Bulletin*, *22*(9), 909–919.

Taylor, S. E. (1991). Asymmetrical effects of positive and negative events – the mobilization minimization hypothesis. *Psychological Bulletin*, *110*(1), 67–85.

Taylor, S. E., & Pham, L. B. (1996). Mental simulation, motivation and action. In P. M. Gollwitzer & J. A. Bargh (Eds.), *The psychology of action: Linking cognition and motivation to behaviour* (pp. 219–235). London: Guilford Press.

Tetlock, P. E., & Belkin, A. (Eds.). (1996). *Counterfactual thought experiments in world politics: Logical, methodological and psychological perspectives*. Princeton, NJ: Princeton 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.

Weiner, B. (1985). ``Spontaneous'' causal thinking. *Psychological Bulletin*, *97*(1), 74–84.

Wells, G. L., & Gavanski, I. (1989). Mental simulation of causality. *Journal of Personality and Social Psychology*, *56*, 161–169.

APPENDIX 1

SCENARIOS USED IN THE EXPERIMENTS

Experiment 1

You're moving house to start a new job in a different city. The night before you leave, you write down your thoughts about the move in your diary: . . . I've got mixed feelings about moving to a place where I know hardly anyone – my friends and social life are *so* important to me.

Unexpected (bad outcome) / Expected (good outcome): *But I'm sure it will be easy to settle in to the new town – I've never had any trouble making new friends.*
Expected (bad outcome) / Unexpected (good outcome): *I'm sure it will be hard to settle in to the new town – I've always had trouble making new friends.*

A lot happens in your first two weeks in the new town. During your first week at work, a staff dinner is held.

Controllable: *You decide to go because you want to get to know your colleagues.*
Uncontrollable: *You have to go because your boss has asked all the staff to be there.*

You enjoy the evening and meet a lot of people.
That weekend, your next-door neighbours invite you to a party. Most of the people who live on your road will be there.

Controllable: *However, that evening you decide to go to the cinema instead.*
Uncontrollable: *However, that evening you're extremely ill with the flu, so you can't go.*

The next week,

Controllable: *you decide to ring an old friend who lives in the town and ask him to show you around. You arrange to go out with him the following evening*
Uncontrollable: *you happen to bump into an old friend who lives in the town and he insists on showing you around. He takes you out the following evening* and he introduces you to a lot of his friends.

A few days later, a colleague tells you there's a membership vacancy at her sports club. You think joining would be a good way to meet people,

Controllable: *but then you decide to spend the money on a new stereo instead.*
Uncontrollable: *but there's no way you can afford the membership fee.*

Six weeks after the move, things have turned out

Unexpected: *nothing like*
Expected: *exactly as*

you had expected.

Bad: *You haven't made any real friends in the new town and you feel very lonely and isolated. You are very upset*
Good: *You've made a lot of good friends in the new town, and you feel quite happy and at home. You are very pleased*
Unexpected: *and very surprised.*
Expected: *but not surprised.*

Experiment 2

Bad unexpected uncontrollable version with additional manipulations:
   Staff dinner: ... *as he always does/which he almost never does.*
   Neighbour's party: However, you suffer from *(very frequent/very occasional)* migraines, and *(as very often occurs/as very rarely occurs),* that evening you have a really bad one, so you can't go.
   Old friend: ... *as regularly happens/as very unusually happens*
   Sports-club: *as you would never have that kind of money/although you would very rarely not have enough money, at that time.*

APPENDIX 2

Examples of spontaneous counterfactual thoughts and causal explanations in each content category produced by participants in Experiment 1

Counterfactual thoughts

Specific inhibitory:
"I should have joined the sports club instead of spending the money on that stereo – I would have got to know so many people"

Specific facilitative:
"I could have tried to make my own new friends instead of just continuing to go out with my old ones"

General internal:
"I wish I had tried harder"

General external:
"If only I could slough off the strains of maturity".

Causal explanations

Specific inhibitory:
"Being sick for the neighbours' dinner party was bad luck and therefore I don't know any of my colleagues"

Specific facilitative:
"Things have turned out well, mainly because I went to the staff dinner"

General internal:
"It's my own fault for not taking all the opportunities to make new friends"

General external:
"Due to unfortunate circumstances beyond my control I have been unable to establish myself in the local social scene".