LEABHARLANN CHOLÁISTE NA TRÍONÓIDE, BAILE ÁTHA CLIATH | TRINITY COLLEGE LIBRARY DUBLIN
Ollscoil Átha Cliath | The University of Dublin

Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

The Role of Context in Counterfactual Thinking

Clare Walsh

Ph. D.

University of Dublin, Trinity College

2001

TRINITY COLLEGE
13 FEB 2002
LIBRARY DUBLIN
THESIS 6607

Declaration

(a) The work contained in this thesis has not been submitted as an exercise for a degree at this or at any other university.

(b) This thesis is the results of my own investigations. The contributions of others are duly acknowledged in the text wherever included.

(c) I give permission to the Library to lend or copy this thesis on request.

Signed: Clare Walsh

Date: 20|12|2001

Acknowledgements

My greatest thanks goes to Ruth Byrne for sparking my in my interest in cognitive science, for her constant interest, encouragement and advice throughout my time in Trinity College and for being a great supervisor.

I would also like to thank Phil Johnson-Laird for allowing me to spend three months in his lab in Princeton. I learned a great deal from him and this visit led to one of the series of experiments reported in this chapter.

Thanks also to everyone who commented on this research, particularly members of our lab, Rachel McCloy and Alice McEleney for lively discussions and visitors to our lab, Orlando Espino, Yvegeniya Goldvarg and Juan Garcia-Madruga.

On a personal level, I would like to thank my family and friends for their support and everyone in the psychology department for making it a very enjoyable environment to work in.

Finally, I would like to thank those who funded me, the University of Dublin, Enterprise Ireland and the Irish Council for the Humanities and Social Sciences and the participants who gave up their time, mostly voluntarily, to take part in our experiments.

Table of Contents

		Page
Chapter 1	**Introduction**	
	Introduction	1
	Counterfactual Thinking	2
	The Mental Representation of Different Possibilities: The Mental Model Theory	11
	Theories of Counterfactual Thinking	16
	Aims of the Thesis	18
Chapter 2	**A Computational Model of Counterfactual Thinking: The Temporal Order Effect**	
	Introduction	20
	The Temporal Order Effect	21
	Two Computational Simulations of the Temporal Order Effect	26
	TempFocus	26
	TempContext	35
	General Discussion	48
Chapter 3	**An Experimental Study of the Temporal Order Effect in Counterfactual Thinking**	
	Introduction	50
	Experiment 1	56
	Experiment 2	65
	Experiment 3	72
	General Discussion	77
Chapter 4	**The Effect of Reasons for Acting on Counterfactual Thinking**	
	Introduction	83
	Experiment 4	94
	Experiment 5	102
	Experiment 6	109
	General Discussion	116

Chapter 5 The Mental Representation of Multiple Possibilities: A study of inferences

Introduction 124

Experiment 7 131

Experiment 8 142

General Discussion 154

Chapter 6 Discussion

Introduction 160

Summary of Findings 161

Implications 174

Future Questions 180

Conclusions 182

References 183

Appendices 195

Appendix 1: TempFocus program code 196

Appendix 2: TempContext program code 215

Appendix 3: Materials used in the temporal order effect experiments 234

Appendix 4: Materials used in the reasons for acting experiments 237

Appendix 5: Materials used in the mental representations experiments 250

List of Figures

Figure	Title	Page
Figure 2.1	A flowchart of the steps in the TempFocus Model	28
Figure 2.2	A flowchart of the steps in the TempContext Model	38

List of Tables

Table	Title	Page
Table 2.1	Outline of the steps in the TempFocus Model	27
Table 2.2	Representation of the Possibility-Models in the TempFocus model and the TempContext model	32
Table 2.3	Outline of the steps in TempContext	37
Table 2.4	Simulations of TempContext using the card selection scenario	47
Table 3.1	Different descriptions of the winning conditions used in the experiments, and the initial set of mental models that represent them. In each of the three experiments, the facts were the same: John picked black and Michael picked black and so they lost.	55
Table 3.2	The percentages of mutations and judgements of guilt and blame in Experiment 1	62
Table 3.3	The percentages of mutations and judgements of guilt and blame in Experiment 2	69
Table 3.4	The percentages of mutations and judgements of guilt and blame in Experiment 3	76
Table 4.1	Mental models of causes and reasons	91
Table 4.2	The percentages of mutations of the action and the reasons in the four counterfactual thoughts generated in the three conditions of Experiment 4	100

Table 4.3	The percentages of mutations of actions and events, and of reasons and causes in Experiment 5	106
Table 4.4	The percentages of mutations of the action and the reasons in Experiment 6	114
Table 5.1	The four sorts of semantic content in Experiment 7	133
Table 5.2	The eight types of inference examined in Experiment 7 together with the correct conclusions	135
Table 5.3	The percentages of correct responses to the eight forms of inference in Experiment 7. Inferences are in bold where the categorical premise is explicitly represented in the initial models of the first premise.	137
Table 5.4	The response times in seconds for correct responses to the eight forms of inference in Experiment 7. The times are in bold where the categorical premise is explicitly represented in the initial models of the first premise	138
Table 5.5	The percentages of correct answers and the response times for each of the four types of semantic content in Experiment 7	139
Table 5.6	The five sorts of semantic content in Experiment 8, and the number of fully explicit models compatible with each content	145
Table 5.7	The percentages of the "Yes", "No" and "Cannot Tell" responses in Experiment 8. The predicted responses are shown in bold, and the correct responses are underlined	150

Table 5.8 The response times in seconds for correct responses to the four forms of inference in Experiment 8 151

Table 5.9 Percentage of correct and predicted responses and question response times in each of the five semantic conditions 153

Summary

The aim of this thesis is to examine the mental representations and cognitive processes involved in counterfactual thinking, that is, in imagining how past events might have happened differently. In chapter 1, we review the functions of counterfactual thinking, the types of counterfactuals people generate most readily and the mental representation of counterfactuals. We describe two broad theories of the processes underlying counterfactual thinking, which we term the focus theory and the context theory. The focus theory proposes that people focus on salient aspects of the facts in selecting an event to mutate. In contrast, the context theory proposes that the context of possibilities also plays a role in the recruitment of alternatives to the facts.

In chapter 2, we describe two computational models, TempFocus and TempContext which we developed to simulate theories of the temporal order effect in counterfactual thinking. We describe how both models simulate the temporal order effect for certain conditions, that is, they mutate the most recent event in an independent sequence. We also describe how the models perform in novel conditions. TempContext produces a reversal of the temporal order effect when the background context is changed and we therefore make novel predictions about human performance on similar tasks.

In chapter 3, we describe three experiments which test the predictions of the theories empirically. The results corroborate the predictions of the context theory and they support the view that the context of possibilities plays a role in the recruitment of counterfactual alternatives.

In chapter 4, we test the context theory in a different domain. In three experiments, we examine the effect of reasons for acting on the mutability of controllable actions. We suggest that reasons form the context in which actions are performed and that

they play a role in the recruitment of alternative possibilities. In support of the context theory, we demonstrate that certain reasons can shift the focus of counterfactual thoughts.

In chapter 5, we study the mental representations underlying counterfactual thinking. In two experiments, we show that the verbal description of a sentence can influence what people make explicit in their mental models. We also show that co-reference can allow reasoners to construct more concise models, even when co-reference occurs across different possibilities.

In chapter 6, we summarise our findings. Our results support the context theory of counterfactual thinking; they show that the context of possibilities is important in the imagination of counterfactual alternatives. We discuss the implications of our results for other aspects of cognition and emotion and we suggest some questions for future research.

Chapter 1 - Introduction

In our everyday lives, many things happen which influence the course of events. We make a choice to visit a favourite place, we have a chance meeting with an old friend or we discover a new book and these actions and events interact in all kinds of expected and unexpected ways. But we live our lives not only in this reality, we perceive these events in the context of the many ways in which they could have happened differently. Our objective in this thesis is to examine the mental representations and cognitive processes underlying the imagination of these alternatives to reality.

This type of thought is referred to as counterfactual thinking. We examine an unexplored aspect of this topic, that is, how context may influence the counterfactual alternatives that people bring to mind. We test experimentally and computationally two broad theoretical approaches to the study of counterfactual thinking. We have used a range of methodologies, including methodologies which are novel to the study of counterfactual thinking. We develop computational models to simulate theories of the representations and processes underlying counterfactual thinking. We use not only standard scenario-based experiments to examine what aspect of a situation people focus on in their counterfactual thoughts but also inference tasks to examine the mental representations underlying counterfactual thinking.

This chapter is divided into three main sections. The first section reviews the literature on counterfactual thinking, including the functions of counterfactual thinking and what counterfactuals people generate most readily. The second section is a review of mental models and the representation of counterfactuals possibilities and third section reviews two broad theories of how people generate counterfactuals.

Counterfactual Thinking

The tendency for people to imagine events beyond those that actually occurred appears to form a central and pervasive part of human thought (e.g., Byrne, 1997). Counterfactual thinking emerges early in development. At pre-school age, children can draw conclusions about what might have happened if antecedent events had been different (Harris, German & Mills, 1996). Counterfactual thinking occurs across cultures, even in the absence of linguistic cues (Au, 1983). In fact, a difficulty in imagining counterfactual alternatives to reality may underlie many of the difficulties experienced by individuals with brain lesions in the dorsolateral prefrontal cortex, such as inflexible behaviour and suppressed emotions (Knight & Grabowecky, 1995). These patients appear excessively bound by environmental cues. They are often uncreative and have difficulty making plans. These difficulties give us some hints to the range of functions that counterfactual thinking may serve.

The Functions of Counterfactual Thinking

The way that people perceive an episode, their judgements about it and their affective reactions to it - each of these is influenced not only by the facts of the episode but also by the availability of counterfactual alternatives to it. Counterfactual thinking allows people to imagine what the world would be like if some aspect of it were different.

Counterfactual thinking may play a role in various cognitive tasks. Much of reasoning involves thinking about possibilities and counterfactuals may provide us with counterexamples to a solution (Johnson-Laird & Byrne, 1991). For example, from the following conditional:

If Susan had some money, then she went to the concert.

and the belief that Susan has money people may infer that she will go to the concert. But, subsequently they may withdraw the conclusion if they consider a counterexample such as:

If she had spent the money on something else, would she have gone to the concert?

Counterfactual thinking may help people to generate sub-goals in problem-solving tasks (Ginsberg, 1986). For example, if the problem is how to build a shelf, you could imagine:

If I had some wood, I could build a shelf

and then the task becomes to find some wood.

Counterfactual thinking also allows us to think about what would follow from an imaginary situation. For example, it makes it possible to infer how past events might have been avoided (Mandel & Lehman, 1996) and so may help people to learn from experience. It enables people to evaluate risks and to make plans for the future (Markman, Gavanski, Sherman & McMullen, 1993; Roese, 1994). In one recent study, students who imagined how they could have performed better after receiving the results of an examination tended to perform better in their next examination (Nasco & Marsh, 1999).

Counterfactual thinking has also been implicated in judgements such as causal inference (Wells & Gavanski, 1989) and in a range of emotions including regret (Kahneman & Tversky, 1982a) and sympathy (Miller & McFarland, 1986). Certain emotions such as disappointment appear to rely on the ability to imagine that events could have turned out differently (Roese & Olson, 1995a). Counterfactual thinking may amplify the emotional reaction to an event. Missing a desired outcome is more upsetting if it almost occurred (Kahneman & Tversky, 1982a) or if its cause is abnormal and hence more mutable (Kahneman & Miller, 1986). For example, people expect an individual to experience more regret when an unwanted outcome follows an exceptional rather than a

routine action and when it follows an action rather than a failure to act. These effects have also been generalised to feelings of sympathy and blame (Turley, Sanna & Reiter, 1995) and to affect-meditated judgement such as perpetrator punishment and victim compensation (Macrae, Milne & Griffiths, 1993; Miller & McFarland, 1986).

The impact of counterfactual thinking on emotions may operate through a contrast effect. People tend to report more negative affect when they imagine how the outcome could have been better (an upward counterfactual) than when they imagine how it could have been worse (a downward counterfactual), regardless of whether the counterfactual is evoked spontaneously through situational factors (Markman et al., 1993) or through instruction (Roese, 1994). In fact, narrowly missing a positive outcome may produce more negative affect than an objectively worse situation (Johnson, 1986). For example, Olympic bronze medallists may feel happier than silver medallists who may be more inclined to imagine the missed gold medal (Medvec, Madey & Gilovich, 1995)

The emotion experienced may also be directly related to the type of antecedent mutated. For example, undoing actions is central to the experience of guilt (Niedenthal, Tangney & Gavanski, 1994) and regret (Gilovich & Medvec, 1995) whereas undoing a personal quality is related to the experience of shame (Niedenthal et al., 1994). Counterfactuals with uncontrollable antecedents are often experienced as disappointing (Zeelenberg et al., 1998) or as unlucky (Teigen 1995). Furthermore, the belief that one is more or less lucky than another individual can lead to feelings of gratitude and envy respectively (Teigen, 1997).

McMullen, Markman and Gavanski (1995) argued that the influence of self-implicating counterfactuals on affect is mediated by perceptions of control. Specifically, they argued that perceived control might moderate negative affect by indicating means for future improvement. Boninger, Gleicher & Strathman (1994) found support for this

proposal by showing that negative affect resulting from counterfactual thinking was reduced for individuals who had a dispositional tendency to consider the future consequences of current behaviour or when people were directed to focus on the future.

This broad range of cognitive and emotional skills may have at their core the ability to imagine counterfactual possibilities. To understand these effects, we need to establish how people engage in counterfactual thinking.

Generation of Counterfactual Thoughts

Counterfactuals have been studied in philosophy (e.g., Lewis, 1973; Stalnaker, 1968), psychology (e.g., Kahneman & Miller, 1986) and artificial intelligence (e.g., Costello & McCarthy, 1999; Ginsberg, 1986). A proper understanding of counterfactuals is essential to the important endeavour in cognitive science to develop and simulate a theory of rationality (e.g., Johnson-Laird & Byrne, 1991; Rips, 1994). In the psychological research, one of the main objectives has been to identify what counterfactuals people generate most readily. Some counterfactual alternatives may be more similar to reality (Lewis, 1973) or more natural than others (Hofstadter, 1979). Imagine the situation described by Hofstadter (1979) in which you are driving down a country road and you run into a swarm of bees. He describes some counterfactuals which may readily spring to mind, for example, "Too bad my window wasn't closed" or "Lucky, I wasn't on my bike". Other alternatives seem very strange, for example, "Too bad those bees weren't dollar bills" or " I bet those bees would rather have had a collision with a rosebush". In fact, research in psychology has shown that there are considerable regularities in the sorts of past events that people mentally undo, despite the numerous ways that things could have happened differently.

The counterfactual thoughts that people entertain may be determined by their motivations. For example, people may be more likely to engage in counterfactual thinking if the outcome is bad or if it is of personal significance (Roese & Olson, 1995a). The selection of a counterfactual may also be goal-directed (Seelau, Seelau, Wells & Windschitl, 1995). For example, if the purpose is to control future outcomes then people may imagine how past events could have turned out better but if the purpose is to console others they may be more likely to imagine how things could have been worse. People's counterfactual thoughts may also be influenced by their cognitive capacity (e.g., Byrne, 1997; Legrenzi, Girotto, & Johnson-Laird, 1993). For example, people tend to make minimal changes to the factual situation (Byrne, 1997; Pollock, 1986) and they may prefer to delete unlikely events because alternatives to them are readily available (Kahneman & Miller, 1986).

Researchers have established that when people think about counterfactual alternatives they tend to focus on focal events, non-ideal events, the most recent event in an independent sequence but the first event in a causal sequence, exceptional events and intentional actions.

In studying counterfactual thinking, the standard methodology has been to present participants with a scenario describing a series of events leading to some outcome. In order to determine what factors contribute to the generation of counterfactual thoughts and their consequences, researchers have typically manipulated some aspect of the scenario, either a characteristic of the outcome, characteristics of some of the antecedent events or the causal structure of events. Counterfactual thinking may be evoked directly by asking participants for an imaginary alternative to the described scenario or may occur spontaneously while participants generate a free-thought listing. The responses may be analysed in a number of ways. These include the number of counterfactuals reported,

their content, their structure, that is, whether they add items or delete them and their direction, that is, whether the imagined outcome is better or worse than the factual one.

Focal Events

One of the original proposals of Kahneman and Miler (1986) was that the mutability of any aspect of a situation will be enhanced if attention toward it is increased whereas unattended aspects of the situation become part of the background. For example, in one study participants were taught the rules of a two-person card game. They were then shown the hands of the two players, A and B and were asked to complete stems such as "A would have won if …..". Participants tended to imagine changes to A's hand even though the same outcome could have been achieved just as easily by changing B's hand (Read, 1985).

Non-ideal Events

People also tend to alter an event by imagining that it was better than by imagining that it was worse. For example, in the card game described above, participants were asked to change the outcome by changing one card, e.g., "The outcome would have been different if the _____ had been a _____ ." Most participants changed the outcome by improving the losing hand rather than by weakening the winning one (Read, 1985). Ideal values tend to be judged as typical (Barsalou, 1985) and therefore may be highly available (Kahneman & Miller, 1986).

Order of Events

People tend to focus on the most recent event in an independent sequence in the events. Consider for example, a scenario in which two individuals, Jones and Cooper, toss a coin.

If both coins come up the same, they each win $1,000, otherwise they win nothing. First, Jones tosses a head, and then Cooper tosses a tail. As a result, they both lose. Most participants in an experiment who read this scenario found it easier to undo the outcome by imagining Cooper tossing a head and he was expected to experience more guilt and to be blamed more by Jones (Miller & Gunasegaram, 1990). Similarly, when participants played the role of students in an examination context, they tended to judge the questions as more unfair when they were set after rather than before they had studied (Miller & Gunasegaram, 1990). The effect may occur because the first event is presupposed (Miller & Gunasegaram, 1990) or because it plays an initialising role in the formation of a model (Byrne, Segura, Culhane, Tasso & Berrocal, 2000).

In contrast, when events are causally related people tend to focus on the first event in a sequence. Wells, Taylor and Turtle (1987) argue that people prefer to undo causes than effects because it is relatively difficult to imagine a cause occurring without its effect unless the observed effect is unusual (Kahneman & Miller, 1986) or perhaps the cause is indeterminate (Kahneman & Varey, 1990). In support of this hypothesis, they found that people prefer to mutate the first rather than subsequent events in a causal chain, regardless of the nature of that event. The order of events is fundamental to our understanding of a situation and plays a very important role in influencing the mutability of events. We will return to it in chapters 2 and 3.

Exceptional Events

Normal events may be more available than exceptional ones. One of the original findings in this area was that people are more likely to undo exceptional than routine events (Kahneman & Tversky, 1982b). In one experiment, participants read a scenario involving a fatal motor accident. In one condition, they were told that the man had left at an unusual

time but had taken his usual route and in the other condition they were told that he had left at his usual time but had taken an unusual route. In both conditions, participants tended to change the event which was unusual. The effect has also been demonstrated in a field study of accident victims (Davis, Lehman, Silver, Wortman, & Ellard, 1996). In this study, individuals whose behaviour was atypical at the time of the accident tended to spend more time thinking about how they could have avoided it.

Exceptional events may be highly mutable because they are salient (Seelau et al., 1995) or because they bring to mind their corresponding norms (Kahneman & Miller, 1986). However, the effect only occurs where the outcome is also exceptional (Gavanski & Wells, 1989). When the outcome is normal people tend to focus on normal antecedents. An alternative explanation is that normal events are constrained by prior causes, such as social rules, legal rules and habits. In contrast, exceptional events are those that occur in spite of these constraints (Wells, Taylor & Turtle, 1987).

Actions

There is a tendency for people to focus on actions rather than failures to act in their counterfactual thoughts (Byrne & McEleney, 2000; Gilovich & Medvec, 1994; Kahneman & Tversky, 1982a). For example, in one study, participants were given a scenario about two individuals own company shares. One individual switches his share holding to another company whereas the other decides not to switch and as a result of their choices both lose money. Most people expect the individual who switched his shareholding to experience more regret (Kahneman & Tversky, 1982a). The finding occurs for good outcomes as well as bad (Byrne & McEleney, 2000) and in the long-term as well as the short-term when the consequences are known (Byrne & McEleney, 2000).

The tendency to mutate actions may be based on the fact that actions are more abnormal than inactions (Kahneman & Miller, 1986) or the fact that people have explicitly represented alternatives to actions but not inactions in their mental models (Byrne & McEleney, 2000). However, the effect appears to be context-dependent. In some contexts actions may be more normal than failures to act. For example, people are more likely to mutate the failure to act than the actions of a football manager when his team is performing badly (Zeelenberg, van den Bos, van Dijk, Pieters, 2001).

Controllable Events

Individual actions or *controllable events* are also mutated more frequently than external events, i.e., *uncontrollable events* (Girotto, Legrenzi & Rizzo, 1991). Girotto et al. define a controllable event as one which depends on the decision of an individual. They presented subjects with a scenario in which a man is delayed on his journey home by a causal chain of events, one of which was controllable, that is, the decision to drink a beer in a bar. The bar event was mutated more often than any other event, regardless of its position in the causal chain.

The focus on controllable events in counterfactual thinking may occur because individual actions tend to be salient and hence people represent them explicitly in their models (Legrenzi et al., 1993). Alternatively, actions may be perceived as less constrained than events which occur in the surrounding background, because these are normally assumed to be independent of external causes (Girotto et al., 1991). This tendency to focus on controllable events has been shown to be a key effect and we will return to it in chapter 4.

We have described various factors that people focus on in their counterfactual thoughts. There may of course be others. We will return to examine some of these factors in later chapters.

The Mental Representation of Different Possibilities: The Mental Model Theory

The mental model theory (Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991; in press) proposes that people represent different states of affairs or possibilities in models. These models correspond to the structure of events in the world rather than to the language which is used to describe it. Mental models may represent factual situations, counterfactual situations, i.e., those situations which were once possible but no longer are, and other hypothetical possibilities. People may construct a different model to represent each possibility. A factual situation may be represented by a single model. For example, the following sentence:

Peter leaves now and he arrives on time

may be represented as follows:

leaves on-time

where 'leaves' represents 'Peter leaves now' and 'on time' represents 'Peter arrives on time'. Conditionals, which are consistent with a number of possibilities, may also be represented by mental models. Take, for example, the following conditional:

If Peter leaves now then he arrives on time

The representation may depend on the interpretation of the conditional. Logicians distinguish two different interpretations of *if*. One interpretation of a conditional is a material implication which is consistent with the following three possibilities:

leaves on-time

¬ leaves ¬ on-time

¬ leaves on-time

where "¬" is a propositional-like tag to indicate negation (see Johnson-Laird, Byrne & Schaeken, 1992). Different models are represented on different lines of the diagram and each model represents a different possibility. Alternatively, the conditional may be interpreted as a material equivalence or a biconditional: 'If p then q' is taken to also imply 'if q then p'. According to this interpretation, the conditional is consistent with the first two models described above but not the third, as follows:

leaves on-time

¬ leaves ¬ on-time

In everyday life, conditionals can have many different meanings depending on the meaning of the antecedent and the consequent and the context in which they are placed (Johnson-Laird & Byrne, in press).

Reasoners rarely construct a fully explicit set of models. Because of working memory constraints, they may represent as little information as possible (Johnson-Laird & Byrne, 1991; Johnson-Laird, Byrne, & Schaeken, 1992). Their initial set of models makes some information explicit and leaves other information implicit. Their construction of mental models is guided by the following principle of truth (Johnson-Laird and Byrne, in press):

"The fundamental representational principle of the theory is:

> The principle of truth: each mental model of a set of assertions represents a possibility given the truth of the assertions, and each mental model represents a clause in the assertions only when it is true in that possibility.

The principle implies that mental models represent only what is true, and not what is false. Moreover, each mental model represents a clause in the premises only when it is true within the possibility that the model represents."

As a result, the initial models of a conditional may contain just one explicit model and an implicit model to denote that there are alternative possibilities. The conditional we described earlier may be represented as follows:

leaves on time

. . .

The ellipsis denotes an implicit model, which acts as a "place holder" for the possibilities in which the antecedent is false. A disjunction may also be represented by models which make just some information explicit. For example, to represent the following sentence:

Either Peter leaves now or else he won't arrive on time.

people may construct two models, as follows:

leaves

¬ on time

The models do not represent the false possibilities in which Peter leaves now and doesn't arrive on time or in which he doesn't leave and arrives on time. In addition, the first model does not explicitly represent the fact that it is false that Peter does not arrive on time and in the second model that it is false that Peter leaves now.

If necessary, the models can be fleshed out to make the implicit information explicit. The models contain annotations, akin to 'mental footnotes' that indicate how they can be fleshed out (see Johnson-Laird, Byrne, & Schaeken, 1992; Johnson-Laird & Savary, 1995; chapter 2, for discussion of the use of footnotes in mental models). The footnotes refer to what can occupy the 'empty' spaces in the models. When people engage in counterfactual thinking they flesh out their models to consider additional

possibilities. In the following section, we examine how people may represent counterfactual conditionals.

Mental Representation of Counterfactual Conditionals

Counterfactuals allow people to imagine what the world would be like if some things were different. For this reason perhaps, people usually generate counterfactuals in the form of a conditional statement. In English, counterfactuals are usually expressed in the subjunctive mood:

If it had been sunny yesterday, then I would have gone to the beach.

Fillenbaum (1974) demonstrated that counterfactual conditionals convey the presupposition that their antecedents and consequents are false. In a memory task, he found that participants who had read counterfactuals such as,

If he had caught the plane, he would have arrived on time

later often mistakenly recognised the negation of the antecedent or the consequent, e.g.,

He did not catch the plane

or

He did not arrive on time

This finding is consistent with the idea that counterfactual conditionals are represented by two explicit mental models from the outset (Byrne & Tasso, 1999). People may construct a model of the counterfactual situation which is explicitly stated, but also a model of the presupposed factual situation, i.e.,

Factual: ¬ plane ¬ on time

Counterfactual: plane on time

. . .

where 'plane' represents 'he caught the plane' and 'on time' represents 'he arrived on time'. The models may be annotated to keep track of their epistemic status (Byrne & Tasso, 1999; Johnson-Laird & Byrne, 1991). The representation of the counterfactual conditional therefore contrasts with the representation of a factual conditional, such as:

If he caught the plane, he arrived on time

which is represented by one explicit model as described earlier (see p. 13).

When people generate a counterfactual, they may mentally undo one of the facts and this may lead to an outcome which is the same as the factual one or which is different. The term semi-factual is often used to describe the situation where the antecedent is false but the consequent is true (Goodman, 1973; McCloy & Byrne, in press). These conditionals can be expressed with the connective 'even if', e.g.,

Even if he had caught the plane, he would have arrived on time.

Semi-factuals may be represented as follows:

Factual:	¬ plane	on time
Semi-factual:	plane	on time

. . .

Further evidence that people represent counterfactual conditionals in this way comes from studies of deductive reasoning. For counterfactual conditionals, the explicit representation of the factual situation (not-p and not-q) allows reasoners to draw the modus tollens inference (not-q, ∴ not-p) and denial of the antecedent inference (not-p, ∴ not-q) more readily than from a factual conditional (Byrne & Tasso, 1994; 1999). Similarly for semi-factuals, the explicit representation of the factual situation (not-p and q) leads reasoners to draw fewer affirmation of the consequent inferences (q, ∴ p) and

denial of the antecedent inferences (not-p, ∴ not-q) than from the corresponding factual conditional (Moreno-Rios, Garcia-Madruga & Byrne, 2001).

Theories of Counterfactual Thinking

In the previous sections, we have reviewed the literature on counterfactual thinking and we have reviewed how people may represent counterfactuals in mental models. In this section, we review two alternative theories of *how* people engage in counterfactual thinking, the focus theory and the context theory.

Focus Theory

One prominent viewpoint, which we term the focus theory, is that when people generate a counterfactual, they focus exclusively on the facts and they select one of those facts which they mentally undo or alter in some way (Gleicher et al., 1990; Kahneman & Tversky, 1982b; Legrenzi, Girotto & Johnson-Laird, 1993; Roese, 1997; Seelau, Seelau, Wells & Windschitl, 1995).

When people think about a situation, they may build a model of the facts (e.g., Johnson-Laird, 1983) and they may restrict the items which they mutate to those that are explicitly represented in this model (Legrenzi, Girotto & Johnson-Laird, 1993). They may mutate an event that is salient (Seelau et al., 1995) or one that is the main object of concern or attention (Kahneman & Tversky, 1982b) and the mutability of an event may be enhanced if attention is drawn to it (Read, 1985). According to the theory, background knowledge is only consulted after an item is mutated in order to decide whether the mutation leads to a new outcome.

The focus theory may explain the tendency for people to mutate focal events (Read, 1985) including the actions of a central character, which tend to be particularly

salient (Legrenzi et al., 1993; Morrow, Bower, Greenspan, 1989). The mutability of exceptional events may also be enhanced because unusual events tend to be more salient than normal ones (Seelau et al., 1995). In summary, the focus theory proposes that the mutability of an event is dependent on the representation and salience of the facts.

Context Theory

An alternative viewpoint, which we term the context theory, is that the mutability of events depends on the availability of alternatives to the facts. The theory is an extension of norm theory (Kahneman & Miller, 1986) but has not been studied systematically to date. There are however, hints in the literature to support this view (Byrne, 1997). An event is more mutable if an alternative to it is explicitly represented in the model of the facts (Byrne & McEleney, 2000; Byrne et al., 2000).

When people experience an event, they often recall similar sequences of events to interpret and make sense of the current situation (e.g., Baddeley, 1999). The context theory proposes that these alternatives may form the basis for the generation of imaginary alternatives to reality. People may use the facts of a scenario to access past experiences which are similar to the factual one. Aspects of the factual situation which are different in the imagined alternative will be mutated. Take the example given by Kahneman & Miller (1986) in which Marty does not cry at his aunt's funeral. People may mentally simulate the alternative that Marty does cry based on their experience that people often cry at funerals. The theory proposes that the context in which an event is placed may play an important role in recruiting alternatives to the facts.

The theory can explain the mutability of exceptional events because exceptional events by definition will be absent in many recalled experiences. The theory can also explain how the mutability of an event may be influenced by other events in the scenario,

for example, the argument that causes constrain the mutability of their effects (Wells, Taylor & Turtle, 1987) and that actions are more mutable because they are perceived as unconstrained by external events in the world (Girotto, Legrenzi & Rizzo, 1991).

To understand how people construct a counterfactual alternative to the present one, the critical questions become which aspects of reality to mutate and which aspects to leave unchanged. These questions may be complementary, however as we aim to show, the focus may give rise to different theories of the processes involved in constructing an alternative. Extensive research in psychology has focused on the former question in an attempt to discover what events are most mutable and why. In this thesis, we aim to readdress this balance by examining the role of the context in counterfactual thinking.

Aims of the Thesis

Our aim in this thesis is to examine the role of context in the mental representations and cognitive processes underlying counterfactual thinking. We aim to explain why some counterfactuals are easier to imagine than others. Most studies of counterfactual thinking have studied *what* people think about when they imagine alternatives to reality. Our aim is to make steps toward the development of a theory at the algorithmic level. In other words, our interest is in understanding *how* people generate these counterfactual alternatives.

In our review, we have grouped together two broad types of explanations for why people tend to focus on some events more than others in their counterfactual thoughts. In general, these explanations do not give a detailed account of the processing steps involved and so we have attempted to give a description at the algorithmic level which is consistent with the general explanations given. We have labelled the two theories as the focus theory and context theory. In chapter 2, we describe two computational models that

we have developed to simulate how these two theories may explain the temporal order effect in counterfactual thinking. The building of mental representations in both programs is guided by the mental model theory (Johnson-Laird & Byrne, 1991). The context model describes a new theory of how people may recruit alternative scenarios. It also describes how the mental representation of these alternatives may influence their accessibility.

The two theories give rise to different predictions about the mutability of events under different conditions. In chapters 3 and 4, we describe two series of experiments designed to distinguish these theories. The studies aim to test whether the representation of the context can affect the mutability of events as the context model predicts. In chapter 3 we report three experiments on the temporal order effect in counterfactual thinking which aim to show that the context can shift the focus of people's counterfactual thoughts from the last to the first event. In chapter 4, we describe three experiments on the effect of reasons on the mutability of controllable actions which also aim to show that background knowledge can shift the focus of people's counterfactual thoughts from the action to the reason. In chapter 5, we describe two experiments of the effects of the nature of verbal descriptions on the mental representations people construct and the inferences they make. These studies show that the mental representations that people construct are influenced by the description of a sentence and the co-referential links contained in it.

In the final chapter, we discuss the implications of our findings for our understanding of the mental representations and cognitive processes underlying counterfactual thinking and related aspects of cognition and emotion. We also outline some new questions that need to be addressed if we are to develop a more complete theory of counterfactual thinking at this level.

In summary, the aim of this thesis is to examine the role of context in the mental representations and cognitive processes underlying counterfactual thinking.

Chapter 2 Two Computational Models of Counterfactual Thinking: The Temporal Order Effect

In Chapter 1, we described two different theories of the mental representations and cognitive processes underlying counterfactual thinking. Our aim in this chapter is to develop these theories in more detail for one aspect of counterfactual thinking, that is the temporal order effect. We describe two computational models which we have developed to simulate these theories of the temporal order effect. There has been just one previous computational model of counterfactual thinking (Byrne, Culhane & Tasso, 1995). The building of these computational models allows us to ensure that we have a coherent account of each theory, that the two models are systematically different from each other and that they lead to different predictions.

Both theories assume that people construct mental representations of the facts and of possible alternatives to the facts which we term the context of possibilities. In our models of the temporal order effect, the context of possibilities is provided by the conditions which determine the outcome of the facts. The focus theory relies exclusively on the representation of the facts in selecting an event to mutate. In the case of the temporal order effect, the theory assumes that people focus on the last event. The context of possibilities is examined *afterwards* to see if the mutated event changes the outcome. In contrast, the context theory assumes that people uses the facts to recruit a counterfactual alternative from the context of possibilities and the difference between the models of the facts and of the selected alternative determines what events they mutate. We developed a context model which extends this theory by assuming that, in some cases at least, the background conditions may provide one or more counterfactual alternatives. The theory predicts that a change in the representation of the background conditions may

influence the accessibility of different alternatives and hence what people mutate. After describing the two models, we demonstrate conditions in which the context model leads to an elimination and a reversal of the temporal order effect. This is important because it allows us to make novel predictions about human performance under different conditions and it allows us to distinguish the two theories. In chapter 3, we will describe empirical tests of these predictions.

The Temporal Order Effect

Research has shown that when a series of events are independent of each other, people tend to mutate the most recent event (Byrne, Segura, Culhane, Tasso & Berrocal, 2000; Miller & Gunasegaram, 1990; Spellman, 1997). Consider the following scenario (from Byrne et al., 2000):

> 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 colour (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 tended to undo the second event, e.g., if only Michael had picked red too, and this finding has been termed the temporal order effect. In addition, the second player, Michael, was usually expected to experience more guilt and to be blamed more by John. This effect has also been demonstrated in a number of practical situations. Miller and Gunasegaram's (1990) original demonstration of this effect (with a coin-toss scenario

similar to the card selection scenario) was also replicated by them in an everyday setting using an examination context. The questions prepared by the student and those set by the teacher must match for a successful outcome, and judgements of the fairness of the questions exhibited a temporal order effect. Sherman and McConnell (1996) also found that when participants were asked to rank team performance in a basketball league, they tended to give greatest weight to the teams' most recent performance. Miller and Gunasegaram (1990) suggest that the mutability of the last event may enhance its perceived responsibility, and that this may also underlie the preference for coaches to place the best player last on a relay team and the tendency for people to wager more on their predictions than on their postdictions.

Explanations of the Temporal Order Effect

Recently, several explanations have been advanced to account for the temporal order effect. One possible explanation is 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 which result in a large change in probability are assigned a greater causal role. Take for example, the card selection scenario described above. There are four possible combinations of card selections. In two of them both players pick the same card and win and in the other two they both select different cards and lose. The probability of winning is therefore 50% at the outset. After John selects a card, there are two possibilities left, one in which Michael picks the same colour card as John and one in which he picks a different colour card. The probability of winning is now one out of two or 50% and therefore the probability has not changed. However, once Michael draws a card, the outcome will be decided. They will either have won, in which case the probability changes to 100% or they will have lost in

which case the probability changes to 0%. Hence, the greatest change in probability occurs after Michael's selection and therefore he will be viewed as most causal. According to the crediting causality theory, counterfactual thinking plays a role in these judgements because the probability calculations rely on the ability to imagine the counterfactual possibilities. The perceived causal role of an event will be increased if undoing the event changes the outcome. However, the theory has difficulty in explaining recent findings by Byrne et al. (2000). They presented participants with a scenario similar to the card selection scenario described above but placed in the context of a television game show. The winning conditions are the same so that both players win £1,000 if they pick the same colour cards. However, while the game is being played there is a technical hitch which occurs as follows:

> Jones goes first and picks a black card from his deck. At this point, the game show host has to stop the game because of a technical difficulty. After a few minutes, the technical problem is solved and the game can be restarted. Jones goes first again, and this time the card that he draws is a red card. Brady goes next and the card that he draws is a black card. Thus, the outcome is that neither individual wins anything.

The results show that 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). Hence an alternative explanation is that the first event in an independent sequence may be relatively immutable because it is presupposed (Miller & Gunasegaram, 1990), acting as a background against which later events are perceived (Sherman & McConnell, 1996), and playing an important contextualising role in constructing a mental representation of a factual situation (Byrne

et al., 2000). Accordingly, it may be the greater mutability of later events which increases their perceived causal strength (Miller & Gunasegaram, 1990).

Our aim is to explain why the first event is presupposed or perceived as immutable. The explanations described here are consistent with the focus theory of counterfactual thinking to the extent that they assume that the mutability of the second event depends on the representation of the facts of the scenario. However, the findings from the technical hitch experiment give a hint that the availability of a contrast may be important in influencing event mutability. Our second theory, the context theory, assumes that the mutability of an event depends not only on the mental representation of the facts, but also on the representation of the counterfactual alternatives to the facts. In the case of the temporal order effect, these alternatives may be provided by the winning conditions.

Mental Representations underlying the Temporal Order Effect

People may understand the card scenario by constructing a set of mental models (Johnson-Laird & Byrne, 1991), and their models may represent certain aspects of the factual situation explicitly in the following sort of representation:

John red Michael black Lose

where 'John red' represents 'John picked a red card', 'Michael black' represents 'Michael picked a black card', and 'Lose' represents the outcome (Byrne et al., 2000). They may also generate models of the counterfactual possibilities. There are a total of four possibilities; two selections with two possible outcomes for each. The fully explicit set of models is as follows:

Factual:	John red	Michael black	Lose
Counterfactual:	John red	Michael red	Win
	John black	Michael black	Win
	John black	Michael red	Lose

The temporal order effect indicates that people construct just a subset of the possible counterfactual models:

Factual:	John red	Michael black	Lose
Counterfactual:	John red	Michael red	Win

. . .

Our aim is to examine why this is the case.

We will describe two computational models which simulate different theories of the processes underlying this effect. First we describe a program to simulate the focus theory of counterfactual thinking. It takes as its starting point the earlier explanations of the temporal order effect which emphasise the role of the representation of the factual event in producing the effect (Miller & Gunasegaram, 1990). In this program the context of possibilities is consulted only after a putative counterfactual scenario is generated from the factual events. Next we describe a second computational model that simulates the context theory of counterfactual thinking. The model gives central importance to the role of the context of possibilities in the selection of a factual event to mutate.

We will show how both theories produce the temporal order effect. We will also examine how the theories deal with different scenarios and we will show that in some cases they lead to different predictions about what people focus on in their counterfactual thoughts. The context theory makes a novel prediction: for a given set of facts, and set of winning conditions, the temporal order effect may either be observed (people will mentally undo the second event) or be reversed (people will mentally undo the first event)

depending on the *description* of the winning conditions, which in turn affects the mental representation of the context of possibilities and hence how accessible the different possibilities are. In the next chapter, we describe the results of three experiments that test these novel predictions.

Two Computational Simulations of the Temporal order Effect

The computational simulations are written in LISP. Full listings of the TempFocus Model and TempContext Model are provided in Appendix 1 and 2 respectively. Both programs take as input a set of facts and a description of the conditions which determine the outcome of the scenario, which we term the *context of possibilities.* In the card selection game these are the actual card selections (e.g., John picked red and Michael picked black) and the winning conditions (e.g., if they both pick red or both pick black they win) respectively. For any given set of facts and conditions, they generate a counterfactual alternative about how the outcome could have turned out differently (e.g., they would have won if Michael had picked red). However, both programs differ in the processes which they use to generate this outcome (Walsh & Byrne, 2001a). In the following sections, we describe the high level functions in each program.

TempFocus Model

Our first program simulates a long standing view of the temporal order effect that it depends centrally on the nature of the factual situation (e.g., Miller & Gunasegaram, 1990). The model assumes that the construction of a counterfactual scenario arises from operations carried out on the factual representation. It selects a mutable event from the facts and removes it from the scenario. The context of possibilities is subsequently

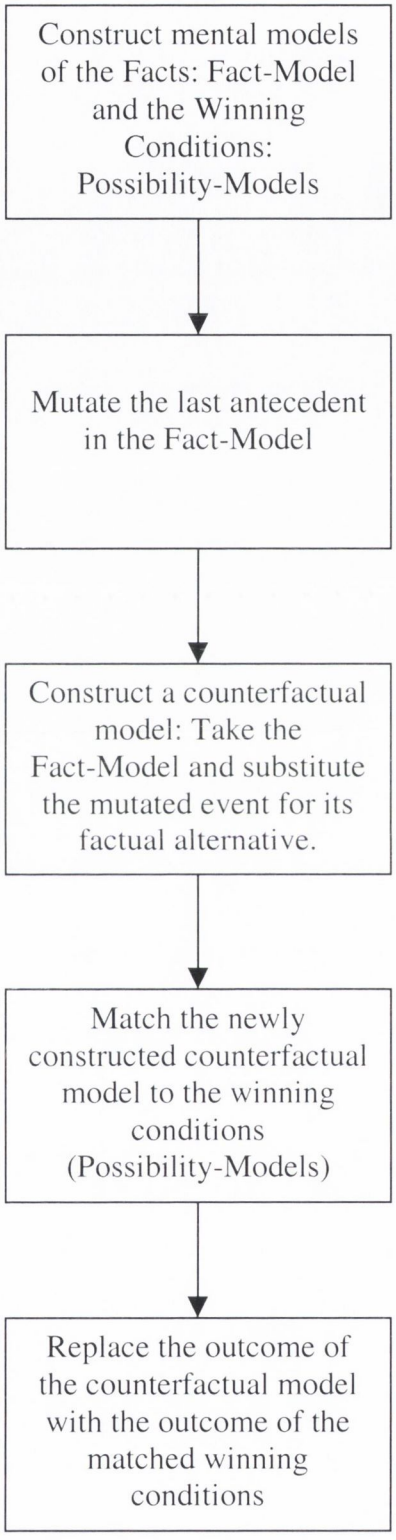
examined to provide a post-hoc check on the newly constructed putative counterfactual alternative.

The program consists of three main suites of functions, that comprise functions to represent the factual events and the context of possibilities (provided by the winning conditions), functions to mutate events and construct a counterfactual scenario, and functions to compare the newly constructed counterfactual scenario to the context of possibilities. The top-level function is called GenerateCounterfactual. It calls each of the three main sets of functions and it prints out the final counterfactual conditional. We will describe each of these suites of functions in turn. The steps in the program are also described in Table 2.1 and Figure 2.1. We will illustrate how the program works with reference to the description of the card selection scenario described earlier and we will show how it simulates the temporal order effect.

Table 2.1 Outline of the steps in the TempFocus model

1. Construct a set of models to represent the context of possibilities (Possibility-Models).
2. Construct a model to represent the facts (Fact-Model)
3. Select the last antecedent event from the Fact-Model.
4. Mutate this event by negating it.
5. Construct a new model by taking the Fact-Model and replacing the mutated event with its negation (Counterfact-Model).
6. Find the outcome of the Counterfact-Model by comparing it to the Possibility-Models.
7. Add the new outcome to the Counterfact-Model. If the new outcome is different from the facts, the model is a counterfactual and if it is the same it is a semi-factual. Exit the program.

Figure 2.1 A Flowchart of the steps in the TempFocus Model [1]



[1] The squares refer to processes.

Representing the Facts and the Context of Possibilities

The input to the program is a list of facts and a conditional which forms the context of possibilities describing the consequent of those facts. The program builds one model to represent the facts, which we call the Fact-Model and a set of models to represent the context of possibilities, which we call the Possibility-Models. Because the model is designed to simulate human performance, it represents as little information as possible explicitly (see Johnson-Laird & Byrne, 1991). The program simulates the construction of mental models which is guided by the principle of truth (Johnson-Laird and Byrne, in press, see also chapter 1, p12-13). For example, take the following conditional which is based on the card selection scenario in which two players, John and Michael, pick cards:

"If one or the other but not both pick cards from a black suit, each wins £1,000"

The program constructs two Possibility-Models to represent this, as follows:

John Black		Outcome Win
	Michael Black	Outcome Win

The program does not represent the losing conditions. Also, in the first model it does not represent explicitly the fact that Michael does not pick a black card and similarly in the second model that John does not pick a black card. However, the models can be fleshed out to be more explicit. The models contain annotations, akin to 'mental footnotes' that indicate how they can be fleshed out (see Johnson-Laird, Byrne, & Schaeken, 1992; Johnson-Laird & Savary, 1995, for discussion of the use of footnotes in mental models). The footnotes refer to what can occupy the 'empty' spaces in the models:

John Black	(NotNext)	Outcome Win
(NotPrevious)	Michael Black	Outcome Win

The footnote 'NotNext' indicates that the slot for Michael's card cannot be given the value it has in the next occurrence of it, likewise, the footnote 'NotPrevious' indicates

that the slot for John's card cannot be given the value it has in the previous occurrence of it. Table 2.2 describes the models that the program constructs to represent the context of possibilities when given different types of input.

The Fact-Model and the Possibility-Models each contain a set of antecedent events or situations and a consequent. Each item in the model is assigned a name and a value. The name is used to match events and the value is the part which is mutated. For example, 'John picked a red card' is named 'John' and it is assigned a value 'Red'. To denote that a value is negative, the event can be tagged with the word 'not'. If the event is not tagged, the function assumes that the value is affirmative. In chapter 5, we will return to the issue of the mental representation of events. The program also assumes that the events are inputted in their temporal order of occurrence.

The facts are inputted to the program as a list of antecedents and an outcome. For example, in the card selection scenario described above the input is as follows:

((John Red) (Michael Black) (Outcome Lose))

The conditions are inputted to the program as a list of events without their values, a set of antecedent conditions and the consequent. For example the conditional:

"If both players pick the same colour card (i.e., both red or both black), each individual wins £1,000."

is inputted to the program as follows:

((John Michael)

((Both Red) or (Both Black))

(Outcome Win))

The program constructs the following set of Possibility-Models to represent this input:

John Red	Michael Red	Outcome Win
John Black	Michael Black	Outcome Win

The program constructs models for different conditionals as described in Table 2.2. The antecedent conditions may directly specify the value to be assigned to a particular event, e.g., 'John picks a red card' and this will be inputted as follows:

((John red)).

Alternatively, the antecedent conditions may specify that both events, one event, or only-one event has a particular value, e.g., 'both players pick a red card' is inputted as follows:

((both red)).

The program also accepts four connectives: and, or (unspecified disjunction), ori (inclusive disjunction) and ore (exclusive disjunction). Each of these connectives is used to combine two sets of models, as described in Table 2.2. Given the connective 'and', the program combines each model in the first set with each model in the second set to construct a new set of models. Any models that contain inconsistencies are eliminated. Given the disjunction 'or', the two sets of models are appended to form one set of models. For the inclusive disjunction, the conjunction of the two models is also appended to this set and for the exclusive disjunction the two sets of models are appended and footnotes are added to record the fact that it is exclusive. In all cases, duplicate models are eliminated.

Table 2.2 Representation of the Possibility-Models in the TempFocus model and the TempContext model

	Events (John Michael) **Input**	**Models**	
1.	((Both Red))	John Red	Michael Red
2.	((One Red))*	John Red Michael Red	
3.	((Only-one Red))	John Red Michael Red	(Not next) (Not previous)
4.	((Michael Red))	Michael Red	

	Connectives	**Models**	**Combined Models**	
1.	And	((John Red)) ((Michael Black))	John Red	Michael Black
2.	Or	((John Red)) ((Michael Black))	John Red	 Michael Black
3.	Ori Inclusive	((John Red)) ((Michael Black))	John Red John Red	 Michael Black Michael Black
4.	Ore Exclusive	((John Red)) ((Michael Black))	John Red Michael Black	(Not Next) (Not Previous)

* The program constructs just one of these two models using a random selection.

Event Mutation and Counterfactual Model Construction

The program selects the most recent antecedent event to mutate. Because the last item in the factual model is the outcome, it selects the second last item. The program mutates this item by negating it. If the value of the item is affirmative it assigns 'not' to the value; if it

is negative it removes the 'not'. For example, the program will change 'Michael Black' to 'Michael not Black'. The program generates a new model, Counterfact-Model, using this mutated event. It takes the Fact-Model and it replaces the mutated event with its negation, e.g., it replaces 'Michael Black' with 'Michael not Black'. For example given the Fact-Model:

Fact-Model: John Red Michael Black Outcome Lose

and the mutated item 'Michael not Black', the program produces the following model:

Counterfact-Model: John Red Michael not Black Outcome Lose

Note that the outcome is still recorded in this model as a loss, until the next suite of functions carry out their work.

Inspection of the Context of Possibilities

The program compares the newly constructed putative Counterfact-Model to the Possibility-Models, in this example, the winning conditions. It treats the values as binary, so for example, it assumes that the value 'Red' is equal to the value 'not Black'. It attempts to find a match between the antecedents of one of the models in the Possibility-Models and the newly constructed counterfactual model. If it succeeds, it replaces the outcome in the new Counterfact-Model with the outcome of the model from the matched Possibility-Model. In our example, the antecedents in the Counterfact-Model match the following Possibility-Model:

John Red Michael Red Outcome Win

and so the newly constructed Counterfact-Model is updated as follows:

Counterfactual: John Red Michael not Black Outcome Win

The program prints the following counterfactual conditional as its final output:

If it had been the case that: Michael not Black

then it would have been the case that: Outcome Win.

If the outcome of the Counterfact-Model is the same as the factual outcome, then the program has in effect generated a semi-factual 'even if' alternative, that is, an alternative with a different antecedent that nonetheless leads to the same outcome (Moreno-Rios, Madruga, & Byrne, 2001; McCloy & Byrne, 1999; in press).

Evaluation of the TempFocus Model

The TempFocus model simulates the temporal order effect in counterfactual thinking. It takes as input a factual scenario in which a number of causally independent events lead to a positive or negative outcome, and it produces as output a counterfactual scenario in which the last event is mutated to lead to a new outcome. We have illustrated how the program works for the card selection scenario but the model is context-independent and therefore can be generalised to many different situations. The program accepts any set of facts and it can generate a counterfactual or semi-factual model provided it receives a conditional which describes the outcome of those facts.

The program encapsulates the primary tenets of the widely held view that the representation of the factual conditions is of critical importance in counterfactual thinking (Legrenzi et al., 1993; Miller & Gunasegaram, 1990). It considers the context of possibilities, in this case provided by the winning conditions, only *after* one of the antecedents has been mutated and a putative counterfactual alternative has been constructed. In the following section, we consider an alternative theory of the temporal order effect based on the context theory. The theory uses the context of possibilities to *select* an event to mutate.

TempContext Model

As described in chapter 1, p17-18, the context theory assumes that people use the facts of a situation to recruit an alternative from the context of possibilities. Items in the recruited alternative which are different in the model of the facts are mutated.

The idea that the context of possibilities may play a role in determining the nature of the counterfactual alternative that is constructed has not been explored systematically before. Our TempContext Model simulates a context theory of the temporal order effect. The theory proposes that in this case, the context of possibilities may be provided by the conditions that determine the outcome - in the card selection scenario these possibilities are the winning conditions. These conditions constrain the possible set of counterfactual alternatives that can be constructed. In this model, the first fact is used to recruit an alternative from the context of possibilities. For example in the card selection scenario, people may use their model of the facts:

John Red Michael Black Lose

to recruit a model from the context of possibilities, which are as follows:

John Red Michael Red Win

John Black Michael Black Win

They may select the first element of the factual model, e.g., John picks red, and find a match for it in the winning models:

John Red Michael Red Win

They may consider only this matching model as a possible counterfactual and conclude that if Michael had picked red they would have won. We suggest that the generation of a counterfactual alternative is driven not only by the 'bottom-up' facts of the actual situation (e.g., Mandel & Lehman, 1996), but also by 'top-down' expectations derived from the context of possibilities (Walsh & Byrne, 2001b).

In the following section, we will describe how the TempContext model simulates this theory of the mental representations and cognitive processes underlying the temporal order effect. The program is similar to TempFocus in its construction of models of the facts and the context of possibilities. However, unlike TempFocus, which only calls on its representation of the context of possibilities to verify the viability of its newly constructed counterfactual model, TempContext uses the context of possibilities from the outset to compare to the facts, in order to construct the model.

A Second Computational Model of the Temporal Order Effect

TempContext simulates the primary tenets of the context theory of the temporal order effect in counterfactual thinking. The program is again written in LISP and a full listing is provided in Appendix 2. Like TempFocus, it takes as input a set of facts (e.g., John picked red and Michael picked black) and the context of possibilities provided by the conditions which determine the outcome of the scenario (e.g., if they both pick red or both pick black they win). For any given set of facts and conditions, it generates a counterfactual alternative about how the outcome could have turned out differently (e.g., they would have won if Michael had picked red). The steps in the program are described in Table 2.3 and Figure 2.2.

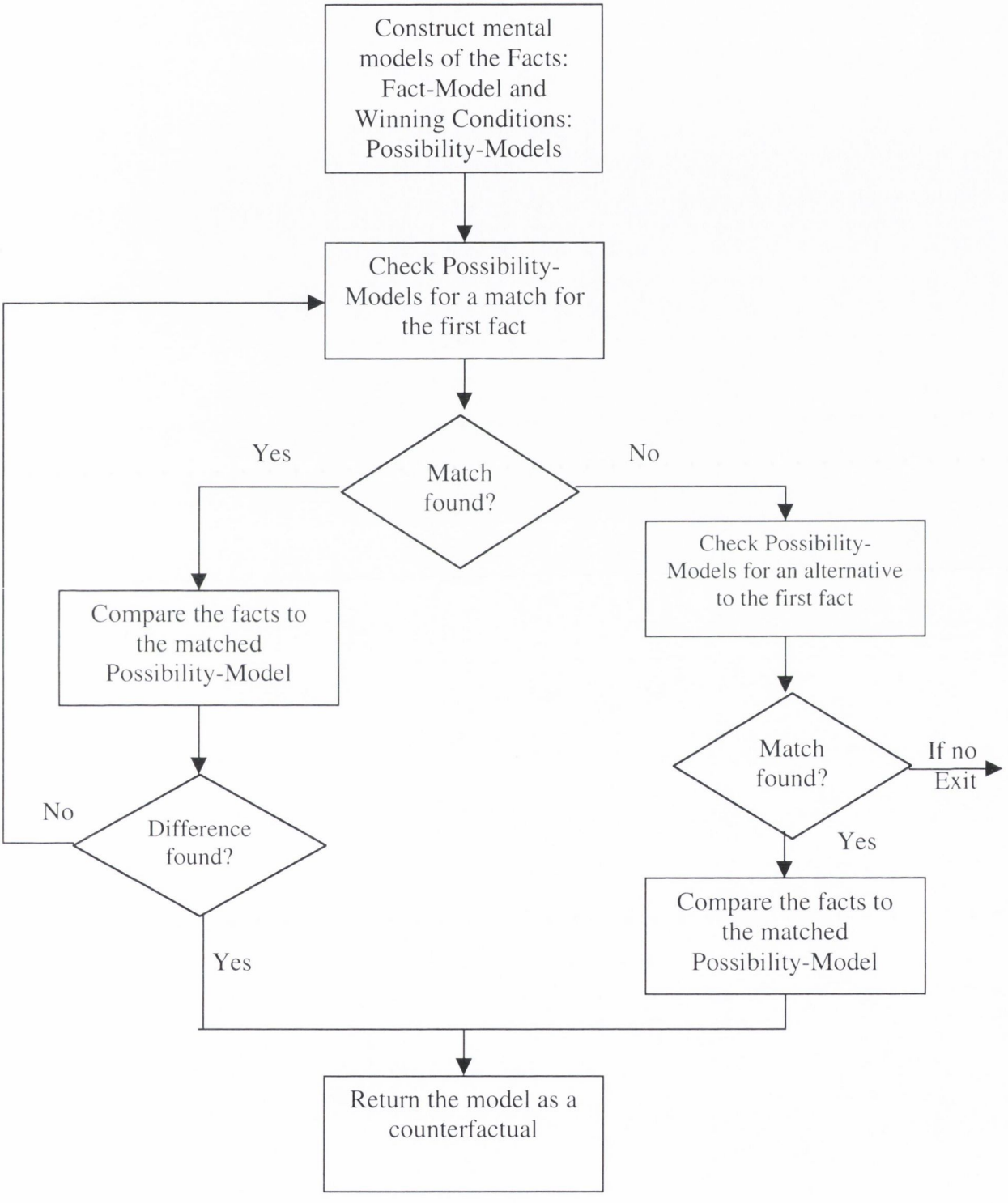
The program consists of three main suites of functions, that comprise functions to represent the factual events and the context of possibilities, functions to match the facts to the context of possibilities, and functions to construct a new counterfactual model. The top-level function is called GenerateCounterfactual and it calls each of the three main sets of functions and it prints out the final counterfactual conditional.

Table 2.3 Outline of the steps in TempContext

1. Construct a set of models to represent the context of possibilities (Possibility-Models)
2. Construct a model to represent the facts (Fact-Model)
3. Place the Possibility-Models in a list.
4. If there are Possibility-Models on the list, search the first model for a match for the first fact.
   - 4.1 If a match is found,
     - 4.1.1 See if the Possibility-Model contains footnotes
       - 4.1.1.1 If so, fleshout the model to be fully explicit
       - 4.1.1.2 Otherwise, go to next step 4.1.2.
     - 4.1.2 Count the number of antecedents in the Possibility-Model which are not contained in the Fact-Model.
       - 4.1.2.1 If there is no difference, the model is factual. Remove the Possibility Model from the list and start at no. 4 again with the reduced list.
       - 4.1.2.2 If the difference is more than 1, put the Possibility-Model onto a stack. Remove it from the list and start at no. 4 again with the reduced list.*
       - 4.1.2.3 If the difference is 1,
         - 4.1.2.3.1 then select the items in the Possibility-Model which are not in the Fact-Model. Create a new counterfactual model by mutating the facts to match these items. Exit the program.
   - 4.2 If no match is found, remove the Possibility-Model from the list and repeat no. 4.
5. If the Possibility-Model list is empty, see if there is a Possibility-Model on the stack.
   - 5.1 If one is found, go to 4.1.2.3.1
   - 5.2 If not, place the set of Possibility-Models on the list again (no. 2)
6. Search the first model for an item which has the same name as the first fact (regardless of the value).
   - 6.1 As 4.1 except replace repeat no.4 with repeat no. 6
   - 6.2 As 4.2 except replace repeat no.4 with repeat no. 6

* The program attempts to find a Possibility-Model with just one antecedent that differs from the Fact-Model to ensure that minimal changes are made in the construction of a counterfactual. Only if none is found, will it use one which differs by more than 1 event.

Figure 2.2 A Flowchart of the steps in the TempContext Model [2]



[2] The squares refer to processes and the diamonds refer to decision points.

We will illustrate how the program works with reference to the description of the card scenario described earlier, and we will show how the program simulates the temporal order effect. Then we will show how it processes other sorts of descriptions, and how these lead to the reversal and elimination of the temporal order effect.

Representing the Facts and the Counterfactual Context

The representation of events in this program is the same as in the TempFocus Model. For example, for the following facts:

> 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.

the program constructs the Fact-Model:

Factual: John Red Michael Black Outcome Lose

The program also takes as input a set of winning conditions:

> If the two cards they pick are of the same colour (i.e., both red or both black), each individual wins £1,000. Otherwise, neither individual wins anything.

which it uses to construct the following Possibility-Models:

John Red Michael Red Outcome Win

John Black Michael Black Outcome Win

The models that the program constructs to represent the context of possibilities for different types of input are exactly those indicated in Table 2.2 for the TempFocus model.

Matching the Facts and the Context of Possibilities

From this point, TempContext operates differently from TempFocus. The program attempts to match its representation of the facts in its Fact-Model e.g.:

John Red Michael Black Outcome Lose

to the Possibility-Models, that is, the models it has constructed of the winning conditions, e.g.:

John Red Michael Red Outcome Win

John Black Michael Black Outcome Win

It selects the first fact from the Fact-Model, John picked a red card, and it searches for a match in the Possibility-Models. It takes one Possibility-Model at a time and compares the fact to each item in the model. If none is found, then it goes on to search the next Possibility-Model. In this example, a match is found in one of the Possibility-Models:

John Red Michael Red Outcome Win

If a match to the first fact is not found, then the program searches through the models for an item which has the same name as the first fact, regardless of the value. For example, it looks for a model which contains, John with any value. We will return to this tactic in the next section. When a match is found, the program fleshes out the Possibility-Model to be fully explicit if necessary.

The program then compares the Fact-Model and the selected Possibility-Model. If there is no difference, then the Possibility-Model is factual (e.g., in the case where the model of the facts represents John picking black and Michael picking black and hence the outcome is a win). In that case, the program continues to look for another Possibility-Model, which it will use to generate a semi-factual conditional, such as 'even if Michael had picked red and John had picked red, they would still have won'.

If there is more than one item in the selected Possibility-Model which differs from the Fact-Model, it records that this model is a potential counterfactual to use, but it continues to inspect the Possibility-Models. It does so because it attempts to find a model in the Possibility-Models which is most similar to the Fact-Model (i.e., which has just one event that differs between it and the Fact-Model). In this way, the program ensures that minimal changes are made to construct the counterfactual alternative (see e.g., Byrne, 1997; Pollack, 1986). Only if there are no models in the Possibility-Models which differ in just one event, does the program use a model which differs in more than one event.

In the current example, the selected Possibility-Model:

John Red Michael Red Outcome Win

contains only one event which differs from the Fact-Model. Hence, it uses the model to generate a counterfactual alternative.

Generating a Counterfactual Alternative

Once a counterfactual model is found, the program identifies which events in the model differ from the Fact-Model (in the current example, the second event differs, Michael picks black in the Fact-Model and red in the Possibility-Model). This is the item which will be mutated. The program then generates a new model, Counterfact-Model by changing this item in the Fact-Model to match the item contained in the Possibility-Model. It describes the newly constructed Counterfact-Model as a counterfactual conditional:

If it had been the case that: Michael had picked a red card

then it would have been the case that: they would have won

The program simulates the temporal order effect, that is, it mutates the second event, when it is given the scenario in the example above, which is typical of those used in such

studies. However, the program also produces a novel reversal of the temporal order effect when it is given certain descriptions, as we will show.

Evaluation of TempContext

We have illustrated how the program works for the card selection scenario. Like the TempFocus model, TempContext may be generalised to other situations which include a set of facts and a conditional that describes the outcome of those facts. TempContext also simulates the temporal order effect in counterfactual thinking. It takes as input a factual scenario and a set of alternative possibilities in the form of the conditions which determine the outcome and it uses these to generate a counterfactual model.

The program differs from TempFocus in one important respect. For any set of independent events, TempFocus mutates the most recent event, regardless of the representation of the context of possibilities. But, this is not the case for TempContext. It uses the facts to recruit a counterfactual model from the context of possibilities. Hence, the program is sensitive to the description of the context of possibilities. It constructs models to represent the context of possibilities but depending on how they are described, some information in the models is made explicit and some information is left implicit. The representation of these models will influence their accessibility. If the models are fully explicit, then each of the models will be accessible. However if some of the models are not fully explicit, then they may be inaccessible. As a result, for some descriptions of the context of possibilities, the program produces the standard temporal order effect (it undoes the second event). However for other sorts of descriptions, the program produces a reversal of the standard effect (it undoes the first event), as we will now show.

TempContext: Simulation of the Temporal Order Effect and its Reversal

The program simulates the temporal order effect, that is, it mutates the second event, when it is given the standard card selection scenario described earlier. We also tested the performance of the TempContext model on novel descriptions, for example, where the individuals have to pick *different* colour cards to win:

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

and where the facts were that John picked a black card and Michael picked a black card and so they both lost. We will show how the program deals with these 'different-card' winning conditions.

We described these 'different-card' winning conditions in several ways, using a disjunction such as the one above that referred to a *different* card (red) from the one that the first player picked (black), or one that referred to the *same* color card, e.g.,:

> If one or the other but not both pick a card from a <u>black</u> suit, each individual wins £1,000,

We also described the 'different-card' winning conditions in a conjunction, e.g.,

> 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.

How the program deals with these descriptions because it reveals important aspects about the nature of the representation of the winning conditions and the effects of different descriptions on these representations. These aspects lead to a novel performance of the program on these tasks and to novel predictions about human performance.

As shown in Table 2.4, given the same facts and the same winning conditions, the computer program produces a very different output, depending on the nature of their description. One sort of description – the different-card 'red' description -- leads to a

reversal of the standard temporal order effect (the program undoes the first event not the second). For the winning conditions:

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

the program constructs two Possibility-Models:

John Red	(NotNext)	Outcome Win
(NotPrevious)	Michael Red	Outcome Win

For the facts:

> John picks a black card and Michael picks a black card and they lose

the program also produces the Fact-Model:

Factual:	John black	Michael Black	Outcome Lose

The program searches for a match for the first fact, *John Black*, in the Possibility-Models. When the Possibility-Models do not contain an explicit match for the first fact, the program selects instead a model which contains the negation of the first fact, John picks not-black (which in the binary context of the colour card game, the program recognises as red). In this case, it finds the model:

John Red	(NotNext)	Outcome Win

which it fleshes out to generate the explicit Possibility-Model:

John Red	Michael not-Red	Outcome Win

The program establishes that it is the first event which is different in the Fact-Model and Possibility-Model. As a result, this event is mutated. It generates a new model, Counterfact-Model by replacing John Black in the Fact-Model with John Red. It produces the counterfactual conditional:

> If it had been the case that: (John Red)
>
> then it would have been the case that: (Outcome Win)

Another sort of description – the different-card 'black' description -- leads to the observation of the standard temporal order effect (the program undoes the second event). For the winning conditions:

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

the program constructs two Possibility-Models:

John Black (NotNext) Outcome Win

(NotPrevious) Michael Black Outcome Win

The facts are the same as in the previous simulation and the program produces the Fact-Model:

Factual: John black Michael Black Outcome Lose

The program again searches for a match for the first fact, *John Black*, in the Possibility-Models. In this case, one of the Possibility-Models contains an explicit match for the first fact, and the program selects this model:

John Black (NotNext) Outcome Win

which it fleshes out the model to generate the explicit Possibility-Model:

John Black Michael not-Black Outcome Win

Now, the second event is different in the Fact-Model and Possibility-Model and it is mutated. The program generates a new model, Counterfact-Model by replacing *Michael Black* in the Fact-Model with *Michael not Black*. It produces the counterfactual conditional:

If it had been the case that: (Michael not Black)

then it would have been the case that: (Outcome Win)

The final sort of description – the different-card 'black and red' description -- leads to the elimination of the standard temporal order effect (the program undoes either the first or the second event). For these winning conditions:

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

the program constructs one of the following set of models at random:

John Black Michael Red Win

. . .

or else:

John Red Michael Black Win

. . .

relying on a random number generator to determine which of these two models it constructs. In this way it simulates the indeterminacy that some people may construct the first model whereas others may construct the second model. When it constructs the first model, the program finds an explicit match for the first fact, *John Black*, in the Possibility-Model. Because the second event is different in the Fact-Model and the Possibility-Model, it is mutated. However, when the program constructs the second model it finds only an explicit alternative to the first event, *John Red,* and hence it mutates it. In this way, the program simulates the elimination of the temporal order effect, by mimicking the two opposing trends as an indeterminacy.

Table 2.4 Simulations of TempContext using the card selection scenario

Simulation of the reversal of the temporal order effect

Winning Conditions: If one or the other but not both pick a card from a red suit, each wins £1,000

Possibility-Models:	John Red	(NotNext)	Outcome Win
	(NotPrevious)	Michael Red	Outcome Win
Factual:	John black	Michael Black	Outcome Lose

No match is found for *John Black* in the Possibility-Models so it looks for an alternative:

Selected Possibility-Model:	John Red	(NotNext)	Outcome Win
Fleshed-out Possibility-Model:	John Red	Michael not Red	Outcome Win

The first event, John Red, differs from the Fact-Model so it is mutated.

Simulation of the production of the temporal order effect

Winning Conditions: If one or the other but not both pick a card from a black suit, each wins £1,000

Possibility-Models:	John Black	(NotNext)	Outcome Win
	(NotPrevious)	Michael Black	Outcome Win
Factual:	John black	Michael Black	Outcome Lose

Match is found for *John Black* in the Possibility-Models.

Selected Possibility-Model:	John Black	(NotNext)	Outcome Win
Fleshed-out Possibility-Model:	John Black	Michael not Black	Outcome Win

The second event, Michael Black, differs from the Fact-Model so it is mutated.

Simulation of an elimination of the temporal order effect

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

Possibility-Models:	John Red	Michael Black	Outcome Win
or:			
	John Black	Michael Red	Outcome Win

If the program produces the first Possibility-Model, then it runs as follows:

Possibility-Models:	John Red	Michael Black	Outcome Win
Factual:	John black	Michael Black	Outcome Lose

No match is found for *John Black* in the Possibility-Models so it looks for an alternative:

Selected Possibility-Model:	John Red	Michael Black	Outcome Win

The first event, John Red, differs from the Fact-Model so it is mutated.

If the program produces the second Possibility-Model, then it runs as follows:

Possibility-Models:	John Black	Michael Red	Outcome Win
Factual:	John Black	Michael Black	Outcome Lose

A match is found for *John Black* in the Possibility-Models:

Selected Possibility-Model:	John Red	Michael Black	Outcome Win

The second event, Michael Black, differs from the Fact-Model so it is mutated.

Overall, the TempContext model simulates the standard temporal order effect and it also makes novel predictions. It makes the novel prediction that given the same facts and the same winning conditions, people should mutate different events depending on how the winning conditions are described. It should be possible to not only eliminate the temporal order effect but also to reverse it. There have been no previous demonstrations of a reversal of the temporal order effect, that is, there is no experimental evidence that people ever undo the first event in an independent sequence of events. We report the results of three experiments that test these predictions in the next chapter.

General Discussion

We have described two theories of the mental representations and cognitive processes underlying the temporal order effect. The focus theory is based on the assumption that people build a model to represent the facts of a situation and they focus on items which are salient or which are the object of attention. Because of their salience, these items are likely candidates for mutation. The TempFocus model assumes that people focus on the last antecedent event and they mutate it.

The context model is based on the assumption that when people understand a situation they build a model to represent the facts but they also use the factual events to recruit alternative possibilities. The TempContext model assumes that these alternatives may be provided by the representation of the conditions which determine the outcome. The theory also assumes that the first fact is used to recruit an alternative model.

We have developed the first computational models of counterfactual thinking to simulate these two theories of the temporal order effect. These models allow us to develop a coherent and well-specified account of the processes underlying the theories. Both of the models simulate the temporal order effect for the standard card selection

game which is typically used to study this effect. The theories also make predictions for scenarios which have not previously been studied and in some cases the two models make different predictions about what people will mutate. The TempFocus model produces the standard temporal order effect regardless of the representation of the context of possibilities, that is, regardless of the representation of the winning conditions. In contrast, the TempContext model depends on the context of possibilities to find an alternative to the facts. The description of the context of possibilities influences the way that the alternatives are mentally represented and hence how accessible they are. We have described examples of scenarios where the theories give rise to different predictions. In the next chapter, we will describe empirical studies which test these predictions.

Chapter 3 An Experimental Study of the Temporal Order Effect in Counterfactual Thinking

Our aim in this chapter is to describe an empirical investigation of the temporal order effect in counterfactual thinking, that is, the tendency for people to mentally undo the most recent event in an independent sequence. The finding is robust (Byrne et al., 2000; Miller & Gunasegaram, 1990; Spellman, 1997) and has been demonstrated in a number of contexts (Miller & Gunasegaram, 1990; Sherman & McConnell, 1996). In Chapter 2, we described two computational models which simulate alternative theories of the mental representations and cognitive processes underlying the temporal order effect. In this chapter, our aim is to report a series of experiments designed to test those theories.

The focus theory of counterfactual thinking proposes that when people generate counterfactuals, they focus on the facts of the situation (e.g., Legrenzi et al., 1993). They select one of the facts and they mentally undo it or mutate it in some way. A fact will be more mutable if it is salient. If the mutated fact together with all of the remaining facts of the scenario change the outcome, then it can be described by a counterfactual conditional. In chapter 2, we developed this theory to explain the temporal order effect. We suggested that in this case, people focus on the last antecedent event. They may negate this event, add it to the remaining aspects of the scenario and then use their knowledge to determine the outcome of the scenario.

The context theory provides an alternative theory of the processes underlying counterfactual thinking (Kahneman & Miller, 1986). When people encounter a new situation, they may think not only about the facts, but they may also recall related situations from past experience. These situations may provide or allow people to construct counterfactual alternatives to the facts. Items which differ in the factual and the

recalled situations are mutated. We have also developed this theory to explain the temporal order effect. The theory proposes that the background conditions to a scenario may provide alternatives to the facts. When people are given a set of conditions which determine the outcome of a scenario, they may construct models of the different possibilities which are consistent with these conditions and they may subsequently use some of these alternatives to generate a counterfactual situation. The theory proposes that people may use the first fact in a sequence to select from these alternatives and items in the selected alternative which are different from the facts will be mutated.

Four main tenets distinguish the two theories. First, the focus theory assumes that some events are more salient or easier to mutate whereas the context theory does not. For example in the temporal order effect, the most recent event may be the most salient. Second, the focus theory emphasises the search for an alternative to one event, whereas the context theory emphasises the search for a similar events. For example, for the temporal order effect, the focus theory is based on the construction of an alternative to the second event, whereas the context theory is based on the search for a match for the first event. Third, the focus theory consults any alternative possibilities, which we term the context of possibilities, *after* an event is mutated to check whether the mutation changes the outcome. In contrast, the context theory uses these alternatives to *select* which event to mutate. Finally, the context theory proposes that the representation of the context of possibilities may influence how accessible they are from the facts and hence may influence the availability of different counterfactual alternatives. In contrast, the focus theory does not use the context of possibilities to mutate an event and so it makes no such predictions.

Mental Representation of the Context of Possibilities

The three experiments that we report test predictions of the two theories about the representation of the context of possibilities. The mental models people construct only represent some information explicitly (Johnson-Laird & Byrne, 1991). The same situation may be represented differently depending on how it is described. For example, the following disjunction in the card game scenario:

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

is consistent with the following fully explicit set of models:

John	Black	Michael	Red	Win
John	Red	Michael	Black	Win
John	Black	Michael	Black	Lose
John	Red	Michael	Red	Lose

However, people usually have difficulty keeping all of these possibilities in mind. Instead they may construct the following initial set of models:

John	black			Win
		Michael	black	Win

The same disjunction with a different description:

> If one or the other but not both picks a card from a red suit, each individual wins £1,000' is represented in the following initial set of models:

is consistent with the same possibilities as the previous disjunction but may be represented differently as follows:

		Michael	red	Win
John	red			Win

There is considerable evidence in the literature on deductive reasoning that people construct mental models of this nature (e.g., Johnson-Laird & Byrne, 1991). The representation of the possibilities may influence their accessibility and hence the generation of counterfactual thoughts. The context theory proposes that people search for a model in the context of possibilities which contains a match for the first fact (e.g., John picked black). But if a match is not explicit, as in the second set of models, people may fail to find it. If no model with a match to the first fact is found, people may look for a model which contains an alternative to the first fact. If the first fact is that John picks a black card then given the black card disjunction, the theory predicts that people will find a match in the first model and they will mutate the second event, whereas given the red card disjunction, it predicts that people will find a match in the second model and they will mutate the first fact.

Our experiments in this chapter address three key questions about counterfactual thinking: the first question is whether the context of possibilities plays a role in the selection of an event for mutation and the second question is whether the description of the context of possibilities can influence the accessibility of those alternatives. These questions distinguish the two theories that we have outlined. The focus theory predicts that the answer to each of these questions is 'no', whereas the context theory predicts that the answer is 'yes'. The third question addresses how alternatives to the facts might be recruited. The context theory predicts that this depends on a search for a similar sequence of events rather than relying on the availability of an explicit alternative to a single event (Byrne et al., 2000). In the following section, we describe three experiments that test these questions.

Three Experiments on the Temporal Order Effect

The experiments that we report were carried out using the colour-card scenario described in the previous chapter, p21. 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 the card that he selects is also black, and the outcome is that both players lose. The context of possibilities provided by the winning conditions was also the same in each of the three experiments. In the four conditions in experiments 1 and 2, both players must pick different cards to win and in the two conditions in experiment 3, both players must pick different cards or both must pick red to win. (We used different card selections as the winning conditions rather than the same card selections as used previously because it gave us greater flexibility in varying the way the winning conditions were represented.) We held constant the facts and the context of possibilities but we varied the *description* of the context of possibilities. These descriptions are outlined in Table 3.1, together with the sorts of mental models that we propose people rely on to represent them.

Our aim in these experiments was to test whether people use these winning conditions to select an event to mutate and whether the mental representation of the winning conditions could influence the accessibility of the different possibilities (Walsh & Byrne, 2001c). If this is the case, as the context model predicts, then people may mutate different events for the different winning conditions described in Table 3.1. As described in chapter 2, the context model of the temporal order effect, TempContext, mutates the second events when there is an explicit match for the first fact, but it mutates the first event if there is only an explicit alternative to the first fact. A reversal of the temporal order effect has never previously been observed. In contrast, the focus model of the temporal order effect mutates the second event in all cases.

Table 3.1 Different descriptions of the winning conditions used in the experiments, and the initial set of mental models that represent them. In each of the three experiments, the facts were the same: John picked black and Michael picked black and so they lost.

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 and red conjunction: predict temporal order effect eliminated (Experiment 1)
   If the two cards they pick are of a different colour (i.e., one from a black suit and one from a red suit), each individual wins £1,000.

John Black	Michael Red	Win
. . .		
OR:		
John Red	Michael Black	Win
. . .		

3. 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

4. Red and black conjunction: predict temporal order effect eliminated (Experiment 2)
   If the two cards they pick are of a different colour (i.e., one from a red suit and one from a black suit), each individual wins £1,000.

John Black	Michael Red	Win
. . .		
OR:		
John Red	Michael Black	Win
. . .		

5. 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

6. Inclusive black and red conjunction: predict temporal order effect eliminated (Experiment 3)
   If both pick a card from a red suit or if the two cards they pick are of different colours (i.e. one from a black suit and one from a red suit), each individual wins £1,000.

John Black	Michael Red	Win
John Red	Michael Red	Win
. . .		
OR:		
John Red	Michael Black	Win
John Red	Michael Red	Win
. . .		

We tested these predictions in the experiments that follow. Participants in the experiments completed a counterfactual mutation task and they also made judgements of guilt and blame. Mutating an event may enhance these attributions (Kahneman & Miller, 1986; Spellman, 1997; Wells & Gavanski, 1989). Previous studies of the temporal order effect have found that these judgements tend to follow the same pattern as event mutability. In the standard card selection scenario, participants mutate the second event and they attribute guilt and blame to the second player (Miller & Gunasegaram, 1990). In the technical hitch scenario, participants mutate the two events with the same frequency and their attributions of guilt and blame to the second player are reduced (Byrne et al., 2000). In the present experiments, we also examined whether guilt and blame judgements would be influenced by changes in the representation of the context of possibilities.

Experiment 1:

Can Different-Card Disjunctions Reverse the Temporal Order Effect?

Our aim in the first experiment in this series was to examine whether the same winning conditions, when described differently, would lead people to mentally mutate different aspects of the factual situation. We constructed two versions of the colour card scenario but we expressed the winning conditions differently in each.

Can the temporal order effect be reversed?

The context theory predicts that it should be possible to bring about a reversal of the temporal order effect by describing the winning conditions in such a way that when people select the first contestant, John, and attempt to match his selection to the winning conditions, they would not find a match. Instead they would find an explicit alternative to John's card selection in the context of possibilities and so they would tend to undo the

first event more. We attempted to bring about such a situation by describing the winning conditions using the disjunction about different coloured cards:

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

In the experiment, the facts were that John picked black and Michael picked black and so they both lost. The exclusive disjunction describing the winning conditions did not refer to the black cards but to red cards. The initial set of models are as follows:

John	Red		Win
		Michael	Red	Win

People think explicitly that the players can win if John picks red or Michael picks red. But the facts are that both players picked black. Hence, people cannot readily match up the fact about the first player, John picked black, with their explicit thoughts about how the players can win, in particular, if John picks red. The context theory predicts that the temporal order effect will be *reversed* given this description. The first factual event (John picks black) is typically immutable, but the models of the winning possibilities contain only an explicit representation of an alternative to this event, John picks red. There is no explicit representation of John picking black so there should be no tendency to maintain John's play as immutable. The availability of an alternative will override the temporal order effect in this condition. The focus theory predicts that the temporal order effect will be *observed* given this description.

Can the same winning conditions eliminate the temporal order effect?

We compared the disjunction to a conjunction:

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

Once again, the facts were that the individuals picked the same colour cards, e.g., John picked black and Michael picked black. The conjunction refers to exactly the same winning conditions as the disjunction. As we outlined earlier, the fully explicit set of models of the winning conditions are as follows:

John Black Michael Red Win

John Red Michael Black Win

If people keep both these models in mind then they should exhibit the standard temporal order effect. However, people find it difficult to keep multiple models in mind (e.g., Johnson-Laird & Byrne, 1991). Instead, we expect that some people may construct the following set of models:

John Black Michael Red Win

. . .

whereas other people may construct the following set of models:

John Red Michael Black Win

. . .

If reasoners construct such initial models then the context theory predicts that the temporal order effect should be *eliminated*: some reasoners will mutate the second event (those who construct the first set of models above), and some reasoners will mutate the first event (those who construct the second set of models above). In contrast, the focus theory again predicts that the temporal order effect will be *observed.*

In the experiment, we gave one group of participants the disjunction and the other group the conjunction. The facts were the same for both conditions: both individuals picked black and they lost. The winning conditions were also the same for both conditions: the individuals could have won if John picked red and Michael picked black, or if John picked black and Michael picked red. The focus theory predicts that the

temporal order effect will be observed in both descriptions of the winning conditions. The context theory predicts that for the disjunction of different cards, the temporal order effect will be reversed, that is, individuals will mutate the first event most often. The initial set of models of the disjunction represents the first individual, John, picking red, which provides an explicit alternative to the fact of John picking black. The context theory also predicts that for the conjunction of different cards the temporal order effect will be eliminated, that is, some people will mutate the first and others will mutate the second event and so there will be no overall tendency for people to mutate one or the other.

Method

Materials and Design

We constructed a scenario based on the colour card scenario (from Byrne et al., 2000). In our scenarios, each player won £1,000 if they picked different rather than the same colour cards. We compared a conjunctive description: 'If the two cards they pick are of a different colour (i.e., one from a black suit and one from a red suit), each individual wins £1,000' to 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 states of affairs, the players could win if John picked black and Michael picked 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 Appendix 3).

Participants completed three tasks, a counterfactual mutation task, and judgements 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 ?

Questions 2 and 3 were forced choice for consistency with previous experiments (Byrne et al., 2000; Miller & Gunasegaram, 1990). They completed the tasks in the fixed order above, on the answer sheet provided. Participants were assigned to one of the two conditions, in a between-participants design.

Participants

The 148 undergraduate students from different departments in the University of Dublin, Trinity College who participated voluntarily in the experiment were 70 women and 77 men and 1 participant did not state their gender. Their ages ranged from 16 years to 37 years with a mean age of 18. Two participants were eliminated from the conjunction condition and three 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 to the conjunction (one black and one red) condition (n = 47) or to the disjunction (one but not both red) condition (n = 96).

Procedure

We tested participants in several large groups. They were given a three-page booklet, see appendix 3. 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-card disjunction reversed the temporal order effect as table 3.2 shows. The results show that for participants who mutated a single event, they exhibited the reverse of the standard temporal order effect when they were given the disjunction (one but not both red), that is, more participants mutated the first event (40%) than the second event (24%), and this difference is reliable (binomial $n = 61$, $z = 1.79$, 1-tailed $p < .04$). This result supports our suggestion that most people keep in mind a set of models that makes some information explicit and leaves other information implicit. Some participants (25%) imagined two counterfactual possibilities suggesting that they kept two fully explicit models in mind. The remaining 11% of participants generated counterfactuals that didn't specify a particular event, for example, 'If only one of the players had picked a red card'.

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 $n = 32$, $z = .18$, $p = .86$). This result for the conjunction of different cards, supports our prediction that most people keep just one explicit model in mind. Two counterfactuals were generated by 17% of participants and the remaining 15% generated a counterfactual that didn't focus on either of the players. The conjunction and disjunction conditions did not differ reliably for those participants who mutated the first or second event only ($chi^2 = 2.04$, $df = 1$, $n = 93$ $p = .15$).[3] The results are consistent with the predictions of the context theory and they do not support the predictions of the focus theory.

[3] The same pattern is observed when we include the first selection of those participants who undid one event after the other. For the disjunction, more participants mutated the first event (64%) than the second event (25%), and this difference is reliable (binomial $n = 85$, $z = 3.90$, $p < .0001$); for the conjunction, as many participants mutated the first event (47%) as the second event (38%, binomial $n = 40$, $z = .47$, $p = .64$); and the conjunction and disjunction did not differ reliably ($chi^2 = 3.42$, $df = 1$, $n = 125$ $p < 0.07$).

Table 3.2: The percentages of mutations and judgements of guilt and blame in Experiment 1

	Red Disjunction One but not Both Red	Conjunction (Black and Red)
Mutations		
First only	40	32
First and then Second *	24	15
Second only	24	36
Second and then First *	1	2
Unspecified	11	15
Guilt		
First	10	0
Second	70	72
Neither	20	28
Blame		
First	62	66
Second	10	6
Neither	28	28

* Refers to the order of mention of counterfactual mutations

Guilt and Blame

The standard temporal order effect occurred in both conditions for judgements 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 conjunction (72% versus 0%, binomial $n = 34$, $z = 5.66$, $p < 0.0001$) and a disjunction (70% versus 10%, binomial $n = 77$, $z = 6.38$, $p < 0.0001$). Participants were no more likely to expect the second player to experience guilt in the conjunction than in the disjunction condition (72% versus 70%). Participants who expected that one player would experience more guilt, were significantly more likely to

expect the first player to experience guilt in the disjunction than in the conjunction condition (10% versus 0%, $chi^2 = 4.85$, df = 1, n = 111 $p < .03$).

A similar pattern emerged for judgements 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 conjunction (66% vs 6%, binomial n = 34, z = 4.63, $p < 0.0001$) and a disjunction (62% versus 10%, binomial n = 69, z = 5.78, $p < 0.0001$). The conjunction and disjunction conditions did not differ reliably (66% versus 62%, $chi^2 = .66$, df = 1, n = 103 $p = .42$).

The experiment provides the first demonstration that the typical temporal order effect can be reversed, that is, participants mutate the first event in the sequence, rather than mutating 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. We suggest that the description contributes to what we have called the context of possibilities, in this case, the mental representation of the winning conditions. It makes some information explicitly available in the mental models that reasoners construct, and renders other information implicit in the representation. An alternative to the first player's choice was made explicitly available in the mental representation (because the winning conditions were framed in terms of picking a red card, and the factual result was that they picked black cards). A corresponding example of the first player's choice in the winning conditions was only implicitly represented. As a result, the presupposition of the first player's selection was undermined by the availability of the alternative, and the temporal order effect was reversed.

The experiment shows a dissociation between mental mutations and judgements 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, 1995b). On the basis of these results we may conjecture that judgements 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, that is, the context of possibilities. We will return to this finding in the general discussion.

The results of this experiment cannot be explained by the focus theory of counterfactual thinking. The first event may be presupposed (Miller & Gunasegaram, 1990) but only if there is an explicit match for it in the context of possibilities. The reversal of the temporal order effect in the red disjunction condition cannot be explained by the availability of an explicit alternative to the first event (Byrne et al., 2000) because there were explicit alternatives to both events. In addition, the result cannot be explained by a change in probability (Spellman, 1997) because there was no change in probability after the first event. However, the results are predicted by the context theory. They suggest that people use the first fact to find a matching model in the context of possibilities. If none is found, they look for an alternative to the first fact and they mutate it. The description of the context of possibilities influences how they are mentally represented and hence how accessible they are from the facts.

Of course, it is possible that the 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.

Other demonstrations of the temporal order effect have occurred under the same constraint. For example, the examination questions scenario we alluded to in the previous chapter, p22 has the constraint that the teacher and student must in effect pick the same questions. The baseball league example has the constraint that the players must perform to the same standard throughout their games. Perhaps the temporal order effect does not occur if the winning conditions require that both players pick a different colour card? In our next experiment, we aimed to rule out this possibility. We also aimed to rule out a second possibility that the reversal arises because of superficial aspects of the linguistic description of the winning conditions, for example, because the disjunction used in the present experiment was difficult to understand.

Experiment 2:

Temporal Order Effects Occur for 'Different-Cards' Scenarios

Our aim in the experiment is to examine whether the temporal order effect can be observed for situations in which the players must pick different cards. Unlike the previous experiment, we described the winning conditions in a way that allowed the first player's selection to be matched to the context of possibilities, that is, the mental representation of the winning conditions. In this experiment we gave 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.

The disjunction has the same linguistic structure as the disjunction used in the first experiment, but in this case, it refers to the black suit, unlike the earlier disjunction which referred to the red suit. The fully explicit models are the same as for the earlier disjunction:

John	Black	Michael	Red	Win
John	Red	Michael	Black	Win

but the initial set of models are different:

John	Black			Win
		Michael	Black	Win

In the first model, 'John picks black' is represented explicitly, but it remains implicit in this model that 'Michael picks red'. People think explicitly that the players can win if John picks black or Michael picks black. The facts are the same as the previous experiment: both players picked black. Hence, people can readily match up the fact about the first player, John picked black, with their explicit thoughts about how the players can win, in particular, if John picks black. The context theory predicts that the temporal order effect should be observed because when reasoners attempt to match John's selection of black with the winning conditions, they will encounter a match. They can flesh out the model to be more explicit, and discover that in this counterfactual alternative, Michael picked red.

We compared this disjunction to a conjunction similar to that used in the first experiment:

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

In this conjunction the order of reference to red and black suits is different from the first experiment, to control for any unforeseen confounding by that order. We expect to replicate the elimination of the temporal order effect for this conjunction, corroborating the suggestion that reasoners keep in mind just a single model, some reasoners keep in mind a model in which the first player picks a red card and the second picks a black card,

whereas other reasoners keep in mind a model in which the first player picks a black card and the second picks a red card.

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 conjunction of antecedents: 'If the two cards they pick are of a different colour (i.e., one from a red suit and one from a black suit), each individual wins £1,000' to a disjunction: 'If one or the other but not both picks a card 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 they did not win the £1000. Participants completed the same three tasks. They were assigned to one of the two conditions, in a between-participants design.

Participants and Procedure

The 152 undergraduate students from different departments in the University of Dublin, Trinity College who participated voluntarily in the experiment were 82 women and 70 men. Their ages ranged from 17 years 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 at random to the conjunction (one red and one black) condition (n = 50) or the disjunction (one but not both black) condition (n = 97). The procedure was the same as in the previous experiment.

Results and Discussion

Counterfactual completions

The different-card 'black' disjunction produces the typical temporal order effect, as table 3.3 shows. The results show that for participants who mutated a single event, they exhibited the standard temporal order effect when they were given the disjunction (one but not both black), 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, 1-tailed p = .06). Of the remaining participants, 21% described two counterfactual alternatives and 16% described a counterfactual that did not specifically refer to either event.

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% versus 34%, binomial n = 36, z = .17, p = .87). Two counterfactuals were generated by 20% of participants and 8% described a counterfactual that did not refer specifically to either event. The conjunction and disjunction conditions did not differ reliably for those participants who mutated the first or second event only (chi^2 = .58, df = 1, n = 97 p = .45).[4]

Guilt and Blame

Once again, the standard temporal order effect occurred in both conditions for judgements 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

[4] Once again, the same pattern is observed when we include the first selection of those participants who undid one event after the other. For the disjunction, more participants mutated the second event (49%) than the first event (35%, binomial n = 82, z = 1.44, 1-tailed p < .075); for the conjunction, as many participants mutated the first event (48%) as the second event (44%, binomial n = 46, z = .15, p = .88); and the conjunction and disjunction did not differ reliably (chi^2 = 1.36, df = 1, n = 128 p = 0.25).

when they were given a disjunction (76% versus 4%, binomial n = 78, z = 7.81, p < 0.0001) and a conjunction (76% versus 4%, binomial n = 40, z = 5.53, p < 0.0001). The conjunction and disjunction conditions did not differ reliably (chi^2 = .0009, df = 1, n = 118 p = .98).

Table 3.3: The percentages of mutations and judgements of guilt and blame in Experiment 2

	Black Disjunction One but not Both Black	Conjunction (Red and Black)
Mutations		
First only	25	34
First and then Second *	10	14
Second only	38	38
Second and then First *	11	6
Unspecified	16	8
Guilt		
First	4	4
Second	76	76
Neither	20	20
Blame		
First	68	74
Second	6	6
Neither	26	20

* Refers to the order of mention of counterfactual mutations

A similar pattern emerged for judgements 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% versus 6%, binomial n = 72, z = 6.95, p < 0.0001) and a conjunction (74% versus 6%, binomial n = 40, z = 5.22, p <

0.0001). The conjunction and disjunction conditions did not differ reliably (chi^2 = .024, df = 1, n = 112 p = .88).

The experiment provides further support for the view that the temporal order effect depends on the context of possibilities. The temporal order effect is observed for the different-card 'black' disjunction, once the description allows the construction of models of the winning conditions that match readily to the facts, in particular the first player's factual situation. In the first experiment, we used the different-card 'red' disjunction:

If one or the other but not both picks a card from a red suit,

each individual wins £1,000.

and we reversed the temporal order effect. Both players picked black, and reasoners mutated the first event, saying if only John had picked red. In the second experiment, we used the same disjunction, but this time referring to a black suit:

If one or the other but not both picks a card from a black suit,

each individual wins £1,000.

and we observed the temporal order effect. Both players picked black, and reasoners mutated the second event, saying if only Michael had picked red. The facts were the same for both experiments; both players picked black. The winning conditions were the same (the contestants would have won if the first picked black and the second red, or vice versa):

John	Black	Michael	Red	Win
John	Red	Michael	Black	Win

The logical form of the description was the same, in that it was an exclusive disjunction. The result therefore ruled out the possibility that the reversal of the temporal order effect in the previous experiment occurred because of a greater difficulty in understanding the

description of the winning conditions. The only difference was in the reference to the colour of the suit, black or red. This small difference of a single word created a large difference in mutation patterns: mutations of the first event versus mutations of the second event. The results again corroborate the predictions of the context theory. The theory proposes that reasoners represent the winning conditions not in a fully explicit set of models but in an initial set of models that makes some information explicit and keeps some implicit. For the disjunction that refers to the red suit, they construct the models:

	Michael Red	Win
John Red		Win

whereas for the disjunction that refers to the black suit, they construct the models:

John Black		Win
	Michael Black	Win

The first set of models leads them to mutate the first event, whereas the second leads them to mutate the second event.

The context of possibilities in the first two experiments is consistent with just two possibilities but in everyday life there may be many alternatives to a situation. In our final experiment, we aimed to replicate and extend the reversal of the temporal order effect to cases where the descriptions of the winning conditions are consistent with more than two models. We described the winning conditions as an inclusive disjunction which is consistent with three possibilities instead of an exclusive disjunction which is consistent with just two. The representation of the inclusive disjunction should place a greater load on working memory and therefore we aimed to see if our results would generalise to these conditions.

Experiment 3:

Inclusive Disjunctions Reverse Temporal Order Effects

Our aim in this experiment was to replicate the reversal of the temporal order effect and to extend it to a set of winning conditions that contained more than two alternatives and therefore placed a greater burden on working memory. In this experiment we studied a scenario which contained the following disjunction in its conditional:

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

This inclusive disjunction, unlike the exclusive disjunctions of the previous experiments, is consistent with three alternative possibilities, which can be represented in the following fully explicit set of models:

John	Red	Michael	Black	Win
John	Black	Michael	Red	Win
John	Red	Michael	Red	Win

In this case, there is only one losing condition, that is, where both players pick black cards. The winning conditions are likely to be mentally represented in a set of models that contains some information represented explicitly, and some information represented implicitly:

John	Red			Win
		Michael	Red	Win
John	Red	Michael	Red	Win

The facts remained the same as in the previous experiment: John picked black and Michael picked black and so they both lost. The context theory predicts that the temporal order effect will be reversed given this description just as it is for the exclusive

disjunction because the context of possibilities does not contain an explicit representation corresponding to the first factual event, John picked black.

We compared this scenario to a second scenario in which the description of the winning conditions was conjunctive and also consistent with three models:

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

The description is similar to the conjunctive description used in the previous experiment, except for the additional condition that both can pick a card from a red suit. It refers to the same states of affairs as the inclusive disjunction above:

John	Red	Michael	Red	Win
John	Black	Michael	Red	Win
John	Red	Michael	Black	Win

But, it should lead to a different representation of those states. The results of the previous experiments indicated that reasoners keep in mind one or other of the two models of the conjunction, and we expect a similar economy of representation with this description. Accordingly we expect that some people will keep in mind the models:

John	Red	Michael	Red	Win
John	Black	Michael	Red	Win

. . .

and others will keep in mind the models:

John	Red	Michael	Red	Win
John	Red	Michael	Black	Win

. . .

Once again reasoners who keep in mind the first of these two sets of models will mutate the second event most, whereas reasoners who keep in mind the second of these two sets of models will mutate the first event most. As a result, the context theory predicts that the temporal order effect should be eliminated once again in this condition. The experiment compared the inclusive disjunction and the conjunction, and we predict the reversal of the temporal order effect for the inclusive disjunction and the elimination of the effect for the conjunction.

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 conditional 'If one or the other or both pick a card from a red suit, 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 colours (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 each 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 155 undergraduate students from different departments in the University of Dublin, Trinity College who participated voluntarily in the experiment were 95 women and 60 men. Their ages ranged from 17 years to 53 years with a mean age of 20. Prior to analysis, one participant was eliminated from the disjunctive condition because he failed to complete all three questions. The remaining participants were assigned to the

conjunctive (one black and one red or both red) condition (n = 69) or to the disjunctive (one or both red) condition (n = 85). 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 for participants who mutated a single event, they exhibited the reverse of the standard temporal order effect when they were given the disjunction (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, 1-tailed $p < .03$). Of the remaining participants, 27% mutated both events and 23% generated a counterfactual that didn't specify a particular event.

The temporal order effect was eliminated when participants were given the conjunction (one black and one red or both red), that is, as many participants mutated the first event (33%) as the second event (41%, binomial n = 51, z = .56, p = .58). Of the remaining participants, 16% mutated both events and 10% mutated neither. As Table 3.4 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 ($chi^2 = 4.33$, df = 1, n = 93 $p < .04$).[5]

[5] Once again, the same pattern is observed when we include the first selection of those participants who undid one event after the other. For the disjunction, more participants mutated the second event (53%) than the first event (24%, binomial n = 65, z = 2.98, $p < .002$); for the conjunction, as many participants mutated the first event (48%) as the second event (42%, binomial n = 62, z = .38, p = .70); and the conjunction and disjunction did not differ reliably ($chi^2 = 3.43$, df = 1, n = 127 p = 0.07).

Table 3.4: The percentages of mutations and judgements of guilt and blame in Experiment 3

	Disjunction One or Both Red	Conjunction Black and Red or Both Red
Mutations		
First only	33	33
First and then Second *	20	15
Second only	17	41
Second and then First *	7	1
Unspecified	23	10
Guilt		
First	17	12
Second	55	77
Neither	28	12
Blame		
First	59	68
Second	14	10
Neither	27	22

* Refers to the order of mention of counterfactual mutations

Guilt and Blame

Once again, the standard temporal order effect occurred in both conditions for judgements 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 conjunction (77% versus 12%, binomial $n = 61$, $z = 5.63$, $p < 0.0001$) and a disjunction (55% versus 17%, binomial $n = 61$, $z = 4.1$, $p < 0.0001$). The conjunction and disjunction conditions did not differ reliably ($chi^2 = 2.0$, $df = 1$, $n = 122$ $p = .16$).

A similar pattern emerged for judgements 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 conjunction (68% versus 10%, binomial n = 54, z = 5.31, $p < 0.0001$) and a disjunction (59% versus 14%, binomial n = 62, z = 4.7, $p < 0.0001$). The conjunction and disjunction conditions did not differ reliably ($chi^2 = .86$, df = 1, n = 116 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 does not refer to the same colour cards as the factual situation. It extends this reversal to inclusive disjunctions as well as exclusive ones. 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 alternatives but also for winning conditions that require multiple models to be held in mind.

General Discussion

In the previous chapter we described two computational models which simulated different theories of the mental representations and cognitive processes that underlie counterfactual thinking. The focus theory proposes that the first event is immutable and therefore people should mutate the second event (Miller & Gunasegaram, 1990). The context theory made the novel prediction that the temporal order effect can be reversed, eliminated, and observed, depending on the description of the context of possibilities, in this case provided by the winning conditions.

In three experiments we corroborated the predictions made by the context theory. In all three experiments, the participants were given the same facts: both players picked black cards. They were also given exactly 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, in the four conditions of Experiments 1 and 2; and in addition, if they both picked red, in the two conditions of Experiment 3. However, we varied the way we *described* the winning conditions. In Experiments 1 and 3, the disjunction referred to the players picking red cards ('If one or the other but not both picks a card from a red suit, each individual wins £1,000'; and 'If one or the other or both picks a card from a red suit, each individual wins £1,000' respectively). The temporal order effect was reversed in both experiments and participants undid the first event most often. In Experiment 2, the disjunction referred to the players picking black cards ('If one or the other but not both picks a card from a black suit, each individual wins £1,000), and the temporal order effect was observed. In all three experiments, a conjunction was also used ('If the two cards they pick are of a different colour (i.e., one from a black suit and one from a red suit), each individual wins £1,000'; 'If the two cards they pick are of a different colour (i.e., one from a red suit and one from a black suit), each individual wins £1,000'; and 'If both pick a card from a red suit or if the two cards they pick are of different colours (i.e. one from a black suit and one from a red suit), each individual wins £1,000' respectively). The conjunctions eliminated the temporal order effect and participants undid the first event as often as the second.

The three experiments support the view that the context of possibilities, that is, the mental representation of the conditions under which the players can win, influences the temporal order effect. We proposed that reasoners represent the winning conditions by constructing models which make some information explicit and leave other information implicit. The winning conditions in the four conditions of Experiment 1 and 2 are:

John	black	Michael	red	Win
John	red	Michael	black	Win

and the winning conditions in the two conditions of experiment 3 also contain a third possibility:

John red Michael red Win

The disjunction in Experiment 1, 'If one or the other but not both picks a card from a red suit, each individual wins £1,000' is represented in the following initial set of models:

Michael red Win

John red Win

The first event in the facts, 'John picked black', does not match anything explicitly represented in this set of models, and instead an explicit alternative is available to it, 'John picks red'. The temporal order effect is reversed and participants undo the first event most often. A similar process occurs in Experiment 3 with the inclusive disjunction, 'If one or the other or both picks a card from a red suit, each individual wins £1,000'

Michael red Win

John red Win

John red Michael red Win

In contrast, the disjunction in Experiment 2, 'If one or the other but not both picks a card from a black suit, each individual wins £1,000', is represented in the following initial set of models:

John black Win

Michael black Win

The first event in the facts, 'John picked black', matches explicitly the information about John's selection in this set of models, and so the model is fleshed out to be explicit and in this model, Michael picks red. The temporal order effect is observed and participants undo the second event most often.

A conjunction was also used in the three experiments, such as, 'If the two cards they pick are of a different colour (i.e., one from a black suit and one from a red suit), each individual wins £1,000'. The winning conditions are:

John	black	Michael	red	Win
John	red	Michael	black	Win

and we propose that some people construct the following initial set of models:

John	black	Michael	red	Win

. . .

whereas others construct the following initial set of models:

John	red	Michael	black	Win

. . .

and as a result of this vagary, the temporal order effect is eliminated.

In chapter 1, we described two broad theoretical approaches which provide an account of the mental representations and cognitive processes underlying counterfactual thinking. In chapter 2, we developed these theories to show how they can explain the temporal order effect and we described two computational models, TempFocus and TempContext, that we developed to simulate these theories. We will now consider how our results address the three questions we posed at the start of this chapter, does the context of possibilities play a role in the selection of an event for mutation, does the description of the context influence the accessibility of those alternatives and does the mutability of an event depend on a search for a similar sequence of events.

First, the results of our experiments support the view that context of possibilities plays a role in the selection of an event for mutation, as predicted by the context theory. By changing the context of possibilities, we have shown that the temporal order effect can be reversed, (so that people mutate the first event more than the second) or eliminated (so

that people mutate both events with equal frequency). The focus theory, which proposes that people search through the facts for a mutable event, cannot account for this finding based on the representation of the facts, as these were the same in each of the experiments.

Second, the mutability of an event may depend in part on the meaning of the context of possibilities which will make different alternatives available. However, our results also show that the description of these conditions and hence what people represent explicitly can influence the accessibility of these alternatives. For example, in our experiments we have shown that a change in the description of the same winning conditions could lead people to mutate the first or second event.

Third, as predicted by the context theory, the results support the view that the generation of a counterfactual depends on a search for a similar sequences of events in the context of possibilities and that factual items which are different in the selected alternative are mutated. The results cannot be explained purely on the basis that some events had explicitly available alternatives. In the red disjunction conditions, there was an explicit alternative to both the first and second events and yet participants mutated the first event more. Also, the black disjunction provides no alternative to the second event and yet most people mutate it in this condition. Overall, the results of our experiments clearly support the context theory over the focus theory at least for the temporal order effect.

All three experiments show a dissociation between mental mutations and judgements 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 results do not support the view that the greater mutability of the second event underlies these judgements (Miller & Gunasegaram, 1990). The attributions of guilt and

blame are consistent with the change in probability was greatest after the second event (Spellman, 1997) and they are also consistent with the conditions under which the players can win and the explicit alternatives to the facts (Byrne et al., 2000). They do not appear to be affected by the nature of the description of, and therefore the mental representation of, the conditions under which the players can win.

To what extent can these findings be generalised to other aspects of counterfactual thinking? The possibility that people may search for episodes which are similar to the facts could be applied to any situation but how this occurs may be different. In our experiments, the situations which made up the context of possibilities, i.e., the winning conditions were explicitly stated. In many everyday situations this is not the case and therefore these possibilities must be recruited from other sources, for example from past experience. What criteria are used to decide what the context of possibilities is for real life events? In the next chapter, we attempt to consider this by examining the mutability of controllable actions.

Chapter 4 The Effect of Reasons for Acting on Counterfactual Thinking

Explanations of counterfactual thinking have tended to emphasise either a search through the facts for a mutable event, which we have termed the focus theory or else a search for an alternative to the facts, which we have termed the context theory. The results of our experiments on the temporal order effect support the view proposed by the context theory that people recruit alternatives to help select an event to mutate. Our aim in this chapter is to examine the extent to which these findings generalise to counterfactual thinking about other types of events. We will examine the mental representations and cognitive processes underlying another important effect in counterfactual thinking, that is the tendency for people to focus on controllable actions more than uncontrollable events (Girotto, Legrenzi & Rizzo, 1991; McCloy & Byrne, 2000). In this chapter, we also aim to extend our study in two further new ways. First, we aim to study antecedent events which are causally related, (whereas the antecedent events studied in the previous series of experiments, i.e., the card selections, were independent). Second, in contrast to the previous chapter where the context of possibilities was explicitly presented in the scenario, we aim to study scenarios for which people must access the possibilities from their background knowledge or experience. In everyday life, the context of possibilities may be provided by knowledge in memory about sequences of typical actions and events that occur in situations (e.g., Baddeley, 1999) or by our knowledge of the different possibilities that are consistent with a causal relation (e.g., Goldvarg & Johnson-Laird, 2001). We propose that *reasons for acting* may form the context in which actions are performed and that different reasons for acting may make different counterfactual possibilities available. If people use these alternatives to generate a counterfactual, then

their availability may influence the mutability of controllable events. We carried out a series of three experiments to address this issue.

The Controllability Effect

Evidence for the greater mutability of controllable actions comes from a study based on a scenario in which a man arrived home unusually late to find his wife dying from a heart attack and he was too late to save her (Girotto et al., 1991). The scenario contained a causal sequence of four events which delayed the man on his journey home. One of the events was controllable, the decision to drink a beer in a bar, whereas the other three were beyond the man's control, for example, a flock of sheep crossing the road. The results showed that people tended to change the controllable action, regardless of its position in the causal chain. Controllable actions are undone more in real life in the counterfactual thoughts of individuals who have suffered a traumatic life event (Davis, Lehman, Wortman, Silver & Thompson, 1995) and in games (Markman, Gavanski, Sherman & McMullen, 1995). Controllable actions are undone more whether they are normal or exceptional (Girotto et al., 1991) or based on traits or not (N'gbala & Branscombe, 1995), and even when the controllable actions are not perceived as causal (Mandel & Lehman, 1996).

As discussed in chapter 1, counterfactual thinking impacts on many aspects of cognition and emotion. The focus on actions may play an important role in enhancing perceived control and in helping people to learn from past mistakes and to develop intentions for the future (Markman, Gavanski, Sherman & McMullen, 1993;1995; Nasco & Marsh, 1999; Roese, 1994). Counterfactual thinking about past actions has also been associated with emotions such as guilt (Niedenthal, Tangney & Gavanski, 1994) and regret (Byrne & McEleney, 2000; Landman, 1987).

Can the focus theory or the context theory explain the controllability effect?

Focus Theory

According to the focus theory, the actions of the central character in a scenario are more salient than other events (Legrenzi, Girotto & Johnson-Laird, 1993). People may build models which are based on the actions of the protagonist perhaps because many narratives are organised around the protagonist (Morrow, Bower & Greenspan, 1989) and they may subsequently mutate these events as they are explicitly represented in their models (Legrenzi et al., 1993).

Context Theory

Alternatively, the mutability of controllable events may depend on the availability of alternatives. There are hints in the literature to support this view. For example, alternatives to actions may be represented explicitly in mental models (Byrne, 1997; Byrne & McEleney, 2000). There is also evidence to suggest that the mutability of actions may be dependent on the context. For example, exceptional actions are more mutable than routine ones (Kahneman & Tversky, 1982b) and socially acceptable actions are more mutable than unacceptable ones (McCloy & Byrne, 2000). Social rules and habits reduce the choice of possible actions and hence may constrain their mutability (Wells, Taylor & Turtle, 1987). But social rules and habits form just a subset of possible reasons for acting. We propose that actions are perceived in the larger context of reasons for acting and we aim to provide the first systematic investigation of the effect of reasons for acting on counterfactual thinking.

There is a long philosophical discussion comparing reasons for acting with other causes that occur in the world (e.g., Davidson, 1963; Melden, 1961). Our understanding

of causal relations and their effect on counterfactual thinking may therefore provide a starting point for the study of reasons for acting. We will first examine evidence to suggest that causes may constrain the mutability of events. We will then consider the nature of reasons for acting and we will compare the mental representation of different types of reasons and causes. People's models of reasons and causes may represent what is possible (Goldvarg & Johnson-Laird, 2001) and we suggest that these models may form the context of possibilities against which the facts are perceived. The context theory proposes that the mutability of actions is dependent on the context of possibilities and therefore leads to novel predictions about the effect of reasons for acting on the mutability of actions. In contrast, the focus theory assumes that the mutability of actions is based on their salience and therefore should be independent of the reasons that lead to them.

Are reasons for actions like causes for effects?

Previous research has shown that people tend to undo causes rather than their effects (Wells, Taylor & Turtle, 1987), perhaps because once the cause is known to have occurred its subsequent effects seem inevitable. Consider a man who is burned after spraying charcoal lighter fluid on an open flame. It is more likely that people will alter the cause, "if only he hadn't sprayed the lighter fuel on the flame", than the effect, "if only he hadn't been burned" (Wells, et al., 1987). Evidence for the greater mutability of causes than effects comes from a study based on a scenario in which a man is impeded by four obstacles on his way to a sale to buy a stereo (Wells et al., 1987). As a result, he arrives too late and the last stereo has just been sold. In the scenario, the obstacles were causally linked, so that each event causally affected subsequent events in the chain (and the order of the obstacles was controlled for). The results showed that participants tended to change the first cause in their counterfactual thoughts.

Controllable actions may be mutable because they are perceived to be independent of external causes. Controllable actions usually occur when the situation allows a choice of possible actions. Actions which are constrained by events beyond the individual's intention, such as stopping the car to cure an asthma attack, are mutated less often than free actions, such as going to a bar for a drink (Girotto et al., 1991). Because controllable actions do not have easily identifiable external causes, they tend to be more mutable than uncontrollable events (Girotto et al., 1991). However, we suggest that although controllable actions may be perceived as independent of external causes, people act for a *reason*. The key question which we want to address is whether reasons constrain the mutability of controllable actions in the same way that causes constrain the mutability of their effects.

Reasons for Acting

Controllable actions may be events which depend on a decision of an individual (Girotto et al., 1991), or which are intentional (Anscombe, 1963). A single action may be defined as intentional or unintentional under different descriptions. For example, if I flip the light switch, I might also light the room, and I might also alert a burglar to the fact that I am at home. Although all three descriptions refer to a single action, only the first two, flipping the switch and lighting the room, were intentional. Intentional actions may be ones which can be explained by stating the actor's reasons for acting (Anscombe, 1963).

To understand an actor's reasons for an action, it may be necessary to understand their beliefs and desires (e.g., Davidson, 1963). Reasons may encompass a broad spectrum, as the following quotation illustrates:

> "Whenever someone does something for a reason, therefore, he can be characterised as (a) having some sort of pro attitude towards actions of a certain kind....Under (a) are to be included desires, wantings, urges, promptings, and a great variety of moral views, aesthetic principles, economic prejudices, social conventions, and public and private goals and values in so far as these can be interpreted as attitudes of an agent directed toward actions of a certain kind. The word 'attitude' does yeoman service here, for it must cover not only permanent character traits that show themselves for a lifetime of behavior, like love of children or a taste for loud company, but also the most passing fancy that prompts a unique action, like a sudden desire to touch a woman's elbow".
>
> (Davidson, 1963, pp. 685-6).

A reason may explain or rationalise an action when an actor has a desire and a belief that the action has the desired property (Davidson, 1963). Consider an individual, James, who carries out an action, going to church, for a particular reason, to please his mother. For this reason to fully explain his action, it is essential to know what property of the action appeals to James, i.e., whether he enjoys pleasing his mother, thinks it is right to do so, or thinks it is his obligation to do so. In addition, a reason may only explain an action if it causes the action (Davidson, 1963). For example, an individual may have a number of beliefs and desires which could make the action comprehensible but only one of which actually led the actor to act (Donnellan, 1967; Smith & Jones, 1986). Consider another individual, Jack, who carries out an action, opening the window, and who may have a number of desires which sensibly lead to such an action, such as wanting fresh air, wanting to speak to a friend outside, wanting to lift in a flower-pot from the window-sill outside. Although any one of the desires may make sense of the action, it may not have been the reason that led him to act. For a reason to explain an action it must do more than make the action comprehensible, it must actually cause the actor to act (Davidson, 1963).

It is a matter of some philosophical debate whether reasons may be equated with causal explanations in the external world (Donnellan, 1967; Smith & Jones, 1986). To

infer a cause or a reason may require different processes. For example, Hume considered that causes may require empirical observation, whereas reasons may require logical reasoning without empirical observation (see e.g., Hume, 1938). If an individual believes that a particular action is required to achieve a particular goal, but the individual shows absolutely no tendency to do the action, it may be inferred logically that the individual doesn't desire the goal more than anything else available (Smith & Jones, 1986). But once causes and reasons are provided, it seems psychologically plausible that they are mentally represented in a similar manner (Johnson-Laird & Byrne, 1991).

The Mental Representation of Reasons

The mental representation of reasons may require people to keep in mind alternative possibilities (e.g., Johnson-Laird & Byrne, 1991). Consider the mental representation of causes as an example. According to one view, people may represent a causal relation between events, e.g., 'the vase's fall caused it to break' by keeping in mind the different possibilities with which the assertion is consistent (Johnson-Laird & Byrne, 1991). Their mental models may represent the factual situation, the vase fell and it broke, and they may also keep in mind the counterfactual situation, in which the vase didn't fall and it didn't break (Goldvarg & Johnson-Laird, 2001; Johnson-Laird & Byrne, 1991; McEleney & Byrne, 2001). These two situations represent a strong causal relation, that is, one in which the cause is both necessary and sufficient for the effect:

Vase Fell Vase Broke

Vase Not Fell Vase Not Broke

Where 'vase fell' represents 'the vase fell', 'vase broke' represents 'the vase broke', and as previously discussed, 'not' is a propositional-like tag to indicate negation (see Johnson-Laird, Byrne, and Schaeken, 1992). There are other sorts of causal relation, as Table 4.1

shows. A conditional causal relation, e.g., 'touching the vase caused it to break' requires reasoners to keep in mind the two possibilities outlined earlier and also a third situation in which the vase was not touched but it broke anyway (perhaps because the wind blew it over):

Vase Touched	Vase Broke
Vase Not Touched	Vase Not Broke
Vase Not Touched	Vase Broke

The cause in this case is sufficient for the effect but not necessary. A third sort of causal relation, e.g., 'the vase's fragility caused it to break', requires reasoners to keep in mind the first two situations outlined earlier and a third counterfactual situation in which the vase is fragile but it didn't break (perhaps it remained untouched). The cause in this case is an enabling condition which is necessary but not sufficient for the effect:

Vase Fragile	Vase Broke
Vase Not Fragile	Vase Not Broke
Vase Fragile	Vase Not Broke

We suggest that a similar analysis can be provided for reasons, as Table 4.1 illustrates. Reasoners may represent a reason for an action by constructing a set of models which represents the factual situation, for example, 'Mary is hungry and she eats a large meal', and they may also keep in mind the counterfactual situation, in which Mary is not hungry and she doesn't eats a large meal. A strong reason is necessary and sufficient for its action. A second sort of reason is a conditional reason, that is, one which is sufficient but not necessary. For example, 'Mary's desire to visit a new restaurant' would be a conditional reason for having a meal if it was the case that other reasons would also cause her to have a meal, even if she didn't want to visit the restaurant, for example, she wants to try out a new recipe or she is just hungry. Reasoners must keep in mind the two

possibilities described for a strong reason, and they must keep in mind a third counterfactual situation in which Mary doesn't want to visit the restaurant and she has a meal anyway.

Table 4.1 Mental models of causes and reasons*

	Cause	*Reason*
Strong relation:	**a caused o**	**x is reason for o**
Factual	a o	x o
Counterfactual	not-a not-o	not-x not-o
Conditional relation:	**b caused o**	**y is reason for o**
Factual	b o	y o
Counterfactual	not-b not-o	not-y not-o
	not-b o	not-y o
Enabling relation:	**c caused o**	**z is reason for o**
Factual	c o	z o
Counterfactual	not-c not-o	not-z not-o
	c not-o	z not-o

* (Adapted from Johnson-Laird & Byrne, 1991; Goldvarg & Johnson-Laird, 2001)

A third sort of reason is an enabling reason, that is, one which is not sufficient but is necessary. For example, 'Mary's love of food' would be an enabling reason to eat a meal if it was the case that other reasons are also needed to bring about the outcome, for example, Mary must have food in the house or have money to buy some. Reasoners must again keep in mind the two possibilities described for a strong reason, and they must keep

in mind a third counterfactual situation in which Mary loves food but she doesn't eat it, perhaps because she cannot acquire any -- actions also rely on external conditions and a reason may not be sufficient if there is no opportunity to act. There may be other types of reasons, such as reasons which are neither necessary nor sufficient for an action, but we will focus primarily on the three sorts of reasons outlined here.

We suggest that the representation of reasons may be somewhat different from causes. We conjecture that people interpret causes more often as strong causes, that is, as necessary and sufficient causes. In contrast, they may interpret reasons more often as a conditional or enabling relation. Common sense understanding of human behaviour differs from commonsense understanding of physical causes. The relation between a reason and its action is not as simple as the relation between a cause and its effect. First, actions can occur without reasons: people may act unintentionally (including actions brought about by absent mindedness, mistakes, and so on), whereas effects are unlikely to occur without a cause. Second, reasons often can occur without actions: people may fail to act despite having a very good reason to do so, whereas effects tend to occur when their cause occurs. Third, reasons often can have a one-to-many mapping to their actions: a single reason can lead to various different actions (for example, a desire or obligation can be realised in a number of ways), whereas causes tend to be tied to their effects in a one-to-one mapping. Fourth, reasons can have a many-to-one mapping to their actions: a set of reasons can lead to a single action (and undoing one of the set will not undo the outcome). Finally, the relation between reasons and actions may be less consistent over time than the relation between causes and effects (Hart & Honore, 1959), because people can give priority to different desires at different times and they can change their beliefs. Of course, there are exceptions to these general tendencies, and some reasons may not fit this pattern whereas some causes may fit it, but overall, these differences may help to

distinguish between reasons and causes. Accordingly, the initial mental representation of reasons may keep more alternatives in mind more than the initial representation of causes. Moreover, different reasons may be represented in different ways: some sorts of reasons may be represented as a strong relation between the reason and the action.

Our primary suggestion is that the tendency to imagine these additional possibilities may have implications for the counterfactuals that people construct; there may be more available alternatives to the facts. We propose that these alternatives form the context of possibilities against which the facts are perceived. According to the focus theory these alternatives should not influence the mutability of actions. However, the context theory assumes that people use the facts to recruit one of these alternatives and then use the alternative to select an event to mutate. It therefore makes novel predictions about the effect of reasons for actions on the mutability of events. In particular if we are right that people tend to represent reasons for actions as a conditional or enabling relation, rather than a strong relation, then it is less likely that they will mutate the reason in order to undo the action in their counterfactual thoughts. Because it is easy to imagine the possibility in which a person has a reason and yet doesn't act, mutating the action by itself seems plausible (unlike say, imagining an effect does not occur even though the cause has occurred). Likewise, because it is easy to imagine a person having a reason and not acting on it, mutating the reason alone will seem an uncertain way to undo the action (unlike say, imagining a cause does not occur and being certain that therefore the effect will not occur).

Our aim in the three experiments that we report was to examine the effects of different types of reasons on people's counterfactual thoughts (Walsh & Byrne, 2001d). We aim to establish whether people undo reasons or actions, and whether different types of reasons affect the focus on actions in counterfactual thinking. In the first experiment

we compared the mutability of actions when people were given two common types of reason for acting.

We also aimed to examine the effect of reasons on judgements about emotion. We focused on people's judgements about the self-blame and regret experienced by the actor. Counterfactual thinking may amplify emotions (Kahneman & Miller, 1986). Previous research indicates that self-blame and regret show a similar pattern to counterfactual mutations (Byrne & McEleney, 2000; Davis et al., 1995), but sometimes diverge from it (see Byrne et al., 2000; N'gbala & Branscombe, 1995; 1997). The results of the experiments that we described in chapter 3, showed a dissociation between counterfactual mutations and judgements of guilt and blame. Changes in the representation of the context of possibilities influenced the mutability of events but not emotional attributions. Our aim was to examine whether any changes in the mutability of actions brought about by a change in the context would lead to a corresponding change in judgements of self-blame and regret.

Experiment 4: Own-sake and Plan-based Reasons

Our aim in the first experiment was to compare the effects of different sorts of reasons for acting, an internal immediate reason and an internal enduring reason, on the mutability of actions. An important sort of reason for acting is a desire to act, which may arise because an actor wants to do the action for its own sake (an immediate reason) or because the actor believes the action is a means to bring about a goal that the actor wants (an enduring reason). The two sorts of desires are sometimes known as intrinsic and extrinsic desires (Goldman, 1970; Mele, 1992), although for simplicity we will refer to them here as 'own-sake' reasons and 'plan-based' reasons. Both are real reasons in that the action is intentional (Smith & Jones, 1986). However, it is likely that only plan-based reasons

adequately explain actions: An own-sake reason asserts only that the action appealed to the actor, whereas a plan-based reason specifies some consequence of the action that appeals to the actor (Davidson, 1963). When people ask 'why are you doing that?' they are interested in the actor's purpose for acting (Anscombe, 1963; Goldman, 1970). If you ask a person standing on a tennis court waving a racket around why they are waving the racket around, you may accept as a genuine explanation that they are warming up or getting the feel of the racket or even swatting flies, but not merely that they want to wave the racket around (Shaver, 1985).

In our first experiment we constructed a scenario based on a leisure sports content in which the action of the protagonist (calling in to a gym) had a reason which was either an own-sake reason (he liked going to the gym), a plan-based reason (he wanted to lose weight), or no reason. The aims of our experiment were twofold. The first aim was to examine whether people mutate reasons more than actions, just as they mutate physical causes more than their effects. The mutability of reasons has not been examined experimentally before and it remains unknown whether people mutate reasons more or less than the actions they bring about. We anticipate that people will not mutate reasons more than actions, in keeping with our conjecture that people may represent the relation between reasons and actions as a weak relation. They may keep in mind several possibilities, including the possibilities that the reason doesn't occur and the action occurs anyway, or that the reason occurs and the action doesn't. Hence to undo the action they may focus directly on it, rather than on the reason that led to it.

Our second aim was to compare the effects of different types of reasons on the mutability of actions. On the one hand, the mutability of an action may be reduced when a reason is provided for it. If so, both sorts of reasons, own-sake and plan-based reasons, should reduce the mutability of the action compared to no reason. On the other hand, the

mutability of an action may be reduced only by a reason which adequately explains the action. If so, a plan-based reason should reduce the mutability of the action more than an own-sake reason or no reason.

Our primary measure was the counterfactual thoughts generated by reasoners. In addition, we examined judgements about emotion. We aimed to examine whether reasons for acting would influence ratings of judgements of self-blame and regret.

Method

Materials and design

We constructed a scenario based on one used by Girotto et al. (1991) about an individual, Tom, who is delayed by several events on his way home from work, only to find he is too late to save his wife who is dying from a heart attack. In our scenario, the uncontrollable events were a flat tyre and a traffic jam, and the controllable event was going for a workout in a gym (rather than going for a drink in a bar, which may be a special sort of controllable event, see McCloy & Byrne, 2000).

We constructed three versions of the scenario. In one version the action (going to the gym) was preceded by an own-sake reason, that is, a desire to do the action for its own sake ('Tom really likes to go to the gym'); in a second version, the action was preceded by a plan-based reason, that is, a desire for a consequence of the action ('Tom is trying hard to lose weight'), and in the third version no reason was given, just a filler item ('It is quite a large gym'). The scenario was as follows (with the different sorts of reasons used in different versions in italics):

Tom has been working in a bank in the city centre for the past few years. He lives with his wife on the outskirts of town. On Thursday, he left the office at the usual time. He was only a short distance down the road, when he drove over some glass and got a flat tyre. After changing it, he headed off again. Because of this delay however, he ran into the rush hour traffic which he usually managed to avoid. This meant that his journey was considerably slower than usual. Tom passes a local gym on his way home every day. It is normally closed but because of his delays that day, when he got there it had already opened for the evening.

[Control: It is quite a large gym.]

[Own-sake reason: He really likes to go to the gym.]

[Plan-based reason: He is trying hard to lose weight.]

He decided to stop off for a quick workout. Thirty minutes later he was finished and he set off again. When he arrived home soon after, he found his wife lying on the floor. He realised that she had had a heart attack and that she was dying. He tried to help her, but his efforts were in vain.

Participants completed three tasks, a counterfactual mutation task, and ratings of self-blame and regret, as follows:

1. As commonly happens in such situations, Tom often thought, "if only..." during the days and weeks that followed the accident. Please list in the order of importance, four ways in which he may have completed this thought so that the outcome would be different.

2. On the following scale, please rate the extent to which you think Tom felt to blame for his wife's death:

1 -------- 2-------- 3-------- 4-------- 5-------- 6-------- 7-------- 8-------- 9
not at all to blame — completely to blame

3. On the following scale, please rate the extent to which you think Tom regretted going to the gym:

1 -------- 2-------- 3-------- 4-------- 5-------- 6-------- 7-------- 8-------- 9
no regret — strong regret

They completed their questions in the fixed order above on the answer sheet provided. Participants were assigned to one of the three conditions at random in a between-participant design.

Participants and Procedure

The 194 undergraduate students who participated in the experiment voluntarily were 78 women and 116 men from a number of departments in the University of Dublin, Trinity College. They ranged in age from 17 to 59 years with a mean of 25 years. Prior to analysis, 4 participants were eliminated because they failed to complete all three questions. The remaining 190 participants were randomly assigned to the plan-based reason condition (n = 61), own-sake reason condition (n = 63) and the control condition (n = 66). The participants were tested in large groups. They were given a four-page booklet, see appendix 4. The first page contained the instructions in which participants were asked to read the scenario carefully and to complete the questions, they were informed that they could refer back to the scenario as often as they wished. The remaining pages contained one of three versions of the scenario, a question-and-answer sheet and a debriefing paragraph.

Results and Discussion

The experiment shows that people mutate actions more than their reasons, whether the action is preceded by a plan-based reason (75% versus 21%, McNemar Test, $chi^2 = 22.76$, $n = 61$, $p < .0001$), an own-sake reason (89% versus 3%, McNemar Test, $chi^2 = 50.16$, $n = 63$, $p < .0001$) or no reason (88% versus 1.5%, McNemar Test, $chi^2 = 55.02$, $n = 66$, $p < .0001$). As Table 4.2 shows, different reasons have a differential effect on the mutability of actions. A plan-based reason (trying to lose weight) reduces the mutability of an

action, compared to an own-sake reason (liking to go to the gym) or no reason (i.e., a descriptive filler: the gym is large). Participants provided on average 3.7 counterfactual thoughts each and overall they mutated the action when it was preceded by a plan-based reason (75%) less often than when it was preceded by an own-sake reason (89%, chi^2 = 3.86, df = 1, 2-tailed $p < .05$), or no reason (88%, $chi^2 = 3.32$, df = 1, one-tailed $p = .03$), as Table 4.2 shows. They mutated an action preceded by an own-sake reason as often as one with no reason (89% versus 88%, $chi^2 = .032$, df = 1, two-tailed $p < .86$). The results show that some reasons are more mutable than others: plans are more mutable than own-sake desires. In fact, an examination of the content of participants' counterfactual thoughts shows that they rarely focused on altering desires. For example, in the plan-based reason condition, participants tended to mutate the necessity of the plan, that is, the protagonist's *need* to lose weight (19.5%), rather than his *desire* to lose weight (1.5%).

A plan-based reason reduces the mutability of an action because people mutate the plan instead. Participants mutated the plan-based reason (21%) more often than the own-sake reason (3%, $chi^2 = 9.587$, df = 1, $p < .002$) or no reason (control filler item: 1.5%, $chi^2 = 12.666$, df = 1, $p < .0004$). There was no difference in the tendency to focus on the reason when it was an own-sake reason or a control filler (3% versus 1.5%, $chi^2 = 0.39$, df = 1, $p < .53$).[6]

[6]There were no difference between conditions in the mutability of the action when only the first counterfactual completions were considered (control: 41%; own-sake reason: 43%; means-end reason 39%). Less than 1.5% of mutations in each condition focused on the reason in the first mutations.

Table 4.2: The percentages of mutations of the action and the reasons in the four counterfactual thoughts generated in the three conditions of Experiment 4

	Plan-based desire (Weight)	Own-sake desire (Enjoyment)	Control (Large gym)
All counterfactuals			
Action (Visit to Gym)	75	89	88
Desire	1.5	3	1.5
Plan	19.5	-	-

Reasons had no effect on ratings of self-blame and regret. Participants were asked to rate the protagonist's self-blame on a scale from 1 (not at all to blame) to 9 (completely to blame). They rated Tom as blaming himself somewhat when his action was preceded by a plan-based reason (M=4.66), an own-sake reason (M=5.14) or no reason (M= 5.18), and there were no differences between the ratings, (Kruskal-Wallis, $chi^2 = 2.01$, df = 2, p = .37). Participants were asked to rate the protagonist's regret for his action on a scale from 1 (no regret) to 9 (strong regret). They rated Tom as regretting his action quite strongly when it was preceded by a plan-based reason (M=7.92), an own-sake reason (M=7.63) or no reason (M= 7.55), and there were no differences between the ratings, (Kruskal-Wallis, $chi^2 = 0.51$, df = 2, p = .77).

The experiment provides the first demonstration that people mutate actions more than the reasons for acting. Unlike the tendency to mutate causes more than their effects, there is no tendency to mutate reasons more than their actions. This result corroborates

our suggestion that reasons for actions may be represented in a weaker conditional relation than causes for effects. People may keep in mind several possibilities to represent reasons and actions, including the possibility in which the reason doesn't occur but the action does, or the situation in which the reason occurs and the action doesn't.

The experiment also shows that some reasons are more effective at reducing the mutability of an action than others. Plan-based reasons are more effective than own-sake reasons at reducing the mutability of an action. People mentally undo an action less often when a plan-based reason is given for it compared to when no reason is given, and they undo an action equally often when an own-sake reason is given as when no reason is given. This result is consistent with the finding that some reasons are more mutable than others. When the plan for acting is explicitly given, then some participants shift their focus from the action to the plan.

Of course, participants may impute an own-sake reason to the actor even in the no reason condition (they may assume he went to the gym because he likes to do so even when they are not explicitly told so). An own-sake reason may be considered to provide little independent information and so participants may not mutate it separately from the action. In the next experiment we will return to examine the effectiveness of different sorts of reasons in reducing the mutability of actions.

The results provide support for the context theory of counterfactual thinking. They provide the first evidence that the same action may differ in mutability depending on the context in which it is placed, that is, depending on whether the reason for acting is interpreted to be strong, conditional or enabling.

The provision of a reason of whatever sort had no effect on ratings of self-blame or regret. Although certain sorts of reasons - plan-based reasons -- may shift the focus of participant's counterfactual thoughts from a focus on the controllable action to a focus on

the reason for that action, it does not affect their judgements about the emotional impact of the event, at least for certain emotions.

We have shown in this experiment that reasons are mutated less often than the actions they lead to. The result contrasts with the finding of Wells, Taylor and Turtle (1987) that causes are mutated more often than the events they lead to. We suggest that this difference may arise from underlying differences in the representation of reasons and causes. Our aim in the second experiment was to compare the mutability of reasons for actions and the mutability of causes for events.

Experiment 5: Reasons for Actions and Causes for Events

In the experiment we described a situation, e.g., a tennis player doesn't train the day before a match. We indicated that this situation arose because of an action (he decides to rest) preceded by a reason (he wants to be fresh for the following day), or because of an event (the courts were closed) preceded by a cause (they were water-logged). In our scenarios, actions are those occurrences that are performed by an individual (e.g., deciding to rest), and events are occurrences that are not performed by a person (e.g., the courts being closed). This distinction between actions and events is somewhat similar to the distinction between controllable and uncontrollable occurrences (e.g., Girotto et al, 1991). The actions are intentional when they are performed for a reason, that is, because of the beliefs and desires of the actor. The reasons used in this experiment were immediate reasons, e.g., wanting to be fresh for the following day. They were plan-based reasons, that is, they state the goal for which the action or event occurred. We examine whether short-term plans are like other immediate reasons (such as the own-sake reason used in the previous experiment) in that they do not affect the mutability of the action or whether they are like other long-term plans (such as the more enduring plan of wanting to

lose weight which was used in the previous experiment) which reduce the mutability of the action. The causes were uncontrollable events or states of affairs, e.g., courts being water-logged, which led to an action or an event.

In the experiment we compared reason-action sequences (e.g., wanting to be fresh was the reason for his decision to rest) and cause-event sequences (the water-logging caused the courts to close). As controls, we also compared reason-event sequences (trying to save the lawns was the reason for the courts to close) and cause-action sequences (being tired after the long journey caused him to take a rest). Note that a cause-action sequence indicates the action was unintentional (because it suggests that there was no desire to act). We compared these four sorts of scenarios to two further controls, one in which the action was described without a reason or cause (and with a filler description instead), and one in which the event was described without a reason or cause (with a filler description).

In order to study the effects for a wider range of materials, we used six scenarios, which were about a tennis player's failure to win a match, a car accident on the way home from work, a car accident on the way home from school, an accident at work, a student's poor performance in an examination and a couple who drive off a collapsed bridge. The scenarios were based on ones in the literature (Kahneman & Tversky, 1982b; Klauer, Jacobsen & Migulla, 1995; Morris & Sim, 1999; N'gbala & Branscombe, 1995; Roese & Olson, 1993; Wells & Gavanski, 1989), and the full set is in Appendix 4.

Method

Materials and design

We constructed six scenarios based on ones used previously in the literature on counterfactual thinking (see appendix 4). The scenarios contained a series of factors

which contributed to a negative outcome. We constructed six versions of each scenario. Half of the versions contained an action and the other half contained an event (and these had the same subsequent consequences). In each case, the action or event was preceded by either a cause, a reason, or a filler item.

Participants received six different scenarios: a reason-action, cause-action, filler-action, reason-event, cause-event, or filler-event scenario, in a within-subject design. They received the six scenarios instantiated in six different contents, which were allocated at random to the scenario versions for each participant. Each participant read the scenarios in a different random order. After reading each scenario, participants completed three tasks, a counterfactual mutation task, and ratings of self-blame and regret, as follows:

1. As commonly happens in such situations, *the protagonist* often thought, "if only..." during the days and weeks that followed. Please list in the order of importance, four ways in which *he/she* may have completed this thought so that the outcome would be different.

2. On the following scale, please rate the extent to which you think *the protagonist* felt to blame for the outcome:

1 -------- 2-------- 3-------- 4-------- 5-------- 6-------- 7-------- 8-------- 9
not at all to blame — completely to blame

3. On the following scale, please rate the extent to which you think *the protagonist* felt regret following the outcome:

1 -------- 2-------- 3-------- 4-------- 5-------- 6-------- 7-------- 8-------- 9
no regret — strong regret

They completed their questions in the fixed order above on the answer sheet provided.

Participants and Procedure

The thirty-six participants who took part in the experiment were members of the department's participant panel (recruited from the general population through national newspaper advertisements) who were paid for their participation, as well as postgraduate and undergraduate volunteers. The participants were 16 men and 20 women who ranged in age from 18 to 78 years (mean age of 35). The participants were tested individually in a quiet room. They were given an eight-page booklet containing instructions that were similar to those used in the first experiment, the six scenarios and a debriefing paragraph, see appendix 4.

Results and Discussion

The experiment demonstrates that actions are more mutable than events (consistent with earlier findings that controllable occurrences are more mutable than uncontrollable ones, Girotto et al., 1991) as Table 4.3 shows. Participants provided on average 3.9 counterfactual thoughts for each scenario, and overall, they mutated actions more than events (78% versus 59%, McNemar Test, $chi^2 = 6.95$, $p < .01$); they did so when both were preceded by reasons (83% versus 56%, McNemar Test, $chi^2 = 4.5$, $p = .03$) but not reliably so when both were preceded by causes (64% versus 56%, McNemar Test, $chi^2 = 0.21$, $p = .65$).[7]

The experiment provides the first demonstration that causes are more mutable than reasons. As Table 4.3 shows, participants mutated causes more than reasons (38% versus 11%, McNemar Test, $chi^2 = 12.96$, $p = .0002$), regardless of whether the

[7] The same pattern of results holds when only the first counterfactual completions are considered. Participants mutated actions (48%) more than events (27%, McNemar Test, chi-squared = 7.93, df = 1, $p < .005$).

antecedent led to an action (42% versus 14%, McNemar Test, $chi^2 = 5.79$, $p < .02$) or to an event (33% versus 8%, McNemar Test, $chi^2 = 5.81$, $p < .02$). These results provide support for our suggestion that reasons and actions are represented as weaker relations than causes and events.

Table 4.3: The percentages of mutations of actions and events, and of reasons and causes, in Experiment 5

	Reason	Cause	Control	Total
Counterfactuals				
Action	83	64	86	78
Antecedent	*14*	*42*	*0*	*19*
Event	56	56	67	59
Antecedent	*8*	*33*	*0*	*14*

The experiment shows that people undo actions whether they are preceded by reasons or by causes, that is, whether the action is intentional or unintentional. But there is a difference: people undo causes for actions more than reasons for actions. For reason-action sequences, people tended to undo the action (83%), and they undid it more than the reason (14%, McNemar Test, $chi^2 = .86$, $p < .0001$). For cause-action sequences, people also tended to undo the action (64%), but not reliably more than the cause (42%)[8] and the difference was not significant (McNemar Test, $chi^2 = 2.23$, $p = .14$). The type of antecedent (reason, cause or none) influenced the mutability of actions (Cochrans Q =

[8] Some participants undid both, and so the percentages do not add to 100%.

6.71, df = 2, 1-tailed p < .02). The cause reduced the action's mutability more than the reason (64% versus 83%, McNemar Test, $chi^2 = 3.27$, 1-tailed p < .04), and the cause reduced the action's mutability more than the filler (64% versus 86%, McNemar Test, $chi^2 = 3.5$, 1-tailed p < .03), unlike the reason compared to the filler (83% and 86%).[9] The results are consistent with the findings of the previous experiment which show that short-term plans like other internal immediate reasons are not effective in reducing the mutability of an action.

For cause-event sequences, people tended to undo the event (56%), and they undid the event more than the cause (33%) although the difference did not reach significance (McNemar Test, $chi^2 = 2.04$, p = .15). For reason-event sequences, people also tended to undo the event (56%) and they undid it more than the reason (8%, McNemar Test, $chi^2 = 11.13$, p = .0005). The type of antecedent did not influence the mutability of the event. Participants mutated the event as often when it was preceded by a cause (56%) a reason (56%) or by a filler (67%, Cochrans Q = 1.52, df = 2, n = 36, p < .47). The tendency to undo the event more than the cause is in contrast with previous research (Wells et al., 1987) which has shown a tendency to undo the first cause more than its effect. Of course, Wells et al. examined a series of four temporally contingent events whereas we studied single direct causes. Nonetheless, causes are undone more than reasons.

Participants rated the protagonist's self-blame on a scale from 1 (not at all to blame) to 9 (completely to blame), and they rated the protagonist as blaming him or herself moderately, and more in the action conditions (mean = 5.33) than in the event

[9] The same pattern of results holds when only the first counterfactual completions are considered. The type of cause (i.e., reason, cause or none) influenced the mutability of actions (Cochrans Q = 5.15, df = 2, 1-tailed p < .04) Participants mutated the action less in the cause condition than in the reason condition (33% versus 58%, McNemar Test, chi-squared = 3.76, df = 1, 1-tailed p < .03) or the control condition (33% versus 53%, McNemar Test, chi-squared = 1.71, df = 1, 1-tailed p < .10) Participants did not mutate the action less in the reason than in the control condition. Also, the type of cause did not influence the mutability of the event.

conditions (mean = 4.6, Wilcoxon Test, $z = 2.44$, $df = 1$, $p < .02$). There was no difference in the ratings of self blame for the action or event depending on whether they were preceded by a reason (mean = 5.10), a cause (mean = 5.04) or neither (mean = 4.75, Friedmans Test, $chi^2 = 0.09$, $df = 2$, $p = .96$).

Participants rated the protagonist's regret following the outcome on a scale from 1 (no regret) to 9 (strong regret), and they tended to judge that he or she would regret the outcome strongly, although there was no difference for whether the outcome resulted from an action (mean = 7.53) or an event (mean = 7.26, Wilcoxon Test, $z = .94$, $df = 1$, $p < .35$). There was also no difference in the ratings of regret for the action or event depending on whether they were preceded by a reason (mean = 7.49), a cause (mean = 7.25) or neither (mean = 7.44, Friedmans Test, $chi^2 = 0.88$, $df = 2$, $p = .64$).

The experiment demonstrates that actions are more mutable than events although this is not the case when both are preceded by causes. The results support the view that actions tend to be perceived as unconstrained by causes in the world. It also provides the first demonstration that causes are more mutable than reasons. These results are consistent with our suggestion that reasons and actions are represented as weaker relations than causes and events. Hence, the representation of reasons and actions may include more possibilities than causes and their effects. The difference has implications for counterfactual thinking. Causes and events may be represented as a strong relation, and so undoing the cause ensures that the event is also undone. However, reasons and actions may be represented in a conditional or enabling relation, and so undoing the reason may provide no guarantee that the action is also undone. Hence people tend to undo actions directly when they think counterfactually about a reason-action sequence, whereas they tend to undo causes when they think counterfactually about a cause-event

sequence. The results also show that people undo causes more than reasons for actions and they undo intentional actions more than unintentional actions.

The results are consistent with the context theory of counterfactual thinking. Actions may be salient but their mutability depends on the context. The same action is more mutable when it is described as intentional than when it is described as unintentional.

In our next experiment we examined more closely the effects of different sorts of reasons on actions. Our aim was to extend the comparison of different sorts of reasons to include external as well as internal reasons.

Experiment 6: Goals and Obligations as Reasons

Our aim in the third experiment was to compare the effect on the mutability of actions of an internal immediate reason, (a short-term goal of availing of an opportunity) and an external immediate reason (a short-term obligation). An individual will have an *internal* reason for acting if there is some end that he wants to reach, and if he thinks that the action is needed or conducive to that end. Internal reasons, such as the 'own-sake' reasons and plan-based reasons examined in the previous experiments, may be considered to be relatively weak, and we have shown that people tend not to mutate them, focusing directly on the action instead. In contrast, an external reason for an action may be represented as a strong relation. External reasons may include communicative challenges such as orders, requests and questions, and prescriptive rules and norms such as community customs, fashions and traditions (Von Wright, 1983).

In an earlier experiment (see Walsh & Byrne, 2001d), we examined the effect of providing one type of external reason for acting, that is, an invitation to act on an action's mutability. We constructed a scenario about an individual, Paul, who was killed in a car

accident on the way to visit his friend, John. In one version of the scenario, no reason was given for Paul's action, deciding to visit John. In the other version, Paul's action was preceded by a telephone call from John inviting him to visit. In accordance with our predictions, we found that an invitation reduced the mutability of the action. Participants told about this external reason may consider that undoing it is sufficient to undo the action: if the invitation had not occurred then the protagonist would not have formed the intention to visit his friend which led to his subsequent act.

In the present experiment we aimed to examine another type of external reason, that is, a social obligation to act. We had two main reasons for studying obligations. First, obligations form an important type of reason for acting because much of people's behaviour is guided by social rules and norms. Second, obligations provide an interesting comparison to other types of reason because they tend to provide a sufficient condition for acting. Previous research suggests that social norms may constrain the mutability of actions. People's counterfactual thoughts tend to focus on immoral or inappropriate actions more than moral or appropriate ones (McCloy & Byrne, 2000; N'gbala & Branscombe, 1995), perhaps because socially exceptional events bring to mind their normal counterparts (Kahneman & Miller, 1986; McCloy & Byrne, 2000). People judge an action to be more unacceptable when it is inconsistent with social obligations than when it occurred accidentally (Green, McClelland, Muckli & Simmons, 1999). Hence, social obligations may exert strong constraints on actions (Wells et al., 1987). We suggest that people may tend not to immediately imagine the situation in which a person has an obligation but does not act on it. For this reason, we expect external reasons such as obligations to have a stronger effect on the mutability of actions than internal reasons such as plans.

Our aim in the experiment was to compare the effects on an action's mutability of

an external reason, i.e., an obligation, and an internal reason, i.e., a short-term plan-based reason. Once again we wished to generalise our results to a further new content, this time a performance related content.

Method

Materials

We constructed three versions of a scenario based on one used by Klauer, Jacobsen and Migulla (1995). In our scenarios, in one version the decision to attend a ball resulted from an obligation to attend, in the second version the decision resulted from a desire to avail of an opportunity, and in the third version no reason was given for the decision to attend (and a descriptive filler item was provided instead). The scenario was as follows:

> Bernard is a famous violinist. At the moment, he is staying in Vienna where he has been invited to perform in an important open-air festival of classical music. In the course of this festival, Bernard intends to perform Beethoven's violin concerto. He knows the concerto well and his previous performances of it have been successful.
>
> Bernard will be accompanied by the well-known Geneva Chamber Orchestra. Since the accompanying orchestra plays a prominent role in the soloists performance, Bernard and his agent had negotiations with several orchestras prior to the festival. The English Polyphonic Orchestra had been a real alternative, however, finally they decided on the Geneva Chamber Orchestra.
>
> The evening before the concert, there is a ball organised by the festival committee.
>
> *[Obligation: The ball is being run to help cover the festival expenses and therefore Bernard is obliged to attend.]*
>
> *[Plan-based reason: Bernard will have a unique opportunity to meet a world-renowned violinist at the ball and this is his goal for attending.]*
>
> *[Control: The ball is being held in the Vienna City Ballroom.]*

Bernard goes to the ball and as a result he has a late night, something he doesn't usually do before a concert.

The day of the concert Bernard thinks carefully about which violin he is going to use. He has the choice between a Stradivari and an Amati. He finally decides on the Amati because although he doesn't use it very often, he thinks that the timber of this violin is best suited to Beethoven's concerto.

At the concert that evening, Bernard's performance of the concerto was poor and the audience did not respond well. He was bitterly disappointed.

Participants completed three tasks, a counterfactual thoughts task, and a self-blame and a regret rating task. They also completed two final questions designed to examine the extent to which the protagonist felt he had a choice or felt compelled to attend the ball. The questions were as follows:

1. Bernard often thought, "if only . . ." in the days that followed the concert. Please list in the order of importance, four ways in which he may have completed this thought so that the outcome would be different.

2. On the following scale, please rate the extent to which you think Bernard felt to blame for his poor performance:

1 -------- 2-------- 3-------- 4-------- 5-------- 6-------- 7-------- 8-------- 9
not at all to blame — completely to blame

3. On the following scale, please rate the extent to which you think Bernard regretted going to the ball:

1 -------- 2-------- 3-------- 4-------- 5-------- 6-------- 7-------- 8-------- 9
no regret — strong regret

4. On the following scale, please rate the extent to which you think Bernard had a choice in going to the ball:

1 -------- 2-------- 3-------- 4-------- 5-------- 6-------- 7-------- 8-------- 9
no choice — completely free choice

5. On the following scale, please rate the extent to which you think Bernard felt compelled to go to the ball:

1 -------- 2-------- 3-------- 4-------- 5-------- 6-------- 7-------- 8-------- 9
not at all compelled — strongly compelled

Participants completed the questions in the fixed order above on the answer sheet provided. They were assigned at random one of the three scenarios, the obligation reason scenario (n = 69), the opportunity reason scenario (n = 73) or the no reason scenario (n = 72) in a between-participants design.

Participants and Procedure

The 217 undergraduate students who participated in the experiment voluntarily were 122 men and 95 women from various departments in the University of Dublin, Trinity College. They ranged in age from 17 to 45 years with a mean age of 21 years. Prior to the analysis, 3 participants were eliminated because they failed to follow the instructions correctly. The participants were tested in several large groups. They were given a four-page booklet containing instructions similar to the ones in the previous experiments, one of three versions of the scenario with a mutation task, a question-and-answer sheet and a debriefing paragraph, see appendix 4.

Results and Discussion

Obligations reduced the mutability of an action more than plan-based reasons. Participants provided on average 3.82 counterfactual thoughts, and overall, they mutated the action when the actor was obliged to do it less often (67%) than when he had no reason (82%), and the difference is reliable, ($chi^2 = 4.33$, df 1, n=141, one-tailed $p < .02$) as Table 4.4 shows. They also showed a tendency to mutate the action when the actor was obliged to do it less often than when he wanted to achieve a goal (67% versus 75%),

and to mutate the action when the actor wanted to achieve a goal less often than when he had no reason (75% versus 82%), although neither of these two differences reaches significance (chi^2 = 1.30, df 1, n=142, one-tailed $p < .13$, and chi^2 = .94, df 1, n=145, one-tailed $p < .33$ respectively). Once again, some of the difference is due to participants' focusing instead on the reason in the obligation condition (6%), and in the plan-based reason condition (7%), more often than in the control condition (0%, chi^2 = 4.30, n = 141, df 1, $p < .025$ and chi^2 = 5.11, n = 145, df 1, $p < .025$ respectively).[10] These results corroborate our suggestion that some reasons are represented as more strongly leading to the action than others. Short-term goals, just like 'own-sake' reasons, may be represented in a weak relation to the action and tend to be relatively immutable and so the action is mutated directly. Obligations, like long-term plan-based reasons, may be represented in a stronger relation to the action, and so the action is mutated less often.

Table 4.4: The percentages of mutations of the action and the reasons in Experiment 6

	Obligation (Fundraising)	Opportunity (Meet composer)	Control (Ballroom)
All counterfactuals			
Action (Attending the ball)	67	75	82
Reason	6	7	0

[10] There was no difference between conditions in the mutability of the action when only the first counterfactual completions are considered (control: 51%; means-end desire: 44%; obligation: 49%). Less

Participants rated whether the protagonist felt to blame on a scale from 1 (not at all to blame) to 9 (completely to blame). They judged him to feel quite to blame for the outcome of the action when he was obliged to act (6.88), wanted to achieve a goal (M= 7.34), or had no reason (M = 6.94) and there were no reliable differences in their ratings (Kruskal-Wallis, $chi^2 = 2.8$, df = 2, p = .24). Participants rated the protagonist's regret on a scale from 1 (no regret) to 9 (strong regret). They judged him to feel quite strong regret for the outcome of the action when he was obliged to act (7.0), wanted to achieve a goal (M= 6.9), or had no reason (M = 7.3) and there were no reliable differences in their ratings (Kruskal-Wallis, $chi^2 = 1.8$, df = 2, p = .41).

Participants rated the protagonist as feeling compelled to attend the ball more when he had an obligation to attend (mean = 7.3) or wanted to achieve a goal (M = 6.9) than when he had no reason (M = 5.8, Mann Whitney U = 1218, one-tailed $p < .00005$ and Mann Whitney U = 1694, one-tailed $p < .0001$ respectively). They considered him to be more compelled to attend when he had an obligation than when he wanted to achieve a goal (Mann Whitney U = 2086, one-tailed $p < .04$). Participants rated the protagonist as feeling he had less choice in attending the ball when he was obliged to attend (M = 4.5) than when he wanted to achieve a goal (M = 6.93, Mann Whitney U = 1170, one-tailed $p < .00005$) or had no reason (M=6.22, Mann Whitney U = 1450, one-tailed $p < .00005$). Although ratings of free choice were lowest in the obligation condition, the actor is still perceived as having a moderate amount of control over his choice of action. Perhaps surprisingly, they rated the protagonist as feeling he had less choice when no reason was specified than when he wanted to achieve a goal (M=6.22 versus M=6.93, Mann Whitney U = 2103, two-tailed $p < .04$). Ratings of the extent to which the actor felt compelled to

---

than 1% of mutations in each condition focused on the reason in the first mutations.

go to the ball and the extent to which he had a choice were correlated (Spearman Rank r = -.18, n = 214, p = .01).

The results show that external reasons such as obligations reduce the mutability of actions compared to goals, they are also judged to make the protagonist feel more compelled to carry out the action than goals, and to feel he has less choice in carrying out the action than goals. Internal reasons such as wanting to achieve a short-term goal do not reduce the mutability of actions compared to no reason, but they are judged to make the protagonist feel more compelled to carry out the action than no reason, yet to make the protagonist feel he had more choice in carrying out the action than no reason.

The results are consistent with our suggestion that these reasons have different representations. It is more difficult to imagine acting in a way that is inconsistent with an obligation than to imagine acting in a way that is inconsistent with a desire. Obligations which are perceived as undesirable may be represented as a strong relation, the action occurs when the obligation is present and does not occur when it is absent. Other obligatory actions, such as 'attending the ball' in the present chapter, may be represented as a conditional reason (see Table 4.1). The action may be desirable and therefore it is possible to imagine that even without the obligation the protagonist may have acted for another reason. In accordance with the predictions of the context theory, the results also suggest that differences in the representation of reasons can influence mutability.

General Discussion

This series of three experiments provides the first systematic experimental investigation of the effects of reasons for actions on counterfactual thinking. The results provide the first empirical demonstration that people mutate actions more than the reasons for acting.

They also provide the first demonstration that the mutability of actions varies depending on the context in which an action is placed, in this case, on the reason for acting.

The first experiment showed actions are mutated more than reasons. It also showed that the mutability of an action depends on the reason for which it was done. An action (going to the gym) was mutated less by participants who were given an internal enduring reason (wanting to lose weight) than those given either an internal immediate reason (liking to go to the gym) or no reason, and there was no difference between an internal immediate reason and no reason. The second experiment compared reasons and causes for acting and reasons and causes of events. As in the previous experiment, actions were mutated more than reasons. The mutability of the action (e.g., inspecting a building) was not reduced when it was done for an internal immediate reason (e.g., wanting to inspect the wiring). People continued to focus on the action. However, the mutability of the action was reduced when it was caused by a mistake (e.g., misreading the rota). In this case, people shift some of their focus from the action to the cause. The mutability of an event (e.g., the building was closed) was not significantly affected by either a cause (e.g., a water leakage) or a reason (e.g., to allow for repairs). Overall, actions were mutated more than events but this difference was reduced when both were accompanied by a cause. The third experiment showed that a reason derived from external sources, such as an obligation, reduced an action's mutability compared to one derived from internal sources, such as a short-term plan-based reason, or no reason. Participants who were given an obligation reason (having to attend a ball because it will raise funds for a classical music festival) mutated the action (attending the ball) less often than participants given a short-term plan-based reason (attending the ball because it provides the opportunity to meet a world-renowned violinist).

We have shown that different reasons have different effects in counterfactual thinking. Some reasons reduce the mutability of actions: we have shown that certain plan-based reasons, causes such as errors and distractions, obligations, and invitations to act, all reduce the mutability of actions. We suggest that people interpret the relation between these reasons and their actions as a strong or conditional relation. The provision of the reason shifts some of the focus onto the reason instead in people's counterfactual mutations. Other sorts of reasons, such as internal immediate reasons, do not reduce an action's mutability. We suggest that people interpret the relation between these reasons and their actions as an enabling relation, that is, they mentally represent the possibility in which the reason occurs and the action does not. Of course, there are many different sorts of reasons but we suggest that the distinction between strong, conditional and enabling is a useful one for the beginning of an analysis of the role of reasons in counterfactual thinking. On this account, an action is less mutable if it is preceded by a reason which is sufficient for its occurrence. Where an action has a reason which is not sufficient or where there appears to be no reason for it, people will mutate the action more. The result is consistent with the conjecture that failures to act may be regretted more over time than actions in part because the reasons for failures to act tend to be forgotten (Byrne & McEleney, 2000; Gilovich & Medvec, 1995; Savitsky, Medvec & Gilovich, 1997).

We have concentrated on scenarios which end in a bad outcome because people spontaneously think counterfactually most often following bad outcomes (e.g., McEleney & Byrne, 2001). Nonetheless people focus on actions even when they think about good outcomes (e.g., Landman, 1987) and we conjecture that reasons may have similar effects on counterfactual thinking about good outcomes as well as bad outcomes.

Our experiments once again demonstrate the importance of context in event mutability. The focus theory proposes that people focus on actions because they are

salient and they mutate them (Legrenzi, Girotto & Johnson-Laird, 1993; Morrow, Bower & Greenspan, 1989). The theory can explain the finding that actions are more mutable than reasons and the finding that actions are more mutable than other uncontrollable events. However, our results also show that the same action can differ in mutability depending on the context in which it is placed, in this case the reason for acting. For example, as our third experiment showed, external immediate reasons reduce the mutability of the action whereas internal immediate reasons do not. A reason for acting can make some possibilities easier to imagine and some more difficult. The focus theory assumes that the context of possibilities is only consulted only after an event is mutated to determine the outcome of the counterfactual scenario and therefore it cannot explain this finding.

Instead the context theory may provide an explanation for the findings. The theory proposes that people generate a counterfactual by searching for alternative possibilities. Items in the factual scenario which are not contained in the recruited alternative will be mutated. The context may provide alternatives to the factual situation. In our studies of the temporal order effect in the previous chapter, this context of possibilities was provided by the winning conditions. By varying the description of these winning conditions we influenced what aspects of the conditions were explicitly represented. In our studies of controllable actions in the present chapter, the context of possibilities may be provided by background knowledge of the relation between reasons and actions and causes and events. People may represent the possibilities that are consistent with these relations and these may form the context of possibilities. People's knowledge of the different types of relation may give rise to different sets of possibilities and may in turn influence the availability of alternatives to the facts.

Our results suggest that the context of possibilities, whether provided explicitly or through background knowledge, may influence the mutability of events. We have demonstrated the effect both for independent sequences of events in the previous chapter and for causally related events in the present chapter. Both effects may arise from the same underlying principles. People may attempt to match the facts to the context of possibilities. For the temporal order effect, they may use the first fact to recruit an alternative, whereas for the controllability effect they may use reasons or goals to recruit an alternative. Our experiments on the temporal order effect showed that when there was no explicit match for the first event, people mutated it most often. Similarly, causes which tend to be represented as a strong relation, for example, the water-logging caused the courts to close:

Factual:	Water-logging	Closed
Counterfactual:	¬ Water-logging	¬ Closed

may have no explicit match in the counterfactual models, i.e., the counterfactual models contain only instances where the cause does not occur. For this reason, people may tend to mutate the cause. The case may be different when there is an explicit match. In the temporal order effect we demonstrated that when there was an explicit match for the first event, people tended to mutate the second. This effect may also explain the shift in focus from reasons to actions. Reasons, particularly desires may be frequently represented as an enabling relation, for example, wanting to be fresh for the match is the reason for resting:

Factual:	Fresh	Rest
Counterfactual:	Fresh	¬ Rest
Counterfactual:	¬ Fresh	¬ Rest

In this case there is an explicit match for the reason, and so people may instead shift some of their focus to the action.

Our studies provide initial steps towards a unified theory of counterfactual mutability, based on the context theory. Our analysis suggests that both the causal and temporal order effects may be guided by the mental representation of the context of possibilities. These possibilities are represented when people comprehend a situation but different factors may influence their representation. The verbal description may make some information explicit and may leave other information implicit. Also, background knowledge of the relation between events will influence what we represent as possible. These different representations can influence the accessibility of counterfactual alternatives to the facts. People may use the facts to recruit an alternative. In the temporal order effect, the first event may be used to select an alternative from the context of possibilities. In scenarios such as the one used in the present scenario, the goal may recruit an alternative.

The provision of a reason had no effect on emotional or social judgments, such as ratings of self-blame, or regret, perhaps because the shift in focus from the action to the action's reason does not change the overall focus on the event itself and social and emotional judgments may be a function of the overall event that is in focus, rather than which specific aspect of that event is in focus. An exception may arise where an individual is induced or coerced by another to act (Hart & Honore, 1959). However, the results are consistent with recent research which suggests that the influence of counterfactual thinking on emotional and social judgements may not be straightforward (e.g. Byrne, Segura, Culhane, Tasso & Berrocal, 2000; N'gbala & Branscombe, 1997). Our results suggest that counterfactual thinking may be influenced by the availability of alternatives. However, as we found with the temporal order effect, judgements of emotion

may not be affected by the availability of these alternatives. Instead, judgements of emotion may be influenced more strongly by the representation of the factual events.

The provision of a reason affected judgements of compulsion and free choice. In the third experiment, judgements of how compelled or how much choice a protagonist had in carrying out a controllable action were influenced by the nature of the reason for acting. The result suggests that the simple distinction between controllable and uncontrollable actions may not capture the fine-grained subtleties in individuals' counterfactual thoughts about controllability.

People tend to dwell on controllable actions when they think about how things might have been different. The tendency is dependent on the context in which the action is placed. It is reliably reduced when a reason for the action is provided. Our results indicate that the focus of counterfactual thoughts on controllable actions, which is a pervasive and much studied aspect of counterfactual thinking, may arise in part because controllable actions seem to be independent of any external constraint. Once such constraints are made apparent, in the guise of the provision of reasons for acting, the mutability of controllable actions is reduced. People shift some of the focus of their counterfactual thoughts from the action to the reason. Different reasons have different effects. Causes, internal enduring reasons, such as long-term goals and external reasons such as obligations and invitations can reduce an action's mutability whereas immediate internal reasons may not. The effect may depend on how these reasons are mentally represented.

Our experiments provide the first experimental support for the suggestion that when people provide a reason for their behaviour, the behaviour seems less mutable, and thereby more inevitable. In fact, in everyday life, people often attempt to excuse their

behaviour by presenting the reasons for it. The efficacy of such excuses may depend on their differential ability to shift the focus of counterfactual thoughts.

Chapter 5: The Mental Representation of Multiple Possibilities: A study of inferences

In the previous chapters we have argued that the mental representations that people construct are of central importance in counterfactual thinking. People construct mental representations of situations and they may also construct or retrieve representations of some alternatives to the facts, which we term the context of possibilities. These representations may make some counterfactuals more available than others.

Our goal in the present chapter, is to examine in more detail the nature of the mental representations underlying the way that people think. We re-examine the idea outlined in chapter 2, that the verbal formulation of a problem or a story can influence the representation that people construct. According to the mental model theory, reasoners build models in accordance with the principle of truth, (see chapter 1, p12-13). As a result, two sentences with the same underlying meaning may be represented differently depending on how they are described. The description can lead people to make some information explicit and leave other information implicit. For example, as discussed in chapter 1, p15, reasoners may draw modus tollens inferences more readily from a counterfactual than from a factual conditional, because the latter conditional is represented more explicitly (Byrne & Tasso, 1999). Information which is made explicit is more accessible and therefore may also influence the availability of counterfactual alternatives (see chapters 2 and 3). In this chapter, we examine further how the verbal description of a problem may influence the inferences that people draw. In Experiment 7, we compare inferences drawn from exclusive disjunctions and biconditionals which have the same logical meaning but a different mental representation. In Experiment 8, we test

the idea that incomplete mental representations can lead people to make systematic and predictable errors.

We also aimed to examine a second factor that can influence the mental representations that people construct, that is, co-reference. Co-reference occurs when two items refer to the same thing. Take for example the following two sentences:

Jack solved the maths problem.

The genius always solves those problems.

Most readers infer that 'Jack' and 'the genius' refer to the same person and our aim is to examine how people represent these co-referential relations in their mental models. Co-reference is of central importance to counterfactual thinking because it forms the bridge between factual and imaginary worlds. Consider the following counterfactual conditional:

If Michael had solved the maths problem, he would have won a prize.

It may be represented by two models, one to represent the factual situation and one to represent the imaginary one (Byrne & Tasso, 1999):

Factual: ¬ solved ¬ won

Counterfactual: solved won

where 'solved' represents 'Michael solved the problem' and 'won' represents 'Michael won the prize'. But both representations share many common referents. Both refer to the same individual, Michael, both refer to the same problem, in one case it is solved and in the other it is not, and both refer to the same prize, which is won or not. Co-reference may play a crucial role in the generation of counterfactual thoughts. For example, in chapter 2, we described how our computational model, TempContext, searches through the context of possibilities for a co-referential match for the first fact. The accessibility of

different counterfactuals depends on the ability to find this match. Our aim is to examine how people represent these co-referential relations.

In this chapter, we use a different means of studying the representations that people construct, that is, through deductive reasoning. There is ample evidence to suggest that people find it easier to draw an inference when the information that they need is explicitly represented in the initial models that they construct. Deductive reasoning may therefore provide another means of corroborating and extending our theory about the mental representations underlying the way that people think.

We first describe previous research on deductive reasoning. We then describe two experiments that we carried out. The experiments examine the idea that people may construct representations that are just partially explicit and that co-reference may allow people to construct more concise representations. Finally, we consider the implications for counterfactual thinking.

Deductive Reasoning

Deduction involves reasoning from a set of premises or statements to a conclusion. For a conclusion to be valid it must follow necessarily from the premises. This means that deductive reasoning does not produce new information beyond that given in the premises. However, the conclusion makes explicit information which was previously implicit in the premises.

Deduction involves reasoning from general principles to particular situations. It allows people to draw inferences from their knowledge about how the world works, from rules and from social conventions. With deduction, people can make predictions about what the consequences of certain events might be and they can draw inferences about the causes of certain outcomes. Deductive reasoning is therefore central to much of human

thought. It makes it possible to evaluate different courses of action and to make decisions. It allows people to make plans based on their expectations of what the future might hold and to solve problems. It enables people to function in a social world; to follow instructions and rules and to pursue arguments and negotiations. Deduction makes it possible to test hypotheses and evaluate different theories, processes which are fundamental for example to the practice of science and law (e.g., Johnson-Laird & Byrne, 1991).

In this chapter, we are concerned with one important type of deduction, that is, propositional reasoning. Propositional reasoning involves reasoning from premises which contain sentential connectives such as *if, and* and *or*. In our studies of counterfactual thinking, we have seen that the context of possibilities can take the form of a conditional or a disjunction of possibilities. Our aim therefore, is to examine what type of inferences people draw from these types of sentences.

Typically in reasoning experiments, participants are presented with two premises. The first premise is a sentence containing a connective, such as a conditional. It is followed by a categorical premise which asserts or denies one of the clauses in the previous sentence. For example:

If Jack got a ticket, then he went to the concert.

Jack got a ticket.

After reading these premises, participants may be asked to generate or evaluate a conclusion, e.g.:

Jack went to the concert.

Mental Models and Deductive Reasoning

As described in chapter 1 (p11), the theory of mental models postulates that individuals who have no training in logic represent the meaning of assertions in mental models (Johnson-Laird & Byrne, 1991). Each mental model represents a possibility. But, the limitations of working memory force individuals to abide by the *principle of truth:* mental models represent only true possibilities and within them the constituent propositions in premises only when they are true (Johnson-Laird & Byrne, in press). For example, an exclusive disjunction such as:

Either Mary is in Brussels or Gino is in Rome, but not both

elicits two mental models of the alternative possibilities:

Mary-Brussels

Gino-Rome

where "Mary-Brussels" denotes a model of Mary in Brussels, and "Gino-Rome" denotes a model of Gino in Rome. Mental models of possibilities do not represent the falsity of the clauses in the disjunction, e.g., that Gino is not in Rome in the first possibility. One consequence which we demonstrated in chapter 3, is that mental models may make some possibilities more accessible when people generate counterfactual alternatives to reality. A second consequence which we aim to examine here, is that mental models may make some inferences easier than others.

According to the model theory, deductive reasoning involves three processes. First, reasoners build an initial set of models of the premises using their knowledge of the meaning of the connectives and also any relevant general knowledge. If possible, they will combine their models of the different premises. Second, they then use the model or models to formulate a conclusion. The conclusion will refer to information which is explicit in their models although it will not reassert one of the premises. If a conclusion is

possible, then its validity can be checked by searching for possible counterexamples. The theory predicts an inference will be difficult if the models of the two premises cannot be easily integrated. Take for example, the following conditional:

If the sun is shining, then Jane goes to the beach.

which has the following initial models:

sun beach

. . .

The categorical premise, 'The sun is shining', can easily be combined with these models because it is explicitly represented and the model yields the modus ponens conclusion that 'Jane goes to the beach'. However, the modus tollens inference is more difficult because the categorical premise, 'Jane does not go to the beach', cannot be combined with the initial models of the conditional. To draw the correct conclusion reasoners must flesh out their models, however, this is difficult and reasoners often fail to draw the correct conclusion, i.e., 'The sun is not shining'. The difficulty of an inference can also be predicted by the number of models that need to be represented. Inferences that require multiple models are more difficult than inferences that require a single model.

The semantic content of premises can also influence deductive reasoning. The meanings of clauses, their co-referential relations, and background knowledge, can all *modulate* the basic meanings of sentential connectives (Johnson-Laird & Byrne, in press) or causal relations (Goldvarg & Johnson-Laird, 2001). One way modulation can occur is when background knowledge facilitates or prevents the construction of a model. We demonstrated this in chapter 4 in our studies of reasons for acting. We described various situations where an individual acts for a reason. The meaning of the relation between the reason and the action determines what alternatives are possible. For example, it may be

easier to imagine that a person has a desire but does not act on it, than to imagine that they do not act on an obligation.

A second example of modulation is through co-reference. The ability to recognise that two expressions refer to the same object or event is essential for our understanding of language as well as our ability to reason and think counterfactually. Bouquet and Warglien (1999) discovered that reasoning from disjunctions was more accurate when the clauses were co-referential, e.g.:

Either Gino is in Brussels or Gino is in Rome, but not both.

Gino is not in Rome.

What follows?

Premises of this sort yielded a greater proportion of valid conclusions, e.g.: Gino is in Brussels, than premises that did not refer to the same individual in both clauses. Co-reference may enable people to construct representations that are more concise. For example, the following sentence without co-referential relations:

Either Mary is in Brussels or Gino is in Rome, but not both.

may be represented by two mental models, as follows:

Mary Brussels

Gino Rome

whereas a sentence with co-referential relations, for example:

Either Gino is in Brussels or Gino is in Rome, but not both.

may yield a more concise representation of the following sort:

Gino: Brussels

Rome

Our aim in the two experiments that we report was to examine how the verbal description of a sentence and the co-referential links between the clauses may influence the way that the sentence is mentally represented and hence how people reason from it.[11]

Experiment 7: Co-reference in reasoning

Our aim in this experiment was to examine how people reason from conditionals and disjunctions with different content (Walsh & Johnson-Laird, 2001). The form of the premises - whether they are based on a disjunction or a conditional - should make different information available. In the experiment, for half the inferences, the first premise was an exclusive disjunction, e.g.:

Rachel is climbing up the stairs or David is cooking at the stove but not both.

which can be represented in the following mental models:

Rachel-climbing

David-cooking

For half the inferences, the same information was expressed as a biconditional, e.g.:

If and only if Rachel is not climbing up the stairs then David is cooking at the stove.

which can be represented in the following mental models:

¬ Rachel-climbing David-cooking

. . .

Both sentences are consistent with the same fully explicit set of models, as follows:

Rachel-climbing ¬ David-cooking

¬ Rachel-climbing David-cooking

[11] These experiments were designed in collaboration with Phil Johnson-Laird during a three month visit to Princeton University funded by the Irish Council for the Humanities and Social Sciences.

The disjunction and the biconditional make different information explicit and hence, each sort of premise should make some inferences easier and some more difficult to draw. An inference should be easier to draw if there is an explicit match for the categorical premise in the models of the previous premise. In this case, the models can be integrated without a need to flesh them out. For example, it should be easier to reason from the disjunction when the categorical premise is 'Rachel is climbing the stairs', and easier from the conditional when it is 'Rachel is not climbing the stairs'. Many studies have demonstrated that it is more difficult to draw the modus tollens inference than the modus ponens inference from a conditional, however, the results for disjunctions are less clear (e.g., Johnson-Laird, Byrne & Schaeken, 1992; Schaeken, Johnson-Laird, Byrne & d'Ydewalle, 1995). Our aim was to re-examine this issue by directly comparing conditionals and disjunctions with the same underlying meaning. We measured response latencies as well as response accuracy.

Our second aim was to examine the effect of co-reference on reasoning. Our conjecture is that co-reference may allow reasoners to build a more concise representation of the premises, thereby reducing the load on working memory. We predicted that such representations should facilitate performance in reasoning tasks particularly if the demands on working memory are high.

Our aim was to extend the study of Bouquet and Warglien (1999) who examined the effect of co-reference on just one type of inference, that is, a denial of the second clause of a disjunction. We aimed to study both affirmative and denial inferences from biconditionals and disjunctions. We also extended the range of materials used. Bouquet and Warglien (1999) used a disjunction in which the two clauses were mutually exclusive; a person cannot be in two places at one time. Background knowledge may facilitate reasoning in this case. In order to manipulate co-reference, we used four types

of semantic content, as described in Table 5.1. The content involved people doing different actions. Three of the conditions contained co-reference and one did not. In the control condition, two people did two actions and there were no shared referents. In the second condition, one person does one of two actions which cannot be performed simultaneously. This is analogous to the example used by Bouquet and Warglien (1999).

Table 5.1: The four sorts of semantic content in Experiment 7

1. *Two-persons, two-actions (simultaneous):* Two persons do two different actions e.g., Brian is standing by the fireplace or Joanne is looking in the mirror, but not both.
2. *One-person, two-actions (non-simultaneous):* One person does two different actions, which cannot be performed simultaneously e.g., Sarah is sitting in the arm-chair or Sarah is opening the front door, but not both.
3. *Two-persons, one-action (non-simultaneous):* Two persons do one action, which they cannot perform simultaneously e.g., Martin is standing on the stool or Susan is standing on the stool, but not both.
4. *Two-persons, one-action (simultaneous):* Two persons do one action, which they can perform simultaneously e.g., Linda is looking out the window or Richard is looking out the window but not both.

We also introduced two new conditions. In one condition, we examined the situation where the shared referent was the action rather than the actor to test whether this sort of co-reference would also facilitate performance. This condition involved one of two people carrying out an action and again both individuals could not perform the action simultaneously. Finally, we added a fourth condition with no pragmatic constraint; one of

two people performs an action, but both could be performed simultaneously. As shown in Table 5.1, these conditions allowed us to examine the effect of co-reference on reasoning for cases with a pragmatic constraint (conditions 2 and 3) and without (conditions 1 and 4) and for cases where the co-referent is the actor (condition 2) and the action (condition 3 and 4).

Method

Participants

We tested individually 30 undergraduates from Princeton University in return for course credit. The participants were 15 men and 15 women. They ranged in age from 18 to 23 years with a mean age of 20.

Materials and Design

Participants were presented with premises in the form of a biconditional or an exclusive disjunction. The second premise was a categorical assertion or categorical denial of either the first or second proposition in the preceding premise. There were accordingly 8 forms of inference, as described in Table 5.2.

In addition, there were four types of semantic content, as presented in Table 5.1. The content varied the number of shared referents (i.e., people carrying out actions) contained in the first premise, and whether the two actions could be performed simultaneously or not. We constructed 32 problems concerning people carrying out various actions. In each case, the major premise contained the same number of words. We also included four filler problems which were inclusive disjunctions and the valid response was that nothing follows. The full set of materials is presented in Appendix 5. The design was therefore 2 forms of major premise x 4 types of inference x 4 types of

semantic content. Participants acted as their own controls and carried out the 36 inferences in a different random order.

Table 5.2: The eight types of inference examined in Experiment 7 together with the correct conclusions.

		Minor Premise			
Major Premise		P	Q	not-P	not-Q
If and only if not P then Q	∴	not Q	not P	Q	P
P or Q but not both	∴	not Q	not P	Q	P

Procedure

The problems were presented on a computer screen. Participants began by reading on-screen instructions. They then completed 5 practice problems followed by 32 experimental problems and 4 fillers. For each problem, the premises were presented together on the screen, followed by the question:

What conclusion, if any, can you draw?

The participants responded by typing their answer, and their latencies were measured from the presentation of the premises to the first key press. They were not told that their responses were being timed.

Results and Discussion

Form

Table 5.3 presents the percentages of correct responses to the eight forms of inference (collapsing over their contents). As the table shows, participants were not more accurate in reasoning from the one-model conditional (77% correct) than from the two-model disjunction (81% correct) (pace Johnson-Laird et al., 1992; Schaeken et al., 1995) probably because the conditional contained a negated clause. As expected, participants were more accurate when the categorical premise matched an event that was represented explicitly in the models of the first premise. When the categorical premise was negative and concerned the first event in the preceding premise, Not-A, the participants were more accurate in reasoning from the biconditional (91%) than the disjunction (68%). But, when the categorical premise was affirmative, A, they were more accurate in reasoning from the disjunction (94%) than from the biconditional (78%; McNemar tests, chi-squared = 17.36, $p < .0001$; chi-squared = 15.04, $p < .0001$ respectively). Participants performed equally well on both descriptions when the categorical premise affirmed the second proposition, B, (94% for both, McNemar Test, chi-squared = 0.07, $p = 1$). We predicted these three results on the basis of the mental models that participants construct, but they might reflect the surface matching of clauses in the premises. One result, however, is more readily explained in terms of models. When the categorical premise negated the second proposition, not-B, participants were more accurate in reasoning from the disjunction (68%) than from the biconditional (46%; McNemar Test, chi-squared = 13.8, $p < .0002$). Not-B mismatches the clauses in both sorts of premises. It is also not represented explicitly in the initial models of either premise. But, reasoners may find it easier to flesh out the disjunction, which already contains two mental models, than to

flesh out the biconditional which is represented by just one explicit mental model (Schaeken et al., 1995).

Table 5.3 The percentages of correct responses to the eight forms of inference in Experiment 7. Inferences are in bold where the categorical premise is explicitly represented in the initial models of the first premise.

		Categorical Premise				
First Premise	Models	*A*	*Not-A*	*B*	*Not-B*	Total
If and only if not-A then B	not-A B	78	**91**	**94**	46	77
	. . .					
A or B but not both	A	**94**	68	**94**	68	81
	B					

We measured latencies from the initial presentation of the premises to the first key press and we analysed the latencies for correct responses only. As Table 5.4 shows, response times varied across the different forms of inference. Responses were slower to biconditionals (mean = 12.5 secs) than to disjunctions (mean = 9.7 secs, Wilcoxon Test, z = 5.5, p <.0001) probably because the conditional contained a negated clause. As predicted, responses were faster when the categorical premise asserted that one of the propositions in the previous premise was true (mean = 9.9 secs) than when it denied a proposition (mean = 13.6 secs, Wilcoxon Test, z = 6.4, p < .0001) and they were faster when the categorical premise referred to the first proposition of the previous premise

(mean = 10.6 secs) than when it referred to the second proposition (mean = 11.5 secs, Wilcoxon Test, $z = 2.1$, $p < .04$). The interactions were not significant.

Table 5.4 The response times in seconds for correct responses to the eight forms of inference in Experiment 7. The times are in bold where the categorical premise is explicitly represented in the initial models of the first premise.

		Categorical Premise				
First Premise	Models	*A*	*Not-A*	*B*	*Not-B*	Total
If and only if not-A then B	not-A B	12.7	**10.7**	**12.4**	15.8	12.5
	. . .					
A or B but not both	A	**7.6**	12.2	**9.1**	10.8	9.7
	B					

Content

Table 5.5 shows the percentage of correct responses and the response times for each of the four types of semantic content. Participants drew more accurate inferences in the one-person, two-actions condition (82%) than in the two-persons, two-actions condition (77%, McNemar Test, $chi^2 = 2.86$, 1-tailed $p < .05$) replicating Bouquet & Warglien's (1999) finding. However, there was no difference between the other conditions. When the co-referent was an action there was no improvement in performance.

We also measured the latencies for correct responses. Participants were faster at reasoning from the two-person, one-action, non-simultaneous (10.9 secs) and simultaneous problems (10.2 secs) than from the two-person, two-action problems (11.6

secs; Wilcoxon test, $z = 2.33, 2.81$, 1-tailed $p < .01$, $p < .003$, respectively, excluding the results of two outliers). They were not significantly faster in reasoning from the one-person, two-action (11.5 secs) than from the two-person, two-action problems (11.6 secs; Wilcoxon test, $z = 0.02$, $p = .98$). Responses were also no faster where the action could not be performed simultaneously by two people (two-person, one-action, non-simultaneous condition = 10.9 secs) than when it could (two-person, one-action, simultaneous condition = 10.2 secs).[12] The interactions between the type of content and type of inference were not significant.

Table 5.5 The percentages of correct answers and the response times for each of the four types of semantic content in Experiment 7

	Correct Responses %	Response Time (secs)
Two-persons, two-actions (simultaneous):	77	11.6
One-person, two-actions (non-simultaneous):	82	11.5
Two-persons, one-action (non-simultaneous):	79	10.9
Two-persons, one-action (simultaneous):	79	10.2

Summary

Overall, the results corroborate the predictions made by the model theory. The findings provide further support for the conjecture that reasoners build models that make some information explicit and leave other information implicit.

[12] Responses were also faster to the two-persons, one-action (simultaneous) problems (mean = 10.7 secs) than to the one-person, two-actions (non-simultaneous) problems (12.2 secs; Wilcoxon test, $z = 2.56$, $p = .01$).

Responses were less accurate and slower for the biconditional which initially require reasoners to construct just one explicit model than for disjunctions which require two explicit models from the outset. The result probably arises because of the extra difficulty in reasoning with negated conditionals (e.g., Evans, Newstead & Byrne, 1993). Consistent with previous findings (Schaeken et al., 1995), responses were faster and more accurate when the categorical premise matched an item explicitly represented in the models of the first premise. This matching allows the models of the two premises to be integrated more easily (Schaeken et al., 1995). Some previous research has found that affirmative inferences are easier than denial inferences from a disjunction, as we do (see also, Johnson-Laird, Byrne & Schaeken, 1992) whereas other studies have not (Schaeken et al., 1995; see also Evans et al., 1993). The difference may lie in the form of the premises. Schaeken et al. (1995) used the form:

P or else Q

In contrast, we and Johnson-Laird et al. (1992) used the form:

P or Q but not both

The addition of the clause 'but not both' is consistent with the following model:

¬ P ¬ Q

If reasoners incorrectly build this model, they may err on the denial inference.

The finding that responses were faster when the categorical premise referred to the first proposition is consistent with the suggestion that forward inferences (i.e., P ∴ Q and not P ∴ not Q) are easier from conditionals of the form "if then" than backward inferences (i.e., not Q ∴ not P and Q ∴ P) (Evans, 1993). Our results suggest that reasoners also find it easier to reason forwards from a disjunction. In the temporal order effect we also found that people may use the first event to recruit an alternative. The

finding may relate to a more general primacy effect in cognition, perhaps because the first event forms the foundation in a representation (Byrne et al., 2000; Gernsbacher, 1990; Tversky & Kahneman, 1982).

The results also show that co-reference can help the process of reasoning. Responses were more accurate to problems in which one person acts than when one of two people act. This result is consistent with the findings of Bouquet and Warglien (1999) and we have demonstrated the effect over a wider range of inferences. Problems in which two persons carry out one action, whether or not they can perform it at the same time, did not significantly improve response accuracy but elicited faster responses than problems in which two persons carry out two different actions. The findings cannot be attributed purely to a pragmatic effect which eliminates the models in which both actions occur because responses are also faster in the two-person, one action simultaneous condition where the two actions can co-occur. The effect may occur because co-reference yields more concise models which contain fewer elements. The faster response times occurred where the co-referent was the action but not when it was the actor, perhaps because actions require a richer representation.

Co-reference may have a different effect on the nature of the mental models for conditionals and disjunctions. Because conditionals may be represented initially by a single model, co-reference may be represented within one model. For example, the following premise:

If and only if, Sarah is not sitting in the arm-chair then Sarah is opening the front door.

allows co-reference to be represented within one initial model:

Sarah: ¬ arm-chair front door

. . .

In contrast, the same sentence described as a disjunction:

Sarah is sitting in the arm-chair then Sarah is opening the front door, but not both may require co-reference to be represented across two models:

Sarah: arm-chair

front door

In the next experiment, we examine how people reason from two model problems, some of which contain co-referents within one model and some of which contain co-referents across models. We also varied the number of co-referential relations contained in the first premise. A further aim of the next experiment was to test the idea that the form of an inference may lead to systematic fallacies.

Experiment 8: Illusory inferences and co-reference

The aims of the experiment were twofold. The first aim was to examine the idea that mental models may give rise to illusory inferences, that is, inferences which are compelling and which most people make but which are invalid. People may draw these inferences because they fail to think about what is false. For example, recent studies have shown that most people draw illusory inferences if they fail to represent the way in which a conditional is false (e.g., Johnson-Laird & Savary, 1999). In one experiment, participants were given the following problem:

If there is a king in the hand then there is an ace in the hand, or else if there is not a king in the hand then there is an ace in the hand.

There is a king in the hand.

What follows?

All the participants drew the conclusion that there is an ace in the hand. Mental models yield this conclusion, but it is wrong. In fact, the sentential connective or else implies that one of the conditionals must be false, that is, that its antecedent is true and its

consequent is false. But, if the first conditional is false, then there is a king and no ace. Similarly, if the second conditional is false, then there is no king and no ace.

Our aim was to test the idea that illusory inferences may arise when people reason about conjunctions as well as conditionals. There are three ways in which a conjunction, e.g., P and Q, can be false, i.e.,

¬ P ¬ Q
P ¬ Q
¬ P Q

but reasoners can have difficulty in representing all of these possibilities (Byrne & Handley, 1992; Handley, 1996). We suggest that this difficulty may give rise to illusory inferences. We examined what inferences people would draw from an exclusive disjunction of the form:

Either P and Q or otherwise R and S.

If one of the conjunctions is true, then the other must be false. According to the principle of truth, reasoners construct two mental models of such a premise:

P Q
R S

It follows that given the categorical premise P, reasoners will infer that Q, not-R and not-S follow. Similarly, given the categorical premise not-P they will infer that not-Q follows. These inferences, however, are illusions. When falsity is taken into account, the disjunctive premise is consistent with six different possibilities, which we present here in *fully explicit* models:

P	Q	¬ R	S
P	Q	R	¬ S
P	Q	¬ R	¬ S
¬ P	Q	R	S
P	¬ Q	R	S
¬ P	¬ Q	R	S

These models show that the four previous inferences are invalid. Given the premise, P, for instance, Q and not-Q are both possible, and likewise R, not-R, S and not-S are all possible (see rows 1, 2, 3 and 5). Similarly, given the premise ¬ P, both Q and ¬ Q are possible (see rows 4 and 6). In contrast, given the categorical premise, not-P, participants can correctly infer that R and S follow: these conclusions follow from the mental models above, and they also follow from the fully explicit models (see rows 4 and 6).

We also varied the semantic content so that in some cases P (e.g., Liz is sitting) and R (Liz is standing) could not meaningfully occur together, (eliminating models in rows 2 and 5 above) and Q and S could not co-occur (eliminating models in rows 1 and 4 above). This left two remaining models, rows 3 and 6, which were consistent with the two initial mental models:

P	Q	¬ R	¬ S
¬ P	¬ Q	R	S

This study also allowed us to test a new prediction about the effect of the number of models on reasoning performance. There is ample evidence to show that a problem becomes more difficult as the number of initial models increases (e.g., Johnson-Laird, Byrne & Schaeken, 1992). In our experiment, the first premise was the same in all problems and is consistent with two initial models. However, on half our trials only one model must be considered (i.e., when given information about P and asked if Q follows –

same-model condition) whereas on half the trials both models must be considered, (i.e., when given information about P and asked whether R follows – *different-model condition*). We predicted that problems would be easier if the number of models to be considered was one rather than if it was two.

Our second aim was to examine the effects of co-reference on these inferences. Experiment 7 showed that co-reference improved accuracy and reduced response times, at least for certain inferences. The present experiment followed up this effect. Table 5.6 presents the five sorts of semantic contents used in the experiment.

Table 5.6: The five sorts of semantic content in Experiment 8, and the number of fully explicit models compatible with each content.

---

1. *Four persons : (six models)*
   Either Jane is kneeling by the fire and Sean is looking at the TV or otherwise Mark is standing at the window and Pat is peering into the garden.
2. *Two persons, one per model: (six models)*
   Either Jane is kneeling by the fire and she is looking at the TV or otherwise Mark is standing at the window and he is peering into the garden.
3. *Two persons, two per model, simultaneous: (six models)*
   Either Jane is kneeling by the fire and Mark is standing at the window or otherwise Jane is looking at the TV and Mark is peering into the garden.
4. *Two persons, two per model, non-simultaneous: (two models)*
   Either Jane is kneeling by the fire and Mark is looking at the TV or otherwise Jane is standing at the window and Mark is peering into the garden.
5. *One person: (two models)*
   Either Jane is kneeling by the fire and she is looking at the TV or otherwise she is standing at the window and she is peering into the garden.

First, these varied the number of shared referents (i.e., people carrying out actions) contained in the first premise. We predicted that co-reference would again facilitate performance and that this facilitation would be greater as the number of shared referents increased.

Second, the contents also varied whether the co-referents occurred within one model (i.e., two-persons, one per model and one-person conditions) or were common to two alternative models (i.e., two-persons, two per model, simultaneous and non-simultaneous conditions and the one-person condition). For example, in the two-persons, one per model condition, P and Q share a common referent:

Either Rose is sitting on the sofa and she is blowing out the candles or otherwise Jim is standing beside the piano and he is drinking through a straw.

and this can be represented within one model:

Rose: sofa candles

Jim: piano straw

However, in the two-persons, two per model conditions, P and R share a common referent, e.g.,

Either Rose is sitting on the sofa and Jim is standing beside the piano or otherwise Rose is blowing out the candles and Jim is drinking through a straw.

this must be represented by making the referents explicit in the two initial models, e.g.,

Rose sofa Jim piano

Rose candles Jim straw

We aimed to examine whether the effect would differ depending on whether co-reference occurred within or across models.

Third, the problems varied whether the categorical premise and question shared a common referent. When the co-referent is represented within models (e.g., Rose is sitting

and blowing out candles or else John is standing and drinking), then the same-model problems, i.e., *P, ∴ Q?* , refer to one individual (e.g., Rose is sitting, Is she blowing out candles?) but the different-model problems, i.e., *P, ∴ R?* , refer to two (e.g., Rose is sitting, Is Jim standing?). In contrast, when the shared referent is across models, then the same-model problems refer to two individuals and the different-model problems refer to one.

The first three sorts of content in Table 5.6 are consistent with all six of the fully explicit models above. However, in the other two cases, the actions P and R cannot occur simultaneously and Q and S cannot occur simultaneously. This pragmatic constraint rules out the models in which P and R occur together, and Q and S occur together. The premise is therefore consistent with just two fully explicit models:

P Q ¬ R ¬ S

¬ P ¬ Q R S

These models yield the same conclusions as the mental models described above, i.e.,

P Q

R S

Logically these conclusions are still illusions but pragmatically they are correct.

Method

Participants

We tested individually 35 participants (25 paid members of the public recruited through national newspaper advertisements and 10 postgraduate volunteers from the University of Dublin, Trinity College). The participants were 14 men and 21 women. They ranged in age from 18 to 78 years with a mean age of 36.

Design

Participants were presented with exclusive disjunctions of the form:

Either P and Q or otherwise R and S.

Each disjunction was paired with a categorical premise, either asserting or denying its first proposition, P. The participants had to evaluate the validity of a one-clause conclusion, either Q which was in the same initial model as P *(same-model condition)* or R which was in a different initial model from P *(different-model condition)*. There were accordingly four forms of inference, as follows:

P, ∴ Q?

P, ∴ R?

Not -P, ∴ Q?

Not -P, ∴ R?

In addition, there were five types of semantic content, as presented in Table 5.6. The content varied whether the number of shared referents (i.e., people carrying out actions) contained in the first premise was four *(four-persons condition)*, two *(two-persons conditions)* or one *(one-person condition)*. It also varied whether the co-referential relations occurred within one model *(two-person, one per model* and *one-person conditions)* or between models *(two-persons, two-per-model, simultaneous and non-simultaneous conditions and one-person condition)*. The design was 4 types of inference x 5 types of semantic content. Participants acted as their own controls and carried out the 20 inferences in different random orders.

Materials

We constructed five sets of materials each containing 20 problems concerning individuals doing four different actions. Each problem contained the same number of words. The materials were rotated so that they were presented equally often with each of the five sorts of semantic content. One full set of materials is presented in Appendix 5.

Procedure

The problems were presented on a computer screen using the computer package, E-Prime. Participants began by reading on-screen instructions. They then completed 5 practice problems followed by 20 experimental problems. For each problem, the premises and questions were presented one-by-one on the screen. After reading each premise, the participant pressed a key to get the next premise or question. The premises stayed on the screen until the question had been answered. Participants responded to the question by pressing one of three keys: yes, no or cannot tell. The program recorded separately the time that it took participants to read each of the premises and to answer the question. The participants were not told that their responses were being timed.

Results and Discussion

Form

Table 5.7 presents the percentages of the "Yes", "No" and "Cannot Tell" responses to the main sorts of problems. As the table shows, the participants succumbed to the illusory (six model) problems (10% correct), but performed well on the (two-model) control problems (78% correct). Of the 35 participants, 34 were less accurate on the illusions than on the control problems (Sign Test, $p < 1$ in 900 million). The pattern of responses was comparable for both types of problem, and so we suspect, as predicted, that the

participants constructed just two mental models for the six-model and the two-model problems. A similar pattern of results occurred in the participants' responses to the first problem that they encountered. They were more accurate with the two-model problems (81% correct) than with the six-model problems (31% correct). This result suggests that the illusions occurred spontaneously and not as a result of the development of a mental set.

Table 5.7: The percentages of the "Yes", "No" and "Cannot Tell" responses in Experiment 8. The predicted responses are shown in bold, and the correct responses are underlined.

	Six models *			Two models **		
1. P & Q, or R & S	Yes	No	Can't Tell	Yes	No	Can't Tell
P						
∴ Q	**87**	4	9	**84**	6	10
2. P & Q, or R & S						
P						
∴ R	22	**69**	9	7	**87**	6
3. P & Q, or R & S						
Not-P						
∴ Q	13	**75**	12	12	**77**	11
4. P & Q, or R & S						
Not-P						
∴ R	**78**	15	7	**83**	10	7

* Four-person, Two-person, one per clause and Two-person inclusive conditions
** Two-person-exclusive and One-person conditions

We analysed response latencies based only on the responses predicted by the mental model theory. We didn't analyse correct responses to the illusory problems because these were significantly lower than chance and it was therefore difficult to know whether

participants were correctly fleshing out their models or whether they got the right answer for the wrong reason. As Table 5.8 shows, participants responded faster when the categorical premise was affirmative (mean = 7.9 secs) then when it was negative (mean = 8.5 secs, Wilcoxon Test, $z = 4.12$, $p < .00005$). The result is consistent with previous research which suggests that models can be combined more easily if there is an explicit match for the categorical premise in the initial models of the previous premise (Schaeken et al., 1995).

Table 5.8 The response times in seconds for correct responses to the four forms of inference in Experiment 8.

	Type of Inference			
	P ∴ Q?	P ∴ R?	not-P ∴ Q?	not-P ∴ R?
Response Times (secs)	6.2	9.8	7.9	9.0

Participants responded faster in the same-model condition (i.e., when the conclusion to be evaluated is Q) (mean = 7.0 secs) than in the different-model condition (i.e., when the conclusion to be evaluated is R) (mean = 9.4 secs, Wilcoxon Test, $z = 5.16$, $p < .00005$). This result provides the first demonstration that problems may be more difficult when the number of models to be considered is two than when it is one, even when the number of initial models is constant. The interaction between these factors was also significant. The difference in response times between the same and different-model conditions was greater when the categorical premise was positive (mean difference = 3.6 secs) than when it was negative (mean difference = 1.1 secs, Wilcoxon Test, $z = 5.16$, $p < .02$). In other words, response times were faster when the conclusion was explicitly

represented, i.e., in the same-model condition when the categorical premise was affirmative and in the different-model condition when it was negative.

Content

There was no reliable difference between the five semantic conditions in response accuracy but participants made more of the predicted responses in the one-person condition (84%) than in the four-persons condition (75%) and the two-persons, one per clause condition (75%, $chi^2 = 3.2$, 1-tailed $p < .04$ for both) but not reliably more than in the other conditions.

There was no difference between the conditions in the time taken to read the first premise. But as shown in Table 5.9, a significant difference in response times to the question occurred across conditions (Friedman Test, chi-squared = 10.66, p = .03). There were three principal results:

1. Responses were faster when the first premise referred to one individual (1-person condition, mean 6.8 secs) than when it referred to 2 or 4 individuals; (2-persons, one per model condition, mean = 8.8 secs, Wilcoxon Test, $z = 2.76$, $p < .003$; 2-persons, two per model, (simultaneous) condition, mean = 8.7 secs, Wilcoxon Test, $z = 3.35$, $p < .0004$; 2-persons, two per model, (non-simultaneous), mean = 8.3 secs, Wilcoxon Test, $z = 2.45$, $p < .01$; 4-persons condition, mean = 8.3 secs, Wilcoxon Test, $z = 3$, $p < .002$). The differences between the two and four person conditions were not significant.

2. There was an interaction between the type of semantic content and whether the conclusion to be evaluated was Q (same-model condition) or R (different-model condition). The difference between the same and different-model conditions was lower in the 1-person condition (mean difference = 1.1 secs), than it was in the 4-person condition (mean difference = 3.9 secs, Wilcoxon Test, $z = 2.28$, $p < .02$) and the 2-person, one

clause condition (mean difference = 4.0 secs, Wilcoxon Test, $z = 3.3$, $p < .0005$). This suggests that problems in the different-model condition may be particularly difficult where the categorical premise and question refer to two individuals, as is the case in the four-person and the two-person, one per clause conditions than when both refer to one. The other interactions were not significant.

3. There was no difference in response times between the two-model problems and the six model problems.

Table 5.9 Percentage of correct and predicted responses and question response times in each of the five semantic conditions

	4 persons	2 persons 1 per model	2 persons 2 per model simultaneous	2 persons 2 per model non-simultaneous	1 person
Response Accuracy					
Predicted responses	75	75	81	82	84
Correct responses to illusory problems	14	8	9	-	-
Response Times					
Question					
Q	6.6	7.0	7.5	7.8	6.3
R	10.5	11.0	10.0	8.8	7.4
Total	8.3	8.8	8.7	8.3	6.8

Overall, the results corroborated the occurrence of illusory inferences, and reasoners seem likely to construct just two models of disjunctions of the form:

Either P and Q or otherwise R and S.

They overlook the different ways in which the conjuncts could be false in the case of the six model problems, i.e., those with a content that does not eliminate any of the possibilities.

The latencies of the responses also bear out our conjecture that co-reference can speed up inferences. The results show that in the one-person condition, when all four clauses share a common referent, the response latencies are significantly reduced. Reasoners can construct more concise models in this case. The effect was strongest in the different-model condition, that is, when reasoners had to consider an alternative model to the one referred to in the categorical premise. The difficulty in thinking about two models appears to be greatest when the categorical premise and question refer to different individuals, that is, in the 4-person and the 2-person, one per clause conditions.

General Discussion

The results of our experiments support the prediction that the verbal description of a problem and the co-referential links within it, can affect the mental representations that people construct. The mental model theory predicts that when people represent a premise, they do so in accordance with the principle of truth. They construct a representation that makes only some information explicit. Experiment 7 corroborated the principle. It showed that reasoners draw different conclusions from given information, depending on whether it is expressed as a biconditional or an exclusive disjunction (Johnson-Laird et al., 1992). They were more accurate when the categorical premise matched an item which was explicitly represented in their models of the previous premise. This match allows the models to be integrated without the need to flesh them out (Schaeken et al., 1995). When the categorical premise did not contain a match, participants were more accurate in reasoning from the disjunction than the conditional. Reasoners may find it easier to flesh

out the disjunction which already contains two explicit models than to flesh out the conditional which only contains one (Schaeken et al., 1995). While the former finding could be explained as a matching bias (Evans, 1998), the latter finding cannot.

Likewise, Experiment 8 showed that the failure to represent falsity can lead reasoners to make illusory inferences. The results suggest that from a premise of the form:

Either P and Q or otherwise R and S

reasoners construct just a subset of possible models. They construct models which represent only what is true in the premise:

P Q

R S

Given the categorical premise P, they falsely conclude that Q follows. They also falsely conclude that not-R follows, either because when R is missing from their representation, they tend to describe it with its negation (Legrenzi, Legrenzi, Girotto & Johnson-Laird, 2000) or because they incorrectly use footnotes to negate what is contained in the first model (see chapter 2). They do not consider all of the possible ways in which a conjunction of two events can be false (see also Byrne & Handley, 1992; Handley, 1996).

The results rule out some explanations which have been proposed to account more generally for illusory inferences from a disjunction of two conditionals, i.e.,

If P then Q or else if R then S.

First, our problems were not unduly complex or ambiguous. In fact, most participants reported that they did not find them very difficult. Second, it is unlikely that participants treated the premises as a conjunction of four elements instead of a disjunction because when given the premise P, they consistently affirmed Q but denied R. Third, our premises do not include conditionals and therefore cannot be explained by an alternative

interpretation of a conditional. In contrast to 'if', the semantics of 'and' are unambiguous. Fourth, it cannot be argued that when reasoners think about one disjunct they forget about the other. Half of our inferences required reasoners to keep both models in mind, i.e., those in the different-model condition. The illusory inferences which arise from a disjunction of conjunctions cannot be explained by these alternative explanations. Formal rule theories, which propose that reasoners make inferences by relying on a mental repertoire of abstract inference rules, may also have difficulty in accounting for the findings (Braine & O'Brien, 1991; 1998; Rips, 1994). They propose that errors occur because reasoners misinterpret the premises or misapply the rules of inference. The meaning of the sentence 'P and Q or else R and S' is unambiguous and the illusory inferences occurred across a range of materials and therefore it is unlikely that reasoners misinterpret the premises. The prevalence of these illusory inferences also suggests that it is unlikely that reasoners misapply their rules of inference. At present therefore, formal rule theories cannot easily explain these findings. However, the mental model theory predicts their occurrence based on the principle of truth and our results corroborate the prediction.

The response latencies are also consistent with the theory that reasoners construct representations which make just some information explicit. In Experiment 7, participants were faster at drawing inferences from a biconditional which is initially represented by just one explicit model than for a disjunction which requires two. In Experiment 8, participants were faster at drawing an inference when the categorical premise and conclusion were represented explicitly in the same model, than when they were represented in different models and both had to be considered. This suggests that response latencies are shorter for inferences which require less information to be studied.

In both experiments, participants were faster when the categorical premise contained an explicit match in the models of the first premise. This should allow the models to be integrated more easily. Also, in Experiment 8, responses were faster when the conclusion was explicitly represented than when it had to be fleshed out. Initially representing less information may be advantageous but subsequently extra work will be required if the models need to be fleshed out to be explicit.

Overall, the results of our experiments provide support for the argument that people build models that make some information explicit and leave other information implicit. The building of partially explicit models influenced the accessibility of different counterfactual models in the card selection scenario leading to a temporal order effect. In this chapter, we have shown that it can also influence the accessibility of inferences in a deductive reasoning task.

In our experiments, we also examined another factor that influences mental models: the occurrence of co-reference within the premises. Experiment 7 showed that co-reference within a premise improved accuracy but not speed of response when the shared referent was an actor (see also, Bouquet & Warglien, 1999). When the shared referent was an action, it speeded up the process of inference although it did not improve accuracy. Experiment 8 showed a facilitation in response time when the shared referent was common to all four clauses. The effect was greater when the inference required the reasoner to consider two models. The inferential task in such cases places a bigger load on working memory, and co-reference evidently facilitates matters.

This finding suggests that reasoners build a more concise representation when there are co-referential relations between the clauses of a premise. Co-reference may be accommodated quite easily in a conjunction because it is often quite easy to combine two items or events in a single model. The following example from one of our experiments:

Jane is kneeling by the fire and she is looking at the TV

can easily be imagined as one situation and may be represented as follows:

Jane: Fire TV

whereas the same sentence without co-referential relations:

Jane is kneeling by the fire and Tom is looking at the TV

requires a model with more elements:

Jane Fire Tom TV

Our results also showed a facilitation in response latencies when the co-referential relations occurred across two models. Disjunctive premises such as:

Either Jane is kneeling by the fire or she is looking at the TV

may also be represented concisely:

Jane: Fire

TV

The results have important implications for the representation of counterfactual conditionals. Co-reference enables us to link factual and imaginary possibilities. They allow us to represent which aspects of these possibilities are the same and where the departures occur. Research shows that people have difficulty in reasoning from situations that require reasoners to keep multiple models in mind, as we have demonstrated in this chapter. Yet we have seen, for example in chapter 4, reasoners readily generate counterfactual possibilities after they read a factual scenario, particularly one with a bad outcome. Our results suggest a possible reason for this apparent inconsistency. In Experiment 8, we demonstrated that co-reference greatly attenuated the time required to reason about two models. The strong co-referential links between factual and counterfactual possibilities may be the key which allows people to imagine these possibilities so easily.

The results provide insight into the nature of the mental models that reasoners construct and they provide further support for the idea that reasoners build models which are not fully explicit. One factor that may influence this is the verbal description of a situation.

The experiments provide a first exploration of how co-reference can influence a range of deductive inferences. Our computer program in chapter 2, relied on matching co-referents, e.g., players in a card game. Co-reference needs to be incorporated into any account of how people access counterfactual alternatives from a factual reality. Our studies make an initial step toward such an account.

Chapter 6 – Discussion

Our aim in this thesis is to examine the mental representations and cognitive processes underlying counterfactual thinking. In this chapter, we will consider how our work contributes to our understanding of this topic both directly and in the broader context of human cognition and emotion.

We used a range of methodologies to study how people imagine alternatives to reality. First, we built two computational models to simulate alternative theories of the processes underlying counterfactual thinking. The focus model simulates recent accounts of how people generate counterfactuals (e.g., Kahneman & Tversky, 1982b; Legrenzi et al., 1993). The theory assumes that people begin by constructing a model of the facts and then they mentally undo one of those facts. The context model assumes that people use the facts to recruit different possibilities which are either explicitly represented or recalled from past experience and these possibilities may provide an alternative to the facts. A change in the context can therefore influence the accessibility of different alternatives. Second, in two series of experiments, we used a scenario-based approach to examine the mutability of temporally ordered and causally related events, in the latter case, reasons and actions. This approach is the dominant method used in the literature on counterfactual thinking. Finally, in a third series of experiments we used an inference based task to examine the nature of the mental representations underlying counterfactual thinking and reasoning in general.

In this chapter, we will first summarise our findings and consider how they inform the understanding of how people generate counterfactual alternatives. Secondly, we will consider at a broader level the implications for the understanding of the processes underlying counterfactual thinking and human cognition, specifically, problem-solving,

decision-making and reasoning. We will outline some new questions that need to be addressed in understanding the processes underlying counterfactual thinking.

Summary of findings

The Temporal Order Effect

In chapters 2 and 3, we examined one factor that can influence counterfactual thinking, that is, the temporal order in which events occur. All events have a temporal relation to other events and our understanding of a series of events regularly requires that we keep this relation in mind. Indeed, time may provide one of the core categories of mental life (Miller & Johnson-Laird, 1976). The temporal order of events can influence the ease with which alternatives to reality come to mind. When people are presented with an independent sequence of events, they tend to mutate the most recent one even though undoing earlier ones could change the outcome (Miller & Gunasegaram, 1990). The effect has been replicated in a range of contexts including scenario-based descriptions of games (Byrne et al., 2000; Spellman, 1997), simulated examination situations (Miller & Gunasegaram, 1990) and judgements of the performance of a basketball team in a league played over 10 weeks (Sherman & McConnell, 1996). Our first aim was to computationally simulate two theories of the mental representations and cognitive processes underlying the temporal order effect. This enabled us to develop complete and coherent models of the theories, to demonstrate clearly the differences between them, and to derive novel predictions that distinguished them. Our second aim was to carry out empirical tests of the models.

The two computational models that we developed are based on what we term the focus theory and context theory of counterfactual thinking. The focus theory assumes that people focus on salient aspects of the factual situation (Kahneman & Tversky, 1982b;

Legrenzi et al., 1993; Seelau et al., 1995). These aspects are mentally undone, for example, imagine John didn't stop at the shops, or mentally altered so that they happened differently, for example, imagine he went to the shops 10 minutes earlier. People then *subsequently* draw on their background knowledge to decide what follows from the imagined antecedent. For example, if he had left earlier what would the consequences have been?

The context theory assumes that people *begin* by drawing on their background knowledge, or on explicitly represented alternatives. These alternatives provide a contrast cases which are used to select an event to mutate (Kahneman & Miller, 1986).

Theories of Counterfactual Thinking

In the introduction, we considered two broad theories of the processes underlying counterfactual thinking. The first, which we term the *focus theory*, is based on recent accounts of event mutability (e.g., Legrenzi et al., 1993; Seelau et al., 1995). This theory assumes that the focus of people's counterfactual thoughts is on the representation of the facts of a situation. The process begins with a search through these facts for a mutable event, based on defining features of the events. For example, people will tend to focus on an event that is salient or is important to the story. One event is selected and mutated. The alternative may be that the event didn't happen (e.g., Bernard did not go to the ball) or it happened differently (e.g., Bernard left the ball early). Background knowledge must subsequently be consulted to ensure that the new alternative event is consistent with the remaining facts of the scenario and to simulate the outcome of the new scenario.

The second theory, which we term the *context theory* advances the ideas of norm theory (Kahneman & Miller, 1986). The theory proposes that people use the facts to select an alternative that is similar to the factual one. In our studies of the temporal order

effect these alternatives, which we term the context of possibilities, were explicitly described. However, as we suggested in our studies of reasons for acting, they may also be provided by background knowledge or previous experiences. The theory assumes that people recruit an alternative from these possibilities. One or some of the facts may be used to recruit an alternative, for example, the first fact in a temporal sequence or in a reason-action sequence. Antecedent events that are different in the factual and counterfactual scenarios selected to be mutated. The two theories differ in an important respect: the focus theory proposes people mutate an event before consulting the context of possibilities, whereas the context theory uses the context of possibilities to provide a contrast to the facts, and items that differ are selected to be mutated.

In chapter 2, we described two computational models to describe how these theories might account for the temporal order effect. Both programs take as input a set of facts and a description of the conditions that determine the outcome, which we termed the context of possibilities. The TempFocus model focuses on the most recent antecedent event, mutates it by replacing it with its negation and then consults the background conditions to determine what follows from the mutated event together with other facts in the scenario. The program mutates the last event regardless of the context. The TempContext model builds models to represent the context of possibilities. The program selects one possibility by matching the first fact to the context of possibilities. It compares this possibility to the facts and if it finds an alternative to one of the facts it mutates that fact. The description of the conditions that determine the outcome influences what information the program makes explicit in the context of possibilities and hence which possibilities are accessible from the facts.

We used one of the standard scenarios to simulate the temporal order effect in both models. In this scenario, the conditions that determine the outcome are explicitly

given, two players play a game in which they select cards from a deck and if both select the same colour card they win, i.e., both pick red cards or both pick black cards. Based on the tenets of the model theory (Johnson-Laird & Byrne, 1991), the program constructs models of the two winning possibilities. The program also builds a model of the facts, for example, John picked a red card and Michael picked a black card and they lost.

TempFocus selects the last antecedent, i.e., Michael picked a black card and it mutates it to produce Michael didn't pick a black card (which in the binary context it understands to be the same as Michael picked red). It then checks the context of possibilities to see what follows when both pick red and it discovers that they would have won. Hence, the counterfactual is 'If Michael hadn't picked black they would have won'.

TempContext uses the first fact, for example, John picked a red card, to select a matching possibility in the context of possibilities, i.e., both players pick red cards, and it eliminates the other possibility, i.e., both players pick black cards. Therefore, it mutates the second event. However, in some situations, a match for the first fact is not explicitly represented in the models of the winning conditions. In this case, the program searches for an explicit alternative to the first fact and hence mutates it. Thus the program produces the temporal order effect for some descriptions, but it produces a reversal for other descriptions (i.e., it mutates the first event).

In chapter 3, we reported the results of three experiments that we carried out to test our theory. We presented participants with versions of the card selection scenario described earlier. Because people only mentally represent what is true (Johnson-Laird & Byrne, in press), we proposed that it should be possible to manipulate the representation of the winning conditions by changing the way that they are verbally described. In Experiment 1, we described the winning conditions as a disjunction of red cards, i.e.,

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

which may be represented as follows:

John red win

Michael red win

The facts were that both players picked black cards and lost. TempFocus mutates the second event in this condition. But, because the models of the winning conditions contain only an alternative to the first fact, TempContext mutates the first event. This reversal of the standard temporal order effect has never previously been observed. In accordance with the predictions of the context theory, the results confirmed that people mutated the first event in this case. In Experiment 2, we gave participants another version of the card selection scenario, but this time we described the winning conditions as a disjunction of black cards, i.e.,

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

which may be represented as follows:

John black win

Michael black win

Again, both players pick black cards and lose. In contrast to the previous example this representation now contains an explicit match for the first fact and the context theory predicts that people should mutate the second event. Again the results support this prediction. The experiment shows that the reversal in the 'red disjunction' condition occurs not because of the meaning of the winning conditions but because of what information they make explicit. Finally, in Experiment 3, we generalised the effect to an inclusive disjunction. We gave participants a version of the same scenario, with the winning conditions described as follows:

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

which may be represented by three explicit models:

John red		win
	Michael red	win
John red	Michael red	win

and once again the facts were that both players pick black and lose. These models give rise to the same predictions as in Experiment 1. TempFocus mutates the second event, whereas TempContext mutates the first event because the context of possibilities contain no explicit match for it. Once again, the results corroborate the predictions of the context theory.

In each of the experiments, participants attributed guilt and blame to the second player regardless of which event was mutated. The finding suggests that the representation of the context of possibilities and the mutability of events are not the basis for these attributions.

Overall, the results of our experiments support the context theory that the representation of the background conditions plays a central role in counterfactual thinking, at least in some situations. In our experiments, the facts remained the same and the winning conditions remained the same. However, the verbal description of those conditions influenced what aspects of the winning conditions were made explicit and hence which counterfactual alternatives were most available. Our study provides the first demonstration that minor changes in the verbal description of a scenario, such as the reference to red cards or black cards, can influence the mutability of events.

The effect of reasons for acting on counterfactual thinking

In our second series of experiments, we studied how the representation of the background conditions may influence the mutability of actions. We suggest that reasons may form the

context in which actions occur and our aim was to examine whether reasons may affect the mutability of actions.

The tendency to mentally undo intentional actions is a pervasive aspect of counterfactual thinking which has been demonstrated in scenario-based studies (e.g., Girotto et al., 1991), in games (Markman et al., 1995) and in real life traumatic situations (Davis et al., 1995). Furthermore, the focus on actions may have particularly important functions, for example, in enhancing perceived control and in helping people to plan for the future. Mentally undoing actions is also associated with specific emotions such as regret and guilt.

According to the focus theory, people mutate actions because they are salient or important to the story (Legrenzi et al., 1993). In this case, the reason for acting should not affect the mutability of the action. In contrast, the context theory proposes that the mutability of an action depends on the context in which it occurs. Actions may be more mutable in some contexts than others depending on what possibilities the context brings to mind. In chapter 4, we described three experiments which examined the effect of reasons for acting on the mutability of actions.

The idea that reasons may constrain the mutability of actions was motivated by an earlier finding that causes tend to be more mutable than their effects (Wells, Taylor & Turtle, 1987). Reasons may explain or rationalise an action in the same way that causes explain physical events in the world. However, despite the similar role played by reasons and causes, they may differ in nature. We suggest that reasons and causes may have a different underlying mental representation. According to the mental model theory, people may represent causal relations by keeping in mind the different possibilities that are consistent with the relation. Causes may be represented most frequently as a strong relation, that is, one which is necessary and sufficient for its effect, for example:

Cause	Effect
¬ Cause	¬ Effect

In contrast, reasons may be represented more often as a weaker relation, that is, one which is unnecessary or insufficient for acting, for example:

Reason	Action
Reason	¬ Action
¬ Reason	¬ Action

The relation between reasons and actions may be perceived as more complex and more variable than causes and their effects which tend to co-vary in the world. We suggest that these representations form the context of possibilities against which the facts are perceived.

This difference led to important predictions about the effect of reasons on the mutability of actions. If a causal relation is strong, then undoing the cause is sufficient to undo the action and it is difficult to imagine the cause occurring without its effect. In this case, people may prefer to mutate causes than their effects. In contrast, if a reason is insufficient for an action then it is possible to imagine that someone has a reason but does not act on it. For example, a person may have an intention to act but the opportunity may not yet be present or they may have a stronger reason for doing an alternative action. If a reason is not necessary to act, then mutating it is not sufficient to undo the action. People frequently have many reasons for doing the same action and so would still have acted in the absence of any specific one.

Our suggestion was that the nature of the mental representation of reasons influences whether people focus on reasons or actions in their counterfactual thoughts. We reported the results of three experiments designed to test this proposal. The experiments examined whether people mutate actions more than reasons and they

compared the effect of different types of reasons, i.e., internal reasons, such as desires and goals and external reasons such as obligations, on the mutability of actions. In the experiments, participants read scenarios about a series of actions and events involving a central character. We focused on one of the actions and studied how often people mutated it when accompanied by different types of reasons. In Experiment 4, we compared the mutability of an action, that is, a visit to the gym, when accompanied by an internal immediate reason ('Tom likes to go to the gym') and when accompanied by an internal enduring reason ('Tom is trying hard to lose weight'). We refer to these as own-sake desires and plan-based reasons respectively. We compared these to a third control condition in which the reason was replaced by a filler ('It is quite a large gym'). The results of the experiment clearly showed that reasons are not more mutable than actions in contrast to the finding that causes are more mutable than their effects (Wells, Taylor & Turtle, 1987). The results also showed that the reasons differed in their effect. When compared to the control condition, the plan-based reason reduced the mutability of the action whereas the own-sake desire did not. The findings support our suggestion that reasons tend to be represented as a weaker relation than causes and that reasons may differ in their representation.

In Experiment 5, we directly compared the effect of reasons and causes on the mutability of actions and events in a scenario-based study. The reasons for acting were internal immediate reasons, such as short-term goals and we compared these to reasons for events, that is, the purpose which gave rise to the event. In addition, we examined causes for acting, such as errors or distractions and causes for events. Overall, actions were mutated more than events. When actions were accompanied by reasons, people tended still to focus on the action; however, when accompanied by a cause they tended to shift their focus somewhat to the cause. The results corroborate the idea that reasons and

causes are represented differently and that actions are highly mutable because they are perceived as unconstrained by events in the world (Girotto et al., 1991).

In Experiment 6, we studied another type of reason, that is, obligations imposed on us by others. Obligations usually provide a sufficient condition for acting and hence might be expected to have a greater effect on the mutability of actions than desires and goals. Indeed, the results bear out this conjecture. The same action was mutated less when preceded by an obligation than an internal reason or no reason. However, obligations were still mutated less than actions.

The results of our experiments show that attributions of guilt and self-blame did not vary depending on whether the item mutated was the action or the cause or reason that lead to it. While these judgements may depend on counterfactual thoughts, our results suggest that they are not sensitive to these shifts in the focus of those thoughts.

Both reasons and causes may help us to make sense of the world by providing us with explanations of how things came about. Despite this, they have a different effect on the way people imagine how events might have happened differently. This finding supports the predictions of the context model that the representation of the context is important in understanding how people generate counterfactuals in the domain of human action. People may understand a causal relation by representing the different possibilities that are consistent with the relation and these possibilities may provide the context. People may use the facts of a situation to recruit one of these alternatives. With causes, people are more likely to imagine an alternative in which neither the cause nor effect occurs. With reasons, people are more likely to imagine an alternative in which the reason occurs but the action does not.

The results generalise support for the context theory to events which are related causally as well as temporally and to situations where the context of possibilities is recruited from background knowledge as well as when it is explicitly given.

Mental Representation of Possibilities

The first two series of experiments showed that the mental representation of events is of central importance in the selection and construction of counterfactual alternatives. In the final series of experiments, we aimed to study two factors which may influence the way that people mentally represent a situation. One factor is the verbal description of the situation, the second is co-reference. These two factors may be particularly important to our understanding of the processes involved in counterfactual thinking. First, as demonstrated in chapter 3, the verbal description of a situation may influence what people make explicit in their models and hence may influence the accessibility of different counterfactual alternatives. Second, the representation of co-reference is important to our understanding of counterfactual thinking because co-reference forms the bridge between factual and imaginary worlds. The accessibility of a counterfactual may depend on the ability to find a co-referential match for the facts. For example, as described in our computational model, people may search the context of possibilities for a co-referential match for the first fact.

We use another method for examining the representations that people construct, that is, deductive inferences. People may reason by constructing models of the different possibilities that are consistent with a set of premises (Johnson-Laird & Byrne, 1991). There is ample evidence to suggest that people find it easier to draw inferences from information which is explicitly represented in their models (e.g., Evans, Newstead &

Byrne, 1993) and therefore studies of reasoning may provide another way of examining these representations.

In Experiments 7 and 8, we showed that the *description* of a problem influences the inferences that people draw. In Experiment 7, we compared reasoning from two types of premise, biconditionals and exclusive disjunctions. Although these premises can describe the same information, they may be represented differently. We presented reasoners with premises such as the following:

If and only if Graham is not standing on the scales then Carol is standing on the scales.

and:

Graham is standing on the scales or Carol is standing on the scales but not both.

Both sentences have the same underlying meaning. The fully explicit set of models for both is as follows:

¬ Graham scales Carol scales

Graham scales ¬ Carol scales

But, people may draw different inferences from these sentences because they have a different initial representation. The biconditional may be represented by a single explicit model and an implicit model, as follows:

¬ Graham scales Carol scales

. . .

whereas the disjunction may be represented as follows:

Graham scales

Carol scales

Our results showed that reasoners were more accurate when the categorical premise matched an item that was explicitly represented in the premise, perhaps because this allows the two premises to be combined more easily (Schaeken et al., 1995). For

example, they were more accurate in reasoning from the disjunction when the categorical premise was 'Graham is standing on the scales' and were more accurate from the biconditional when the categorical premise was 'Graham is not standing on the scales'.

In Experiment 8, we also examined how the explicit representation of a premise affects reasoning. We gave participants premises of the following form:

Either P and Q or otherwise R and S

which they may represent as follows:

P Q

R S

The problems are consistent with six fully explicit models:

P	Q	¬ R	S
P	Q	R	¬ S
P	Q	¬ R	¬ S
¬ P	Q	R	S
P	¬ Q	R	S
¬ P	¬ Q	R	S

However, participants failed to represent all of these possibilities and as a result they succumbed to illusions on most problems. For example, when told that P was the case, they inferred that Q must also be the case. This result follows from the incomplete mental models but not from the fully explicit models (see row 5 of the fully explicit models for a counterexample).

The results of both experiments support the idea that the verbal description of a premise can have a dramatic influence on the inferences that people draw. The results have important implications for counterfactual thinking. The information that people make explicit in their models of the facts and the context of possibilities determines what alternatives they bring to mind. If an event is not explicitly represented in the facts of a

situation it cannot be used to recruit alternatives. Also, if an event is not explicitly represented in the context of possibilities it will tend to be inaccessible to the facts.

Experiments 7 and 8, also showed that *co-reference* influences the inferences people draw. In Experiment 7, responses were more accurate when both clauses in the premise referred to the same actor, but not when they referred to the same action. In Experiment 8, participants also made more of the predicted responses when all four clauses referred to the same actor. Co-reference also affected response times. Overall, responses tended to be faster when there were co-referential relations between the clauses. In Experiment 7, responses were faster when the shared co-referent was an action but not when it was the actor, perhaps because actions require a richer representation. In Experiment 8, responses were faster when all four clauses in the first premise referred to the same person than when they referred to two or four people. The effect was greatest when the shared referents were in different models.

The results support the conjecture that co-reference may allow reasoners to construct more concise representations. In general, people have difficulty in keeping multiple models in mind, but they appear to generate counterfactual possibilities with ease. Factual and counterfactual situations share many co-referents and it may be these links that allow people to keep multiple counterfactual possibilities in mind.

Implications

In this section we examine the implications of our results for our understanding of counterfactual thinking and for human cognition in general.

Implications for the generation of counterfactual thoughts

The results of our experiments provide support for the context theory of counterfactual thinking. They provide the first empirical evidence that the context of possibilities may play a role in selecting an event to mutate. Our experiments on the temporal order effect show that it is possible to shift the focus of people's counterfactual thoughts by changing the context of possibilities while leaving the facts unchanged. The focus theory assumes that people focus on the facts in selecting an alternative, and therefore it cannot explain our findings. The results lend support to the idea that people use the facts to recruit or chose among alternatives (Kahneman & Miller, 1986). We also suggested that reasons for acting may form the context in which actions occur and that the representation of reasons may provide the context of possibilities. The experiments showed that actions are more mutable than reasons and that different reasons shift the focus from the action. The finding that actions are more mutable is consistent with the context theory and the focus theory. According to the focus theory regardless of the reason, people focus on the salient item, that is, the action and mutate it. However, the differential effectiveness of different reasons is not compatible with the focus model and it was predicted by the context model.

The context theory is consistent with the results of our experiments on the temporal order effect and the effect of reasons on the mutability of actions. It is also consistent with other phenomenon in counterfactual thinking.

The idea that people think about the context of possibilities may explain the tendency for people to undo exceptional rather than routine events. Whether an event is viewed as exceptional will depend on the context in which it occurs and therefore what alternative is brought to mind. Imagine you normally drive to work but today you take your bike. If you call to mind other situations in which you go to work, then taking your bike will seem exceptional. But if you call to mind situations in which you ride your bike

then it will seem normal. In other words, the alternatives which are brought to mind will define whether an event is viewed as exceptional or normal (Kahneman & Miller, 1986).

Second, the suggestion that the context of possibilities is important may explain why actions are more mutable than failures to act (Byrne & McEleney, 2000; Kahneman & Miller, 1986). Actions may be more exceptional, although again it depends on the context. For example, when it is a person's job to act, then inactions become the exception (Zeelenberg et al., 2001).

The context theory proposes that people bring to mind an alternative sequence of events rather than an alternative single event (Byrne & McEleney, 2000). The context of possibilities is also evident in the causal order effect, that people are reluctant to undo an event if its cause is known to have occurred. If people mutate a single event in isolation, then they must do a post-hoc check to ensure that the mutated event is consistent with the other factual events. In contrast, if they bring an alternative sequence of events to mind then the events are already consistent with each other and the sequence may provide a readymade counterfactual.

Implications for human thinking

More generally our finding that the context of possibilities is important for counterfactual thinking illustrates some of it commonalities with other aspects of human cognition, namely problem-solving, decision-making and reasoning. The imagination of alternatives is often important in these domains. In many domains it is clear that the *context of possibilities* may be more important in recruiting these alternatives than *focusing* on a salient aspect of the facts.

Counterfactual thinking may play a role in problem-solving. Consider for example, the "two-string" or "pendulum" problem (Maier, 1931) which involves bringing

people into a room with two strings hanging from the ceiling. The task is to tie the two strings together but it is not possible to reach one while holding the other. Participants are given other objects including a pliers. The normal use of a pliers is to hold wires and other objects and people have difficulty in imagining alternative uses for it in this task. Focusing on the item and its salient functions may not be useful in constructing these imaginary alternatives. Identifying the context of alternative possibilities is critical in solving the problem and the creation of this set of possibilities may be what is brought about when the experimenter knocks against one of the strings, apparently by accident and it swings. Frequently, this allowed participants to imagine an alternative use for the pliers, that is, to construct a pendulum.

The imagination of alternatives is clearly important in making a decision. However, the tendency to imagine alternatives may depend on the context in which a choice is placed. In one study (Legrenzi et al., 1993), participants were given a choice of whether to do various actions, such as attending a movie or having dinner in a particular restaurant. They were allowed to ask as many questions as they liked about the choice until they made a decision. In one group, the choice was made without any background context and participants only asked questions about the action. However, for a second group the action was placed in a context, for example, participants were told that they were visiting Rome and the person they are asking is an expert on the local attractions. In this case, many participants asked about alternative choices. The context encouraged participants to consider alternatives to the item of focus. Furthermore, there was an association between thinking of alternatives and reaching a negative decision.

Context may also play an important role in imagining alternatives to factual events (Byrne, 1989). The conditional:

If Mary meets her friend then she will go to a play.

may be represented as follows:

friend play

. . .

where 'friend' represents 'Mary meets her friend' and 'play' represents 'she will go out to play'. When people are then told that Mary meets her friend, they usually infer that she will go out to play. However, a change in context can suppress this inference (Byrne, 1989; Byrne, Espino & Santamaria, 1999). When people are presented with a second conditional which suggests that additional preconditions must be met, for example:

If she has enough money then she will to a play

many people fail to draw the inference. The additional conditional leads reasoners to imagine a new alternative in which Mary meets her friend but does not go out to play. The context of possibilities in general is clearly important in many domains.

Implications for emotions and social judgements

One frequently studied consequence is the effect of counterfactual thinking on emotions. People's emotions may be amplified by (Kahneman & Miller, 1986) or even dependent on (Roese & Olson, 1995a) the availability of a counterfactual alternative and their emotions may vary depending on what is mutated (e.g., Byrne et al., 2000; Niedenthal et al., 1994). However, some recent studies have suggested that the effect may not be as pervasive as originally conceived (e.g., N'gbala & Branscombe, 1997). Many earlier studies examined the effect of counterfactual thinking on judgements and emotions indirectly. For example, they manipulated the mutability of an antecedent and measured the emotional reaction (e.g., Kahneman & Tversky, 1982a; Macrae, 1992). This may influence what people focus on in their counterfactual thoughts but may not change the extent to which they engage in counterfactual thinking (Roese, 1997). Even when

counterfactual thoughts are measured, the effect on emotions may not be causal. For example, in the standard card selection scenario used to study the temporal order effect, the focus of both counterfactual thoughts and emotions is generally placed on the second player. However, our experiments in chapter 3, showed that it was possible to shift the focus of people's counterfactual thoughts to the first player, but guilt and blame were still attributed to the second player. In chapter 4, we showed it was possible to shift the focus of counterfactual thoughts somewhat from actions to reasons but there is not a corresponding reduction in attributions of regret or blame. Our results suggest that attributions of emotions, at least in certain cases may depend more strongly on the representation of the facts than alternatives to those facts.

Implications for learning and planning for the future

Counterfactual thinking has also been implicated in learning and planning for the future. For example, in one study participants generated more counterfactuals after playing a game when they expected to play again (Markman et al., 1993) and counterfactual thinking has been found to enhance subsequent performance, for example, on an anagram task (Roese, 1994). One of the functions of counterfactual thinking may be the prevention of future unwanted outcomes (Mandel & Lehman, 1996; McEleney, 2000). The tendency for people to focus on actions in their counterfactual thoughts is consistent with this argument (Girotto et al., 1991; McEleney & Byrne, 2001). The results of our experiments provide further support for the idea that counterfactual thinking plays a role in learning and planning. People focused on actions more frequently when they were intentional than when they were accidental and they were mutated more when the reason for acting was internal, for example to achieve some goal, than when the reason was external, for example, the action was obligatory. Mutating actions which are outside control may have

little value in changing behaviour as people may be unable to avoid acting again in the future. On the other hand, mutating actions under their control may help them to make a different choice in the future.

Future Questions

We will consider two new questions that need to be addressed if the understanding of the processes underlying counterfactual thinking is to be further developed. First, what situations can contribute to the context of possibilities? Second, what factors do people use to recruit these possibilities and what factors might change their accessibility? We suggest that these questions may provide a framework for future research.

What situations may contribute to the context of possibilities?

Just as the context of possibilities was explicitly described in the card selection scenario, it may also be explicitly stated in many real life cases, especially when actions are involved. For example, people are often given rules, instructions or advice which may provide a potential contrast case when people deviate from the recommendations. People are presented with options everyday and in retrospect they may consider paths not taken. In addition, people are often given predictions or forecasts about what will happen, for example, what the weather or traffic will be like and these predictions provide a salient contrast when they go wrong. But many counterfactuals must also rely on past experiences (Kahneman & Miller, 1986). For example, the tendency to mutate exceptional events requires the imagination of what normally occurs. People may use consensus, distinctiveness or consistency information (Kelley, 1967), that is, information about other people in the same situation (social norms) or about the character in other situations or in the same situation on different occasions (personal norms). In this case,

people may draw on previous experiences or knowledge structures that encapsulate typical sequences of events (Hilton & Slugoski, 1986; Schank & Abelson, 1977). People may use background knowledge to make judgements about particular situations, for example, the service in this restaurant is slow, and therefore may also use them to generate an alternative to a particular case, for example, if only the service had been faster. Further research is clearly needed to clarify how background knowledge can contribute to the context of possibilities.

How do people recruit alternatives?

If the context of possibilities is vital to counterfactual generation then people must search for alternatives whenever they generate counterfactuals. The second question we wish to address therefore is how people recruit these alternatives. People may do this by searching for similar situations to the factual one. It may be that any event can be used to recruit these situations. But, universal items, such as gravity, may not be useful in finding similar experiences because they include too many possibilities. Instead, some of the features will need to be distinctive in some way. In our experiments on the temporal order effect, we have shown that in some cases at least, people use the first event in an independent sequence to search for a potential counterfactual. In our study of reasons for acting, we have also shown that reasons may also play a role. There are likely to be other features that people use and there are some hints in the literature to suggest what they might be. For example, people are more likely to focus on controllable events in their counterfactual thoughts which suggests that people draw on alternatives that include the central character.

Other factors besides the facts of the scenario may influence what alternatives are recruited. Different people may select a different alternative contrast case about the same

situation depending on their experience or interest. For example, given the case of a man who eats parsnips and suffers indigestion, the doctor may compare the man to other men whereas his wife may compare the parsnips to other vegetables (Hart & Honore, 1959). A change in perspective may bring different alternatives to mind and has been shown to change the focus of people's counterfactual thoughts (Mandel & Lehman, 1986). Also, the type of question asked may lead people to draw on different alternatives. Questions about the cause of an outcome tend to evoke different alternatives from counterfactual questions about how the outcome could have happened differently (Mandel & Lehman, 1986; McEleney & Byrne, 2001). Future research is clearly needed to clarify how people access and construct the context of possibilities.

Conclusion

In this thesis we have described a study of the mental representations and cognitive processes underlying counterfactual thinking. The imagination of counterfactual alternatives tends to follow certain regularities. The results of our experiments, support the view that the context plays a central role in the imagination of these alternatives. In our experiments, we demonstrated that by changing the mental representation of the context it is possible to change what people focus on. This finding has important implications, not least, because it means that by changing the context it may be possible to change the usual pattern that counterfactual thinking follows.

References

Anscombe, G. E. M. (1963). *Intention.* 2nd Ed. Oxford: Basil Blackwell.

Au, T.K. (1983). Chinese and English Counterfactuals: The Sapir-Whorf Hypothesis revisited. *Cognition, 15*, 155-187.

Baddeley, A. D. (1999). *Essentials of Human Memory.* Sussex: Psychology Press.

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 629-654.

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

Bouquet, P. & Warglien, M. (1999). Mental Models and Local Models Semantics: the Problem of Information Integration. *Proceedings of the European Conference on Cognitive Science* (ECCS'99), pp.169--178, Siena, Italy.

Braine, M.D.S., & O'Brien, D. P. (1991). A theory of 'if': A lexical entry, reasoning program, and pragmatic principles. *Psychological Review, 98*, 182-203.

Braine, M.D.S., & O'Brien, D. P. (1998). *Mental Logic.* Mahwah, NJ: Lawrence Erlbaum Associates.

Byrne, R. M. J. (1989). Suppressing valid inferences with conditionals. *Cognition, 31*, 61-83.

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. 105-54. San Diego, CA: Academic Press.

Byrne, R. M. J., Culhane, R. & Tasso, A. (1994). The temporality effect in thinking about what might have been. In J. Moore & J. Lehman (Eds.), *Proceedings of the seventeenth annual conference of the Cognitive Science Society* (pp. 385-390). Hillsdale: Erlbaum.

Byrne, R. M. J., Espino, O. & Santamaria, C. (1999). Counterexamples and the Suppression of Inferences. *Journal of Memory and Language, 40,* 347-373.

Byrne, R. M. J. & Handley, S. J. (1992). Reasoning strategies. *Irish Journal of Psychology: Trinity 400 Special Issue. 13,* 111-124.

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

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 and Cognition, 28,* 264-281.

Byrne, R. M. J. & Tasso, A. (1994). Counterfactual reasoning: inferences from hypothetical conditionals. In A. Ram & K. Eiselt (Eds.), *Proceedings of the sixteenth annual conference of the Cognitive Science Society* (pp. 124-129). Hillsdale: Erlbaum.

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

Costello, T., & McCarthy, J. (1999). Useful Counterfactuals. *Electronic Transactions on Artificial Intelligence, 3,* 51-76.

Davidson, D. (1963). Actions, Reasons, and Causes. *Journal of Philosophy*, 60, 685-700.

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

Davis, C. J., Lehman, D. R., Silver, R. C., Wortman, C. B., & Ellard, J. H. (1996). Self-blame following a traumatic event: The role of perceived avoidability. *Personality and Social Psychology Bulletin, 22*, 557-567.

Donnellan, K. S. (1967). Reasons and Causes. In P. Edwards (Ed.), *The Encyclopedia of Philosophy*. New York: The Macmillan Company.

Evans, J.St.B.T. (1993). The model theory of conditional reasoning: Critical appraisal and revision. *Cognition, 48,* 1-20.

Evans, J.St.B.T. (1998). Matching Bias in Conditional Reasoning: Do we understand it after 25 years? *Thinking and Reasoning, 4,* 45-82.

Evans, J.St.B.T., Newstead, S.E. & Byrne, R.M.J. (1993). *Human reasoning: The Psychology of Deduction.* Hove, UK: Lawrence Erlbaum Associates.

Fillenbaum, S. (1974). Information amplified: memory for counterfactual conditionals. *Journal of Experimental Psychology, 25,* 44-49.

Gavanski, I. & Wells, G. L., (1989). Counterfactual Processing of Normal and Exceptional Events. *Journal of Experimental Social Psychology, 25*, 314-325.

Gernsbacher, M. A. (1990). *Language Comprehension as Structure Building*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Gilovich, T., & Medvec, V. H. (1994). The temporal pattern to the experience of regret. *Journal of Personality and Social Psychology*, 67, 357-365.

Gilovich, T., & Medvec, V. H. (1995). The experience of regret: What, When 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.

Gleicher, F., Kost, K. A., Baker, S. M., Strathman, A. J., Richman, S. A., & Sherman, S. J. (1990). The role of counterfactual thinking in judgements of affect. *Personality and Social Psychology Bulletin, 16*, 284-295.

Goldman, A. (1970). *A Theory of Human Action.* Englewood Cliffs, N.J.:Prentice-Hall.

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

Goodman, N. (1973). *Fact, Fiction and Forecast.* 3rd Edition. New York: Bobbs-Merrill.

Green, D.W., McClelland, A., Mucki, L., & Simmons, C. (1999). Arguments and Deontic Decisions. *Acta Psychologica, 101*, 27-47.

Handley, S. J. (1996). Explicit Models, Disjunction Alternatives and Deductive Reasoning. *PhD Thesis, University of Wales, Cardiff.*

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

Hart, H. L. A. & Honore, A.M. (1959). *Causation in the Law*. Oxford: Clarendon Press.

Hilton, D. J. & Slugoski, B. R. (1986). Knowledge-based Causal Attribution: The Abnormal Conditions Focus model. *Psychological Review, 93,* 75-88.

Hofstadter, D.R. (1979). *Godel, Escher, Bach: An Eternal Golden Braid.* New York: Basic Books.

Hume, D. (1938). An abstract of a treatise of human nature. In J. M. Keynes & P. Sraffa (Eds.). *An abstract of a treatise of human nature.* Cambridge: Cambridge University Press.

Johnson, J. T. (1986). The knowledge of what might have been: Affective and attributional consequences of near outcomes. *Personality and Social Psychology Bulletin, 12*, 51-62.

Johnson-Laird, P. N. (1983). *Mental Models*. Cambridge: Cambridge University Press.

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

Johnson-Laird, P. N. & Byrne, R. M. J. (in press). Conditionals: a theory of meaning, pragmatics and inferences. *Psychological Review.*

Johnson-Laird, P. N. & Byrne, R. M. J. & Schaeken, W. S., (1992). Propositional reasoning by model. *Psychological Review, 99,* 418-439.

Johnson-Laird, P. N. & Savary, F. (1995). How to make the impossible seem probable. In Moore, J.D., and Lehman, J.F. (Eds.) *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates, pp. 381-384.

Johnson-Laird, P.N, & Savary, F. (1999). Illusory inferences: A novel class of erroneous deductions. *Cognition, 71,* 191-229.

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

Kahneman, D., & Tversky, A. (1982a). The psychology of preferences. *Scientific American,* 246, 160-173.

Kahneman, D., & Tversky, A. (1982b). 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.

Kahneman, D., & Varey, C. A. (1990). Propensities and Counterfactuals: The loser that almost won. *Journal of Personality and Social Psychology, 59,* 1101-1110.

Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska Symposium on Motivation* (Vol.15). Lincoln: University of Nebraska Press.

Klauer, K. C., Jacobsen, T., & Migulla, G. (1995). Counterfactual processing: test of an hierarchical correspondence model. *European Journal of Social Psychology,* 25, 577-595.

Knight, R. T & Grabowecky, M. (1995). Escape from linear time. Prefrontal cortex and conscious experience. In M.S. Gazzaniga (Ed.), *The Cognitive Neurosciences,* (pp. 1357-1371). Cambridge, MA: MIT Press.

Landman, J. (1987). Regret and elation following action and inaction: Affective responses to positive versus negative outcomes. *Personality and Social Psychology Bulletin,* 13, 524-536.

Legrenzi, P., Girotto, V. & Johnson-Laird, P.N. (1993). Focusing in reasoning and decision-making. *Cognition,* 49, 37-66.

Legrenzi, M.S., Legrenzi, P. Girotto, V., & Johnson-Laird, P.N. (2000). Reasoning to consistency: a theory of naïve nonmonotonic reasoning. *In submission.*

Lewis, D. (1973). *Counterfactuals*. Oxford: Blackwell.

Macrae, C. N. (1992). A tale of two curries: counterfactual thinking and accident-related judgements. *Personality and Social Psychology Bulletin,* 18, 84-87.

Macrae, C. N., Milne, A. B., & Griffiths, R. J. (1993). Counterfactual thinking and the perception of criminal behaviour. *British Journal of Psychology,* 84, 221-226.

Maier, N. R. F. (1931). Reasoning in humans II: The solution of a problem and its appearance in consciousness. *Journal of Comparative Psychology, 12,* 181-194.

Mandel, D. R., & Lehman, D. R. (1996). Counterfactual thinking and ascriptions of cause and preventability. *Journal of Personality and 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.

Markman, K. D., Gavanski, I., Sherman, S. J. & McMullen, M. N. (1995). The impact of perceived control on the imagination of better and worse possible worlds. *Personality and Social Psychology Bulletin, 21*, 588-595.

McCloy, R. & Byrne, R.M.J. (1999). Thinking about what might have been: "if only", "even if", causality and emotions. *Proceedings of the Twenty-first Annual Conference of the Cognitive Science Society* (pp. 373-377). Mahwah, NJ: Lawrence Erlbaum Associates.

McCloy, R. and Byrne, R. M. J. (2000). Counterfactual thinking about controllable events. *Memory and Cognition, 28,* 1071-1078.

McCloy, R. & Byrne, R.M.J. (In press). Semifactual 'even if' thinking. *Thinking and Reasoning.*

McEleney, A. (2000). Why it happened and how it could have been different: A comparison of causal and counterfactual thinking. *PhD thesis., University of Dublin, Trinity College.*

McEleney, A. and Byrne, R. M. J. (2001). Counterfactual and Causal Thoughts. *In submission.*

McMullen, M. N., Markman, K. D. & Gavanski, I. (1995). Living in Neither the Best Nor Worst of All Possible Worlds: Antecedents and Consequences of Upward and Downward Counterfactual Thinking. In N. J. Roese & J. M. Olson (Eds.), *What might have been: The social psychology of counterfactual thinking.* Hillsdale, NJ: Lawrence Erlbaum.

Medvec, V. H., Madey, S. F. & Gilovich, T. (1995). When less is more: Counterfactual thinking and satisfaction among Olympic athletes. *Journal of Personality and Social Psychology, 69*, 603-610.

Melden, A. I. (1961). *Free Action*. London: Routledge and Kegan Paul.

Mele, A. R. (1992). *Springs of Action: Understanding Intentional Behavior.* New York: Oxford University Press.

Miller, D. T. & Gunasegaram, S. (1990). Temporal order and the perceived mutability of events: Implications for blame assignment. *Journal of Personality and Social Psychology, 59*, 1111-1118.

Miller, D. T., & McFarland, C. (1986). Counterfactual thinking and victim compensation: A test of norm theory. *Personality and Social Psychology Bulletin,* 12, 513-519.

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

Moreno-Rios, S., Garcia-Madruga, J. A. & Byrne, R. M. J. (2001). The effects of Linguistic Mood on 'if': Semifactual and Counterfactual Conditionals. *In submission.*

Morris, M. W., & Sim, D. L. H. (1999). Choosing remedies after accidents: Counterfactual thoughts and the focus on fixing "human error." *Psychonomic Bulletin and Review, 6*, 579-585.

Morrow, D. G., Bower, G.H. & Greenspan, S. L. (1989). Updating situation models during narrative comprehension. *Journal of Memory and Language, 28*, 292-312.

Nasco, S. A., & Marsh, K. L. (1999). Gaining Control through Counterfactual Thinking. *Personality and Social Psychology Bulletin, 25,* 556-568.

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 from guilt in counterfactual thinking. *Journal of Personality and Social Psychology*, 67, 585-95.

Pollack, J. L., (1986). *Subjunctive Reasoning.* Dordrecht Reidel.

Read, D. (1985). *Determinants of relative mutability.* Cited in Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review*, *93*, 136-153.

Rips, L. J. (1994). *The Psychology of Proof.* Cambridge, M.A.: MIT Press.

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

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

Roese, N. J., & Olson, J. M. (1993). The structure of counterfactual thought. *Personality and Social Psychology Bulletin, 19*, 312-319.

Roese, N. J., & Olson, J. M. (1995a). Counterfactual thinking: A critical overview. In N. J. Roese & J. M. Olson (Eds.), *What Might Have Been: The Social Psychology of Counterfactual Thinking.* Hillsdale, NJ: Lawrence Erlbaum. Pp 1-55.

Roese, N. J., & Olson, J. M. (1995b). *What Might Have Been: The Social Psychology of Counterfactual Thinking.* Hillsdale, NJ: Lawrence Erlbaum.

Savitsky, K., Medvec, V. H., & Gilovich, T. (1997). Remembering and regretting: The Zeigarnik effect and the cognitive availability of regrettable actions and inactions. *Personality and Social Psychology Bulletin, 23*, 248-257.

Schaeken, W., Johnson-Laird, P. N., and Byrne, R. M. J., d'Ydewalle, G. (1995). A comparison of conditional and disjunctive inferences: A case study of the mental model theory of reasoning. *Psychologica Belgica, 35*, 57-70.

Schank R. C., & Abelson, R. P. (1977). *Scripts, plans, goals and understanding.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Seelau, E. P., Seelau, S. M., Wells, G., & Windschitl, P. D. (1995). Counterfactual constraints. In N. J. Roese & J. M. Olson (Eds.), *What Might Have Been: The Social Psychology of Counterfactual Thinking.* Hillsdale, NJ: Lawrence Erlbaum. Pp 57-80.

Shaver, K. G. (1985). *The Attribution of Blame: Causality, Responsibility and Blameworthiness.* New York: Springer-Verlag.

Sherman, S. J., & McConnell, A. R. (1996). Counterfactual Thinking in Reasoning. *Applied Cognitive Psychology, 10*, 113-124.

Smith, P., & Jones, O. R. (1986). *The Philosophy of Mind: An Introduction.* Cambridge: Cambridge University Press.

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*. Oxford: Basil Blackwell.

Teigen, K. H. (1995). How good is good luck? The role of counterfactual thinking in the perception of lucky and unlucky events. *European Journal of Social Psychology, 25*, 281-302.

Teigen, K. H. (1997). Luck, envy and gratitude: it could have been different. *Scandinavian Journal of Psychology, 38,* 313-323.

Turley K. J., Sanna, L. J., & Reiter, R. L. (1995). Counterfactual thinking and perceptions of rape. *Basic and Applied Social Psychology, 17*, 285-303.

Tversky, A. & Kahneman, D. (1982). Judgement under uncertainty: Heuristics and biases. *Science, 125,* 1125-1131.

Von Wright, G. H. (1983). *Practical Reason.* Oxford: Basil Blackwell.

Walsh, C.R. & Byrne, R.M.J. (2001a). 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: Lawrence Erlbaum Associates.

Walsh, C.R. & Byrne, R.M.J. (2001b). Counterfactual Thinking: The Temporal Order Effect. In D. O'Donoghue (Ed.) *Proceedings of the Twelfth Irish Conference on Artificial Intelligence and Cognitive Science Society* (pp. 233 - 242). NUI, Maynooth.

Walsh, C.R. & Byrne, R.M.J. (2001c). Counterfactual Thinking and The Temporal Order Effect. *In submission.*

Walsh, C. R. & Byrne, R.M.J. (2001d). Counterfactual Thinking about Reasons for Acting. *In submission.*

Walsh, C.R. & Johnson-Laird, P.N. (2001). The Semantic Modulation of Deductive Premises. In J.D. Moore & K. Stenning (Eds.) Proceedings of the 23rd Annual Conference of the Cognitive Science Society (pp. 1084 - 1088). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Wells, G. L., Taylor, B. R. & Turtle, J. W. (1987). The undoing of scenarios. *Journal of Personality and Social Psychology*, 53, 421-430.

Zeelenberg, M., van Dijk, W. W., van der Pligt, J., Manstead, A. S. R., van Empelen, P., & Reinderman, D. (1998). Emotional reactions to the outcomes of decision: The role of counterfactual thought in the experience regret and disappointment. *Organizational Behavior and Human Decision Processes, 75*, 117-141.

Zeelenberg, M., van den Bos, K., van Dijk, E., Pieters, R. (2001). The inaction effect in the psychology of regret. *Presented at the EAESP Small Group Meeting on Counterfactual Thinking.* Aix-en-Provence, France.

Appendices

		Page
Appendix 1:	TempFocus program code	196
Appendix 2:	TempContext program code	215
Appendix 3:	Materials used in the temporal order effect experiments	234
Appendix 4:	Materials used in the reasons for acting experiments	237
Appendix 5:	Materials used in the mental representations experiments	250

Appendix 1: TempFocus program code

;;; A simulation of the focus theory of the temporal order effect in
;;; counterfactual thinking.
;;; The aim of this program is to construct a counterfactual conditional from a factual
;;; scenario and a conditional. It simulates a theory of the processes underlying the
;;; temporal order effect.

;;; ***Representation of the Facts***
;;; The program accepts a factual scenario as a series of antecedent events and an
;;; outcome. Each event is assigned a name and a value. For example, an event might be
;;; named 'Johns-card' and might be assigned a value 'Red'. To denote that the value is
;;; negative, the model can be tagged with the word 'not' e.g., (John not Red). If the
;;; event is not tagged, the program assumes that the value is positive. The program also
;;; assumes that the events are presented in their temporal order of occurrence. For
;;; example, given the following facts:

;;; 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.

;;; The program accepts the following list of facts:

;;; ((Johns-card Red) (Michaels-card Black) (Outcome Lose))

;;; ***Representation of the Context of Possibilities***
;;; The program also accepts a set of conditions that determine the outcome. The
;;; conditions contain a list of events without their values, a set of antecedent
;;; conditions and the consequent. Take for example, the following conditional:

;;; If the two cards they pick are of the same colour (i.e., both red or both black),
;;; each individual wins £1,000. Otherwise, neither individual wins anything.

;;; The program accepts this conditional in the following format:

```
;;;  (((Johns-card) (Michaels-card))    ..... the events
;;;   ((Both Red) or (Both Black))      ..... the antecedent conditions
;;;   ((Outcome Win)))                  ..... the outcome.
```

;;; From this conditional, the program constructs the following mental models:

```
;;;       Johns-card  Red      Michaels-card  Red      Outcome Win
;;;       Johns-card  Black    Michaels-card  Black    Outcome Win
;;;                                                    Outcome Lose
```

;;; The program aims to minimise the amount of information that it represents explicitly.
;;; For this reason, it does not represent the losing conditions explicitly. In order to
;;; represent the fact that the winning conditions are exhaustively represented, the
;;; program constructs a third model which only contains a consequent, in this case that
;;; the outcome is a loss. This indicates that any antecedents other than those represented
;;; in the first two models will lead to a loss.

;;; TOP-LEVEL FUNCTION - GenerateCounterfactual

;;; The top-level function is GenerateCounterfactual. This function goes
;;; through the steps involved in generating a counterfactual conditional.

;;; The function begins by calling RepresentFacts and RepresentContext
;;; to construct models of the facts (Fact-Model) and winning conditions
;;; (Possibility-Models) respectively. RepresentFacts calls PrintModel to
;;; print a model of the facts to the screen, e.g.,

;;; Factual: Johns-card Red Michaels-card Black Outcome Lose

;;; The function then calls MutateFacts to mutate one of the antecedent events
;;; in the Fact-Model. As the objective is to simulate the temporal order effect,
;;; it mutates the most recent antecedent event. When this is done, the function
;;; creates a counterfactual model 'Counterfact-Model' by taking the Fact-Model
;;; and substituting the mutated event for its factual alternative.

;;; The function then calls CheckContext to check through each of the
;;; conditions to see whether the counterfactual scenario leads to a new outcome.
;;; If it does, then CheckContext returns the scenario with the new outcome.

;;; Finally GenerateCounterfactual prints the counterfactual model and it
;;; prints a conditional in the following form: If it had been the case
;;; that "the mutated event" then it would have been the case that
;;; "the new outcome".

```
(defun GenerateCounterfactual (facts conditional)

 (let* ((Possibility-Models (RepresentContext (first conditional)
                    (second conditional) (third conditional)))
        (Fact-Model (RepresentFacts facts))

        (MutatedEvent (MutateFacts Fact-Model))
        (Counterfact-Model (replace1 MutatedEvent Fact-Model))

        (Counterfact-Model  (CheckContext Possibility-Models Counterfact-Model))
        (CounterfactConsequent (first (last Counterfact-Model)))
        (FactualConsequent (first (last Fact-Model)))
        (CounterfactStatus
              (if (same-value FactualConsequent CounterfactConsequent) 'sf 'cf)))

  (dolist (Counterfactual Possibility-Models) (PrintModel Counterfactual 'p))
  (PrintModel Counterfact-Model CounterfactStatus)
  (format t "~&~%~%If it had been the case that ~S
          ~% then it would have been the case that ~S"
        (remove nil MutatedEvent)
        (remove nil CounterfactConsequent))
))
```

```
;;; CONSTRUCT MODELS TO REPRESENT THE CONTEXT

;;; RepresentContext is the top-level of a set of functions which
;;; constructs a set of models to represent the Context of Possibilities. It is called
;;; in GenerateCounterfactual. It accepts a set of events, a set of antecedent
;;; conditions and an outcome.It calls ModelContext to construct a
;;; a set of models to represent the antecedent conditions.
;;; RepresentContext then adds the consequent to each model and adds a
;;; sign to events with positive values. Finally the function records the
;;; fact that the conditional is a bi-conditional. To represent this, it
;;; adds another model which only contains a consequent. This model
;;; indicates that any conditions other than those explicitly represented
;;; will lead to a loss. For example, given the event-list
;;;       (johns-card michaels-card)
;;; the antecedent conditions
;;;       ((both red) or (both black))
;;; and the outcome
;;;       (outcome win)
;;; the function returns the following set of models:
;;;
;;;  (((johns-card nil red) (michaels-card nil red) (outcome nil win))
;;;   ((johns-card nil black) (michaels-card nil black) (outcome nil win))
;;;   ((outcome nil lose)))


;;; Note, in the following functions:
;;; An event is represented by a single list, e.g.,
;;;       (Johns-card red),
;;; a model contains a list of events, e.g.,
;;;       ((Johns-card red) (Michaels-card red))
;;; and a set of models contains a list of models, e.g.,
;;;       (((Johns-card red) (Michaels-card red))
;;;        ((Johns-card Michaels-card red))).

(defun RepresentContext (events antecedent-conditions outcome)
  (append
   (add-signs (add-outcome (ModelContext events antecedent-conditions) outcome))
   (list (list (negate-antecedent (enter-sign outcome))))
))
```

```
;;; ModelContext is called in RepresentContext. From the
;;; conditional it accepts the list of events without their values,
;;; e.g.,
;;;      (johns-card michaels-card)
;;; and the antecedent conditions. These conditions may directly specify
;;; the value to be assigned to a particular event, e.g., 'John picks a
;;; red card' and this will be represented as follows:
;;;      ((johns-card red)).
;;; Alternatively the conditions may specify that both, one or only-one
;;; event/s have a particular value, e.g., 'both players pick a red card' is
;;; represented as follows:
;;;      ((both red)).
;;; ModelContext also accepts four connectives: and, or, ori
;;; (or inclusive) and ore (or exclusive). The antecedent conditions may
;;; take the following format:
;;;      (((one red) and (one black)) or (both black))
;;;  or
;;;      ((john red) or (michael red)).
;;;
;;; In order to interpret the conditional, ModelContext takes one
;;; item from the antecedent conditions at a time. If the item contains an
;;; embedded list, e.g.,
;;;    ((one red) and (one black))
;;; then the function calls itself again and puts the antecedent conditions
;;; equal to this item. If the item is a connective, then it is put on the
;;; stack. Otherwise the function calls CreateModels to generate models to
;;; represent the item. It assumes that people minimise the amount of
;;; information that they make explicit in their models because of working
;;; memory limitations (Johnson-Laird & Byrne, 1991). It puts these models
;;; onto a list called Possibility-Models.
;;;
;;; Next, the function looks to see if there is a connective on the stack.
;;; If there is and the connective has not just been put there, it takes
;;; the two sets of models in Possibility-Models and it joins them with the
;;; connective. It replaces the two sets of models with the new combined
;;; set of models. It then reads the next item on the antecedent condition
;;; list and repeats this process until the list is empty.

;;; For example, given the event-list
;;;      (johns-card michaels-card)
;;; and the antecedent conditions
;;;      ((both red) or (both black))
;;; the program constructs the following set of models:
;;;      (((johns-card red) (michaels-card red))
;;;       ((johns-card black) (michaels-card black))).
```

```
(defun ModelContext (events antecedent-conditions)
  (do* ((conditions antecedent-conditions (rest conditions))
        (item (first conditions) (first conditions))
        (stack nil)
        (Possibility-Models nil))
       ((null conditions) (remove-duplicates1 (first Possibility-Models)))

    (cond ((member item '(or and ori ore)) (push item stack))
          ((eq (type-of (first item)) 'cons)
           (push (ModelContext events item) Possibility-Models))
          (t (push (CreateModels events item) Possibility-Models)))

    (if (and stack (not (member item '(or ori ore and))))
      (push
       (funcall
        (cond ((eq (first stack) 'and) 'and-x)
              ((eq (first stack) 'ore) 'ore-x)
              ((eq (first stack) 'or) 'or-x)
              ((eq (first stack) 'ori) 'ori-x))
        (pop Possibility-Models)(pop Possibility-Models))
       Possibility-Models))
))
```

```
;;; CreateModels is called in RepresentContext. It takes the list of
;;; events before they have been assigned values, e.g.,
;;;      (johns-card michaels-card)
;;; and one item from the antecedent conditions, e.g.,
;;;      (both red).
;;; If the item contains one, only-one or both, then it calls the
;;; corresponding function with that name. Otherwise, if the item refers to
;;; a specific event, e.g.,
;;;      (johns-card red)
;;; then the function returns the item embedded in two more lists, i.e.,
;;;      (((johns-card red))).

(defun CreateModels (events item)
  (cond ((member (first item) '(one only-one both))
         (funcall (first item) events (rest item)))
        (t (list (list item)))
))
```

```
;;; Only-one is called in CreateModels. It accepts a list of two events,
;;; e.g.,
;;;      (johns-card  michaels-card)
;;; and the value to be assigned to those events, e.g.,
;;;      (Red).
;;; It constructs two models; each one contains an explicit item in which
;;; the value is assigned to one of the events. Each model also contains a
;;; footnote to note that the other model is false. For example, given
;;; the input described above, the function returns
;;;      (((johns-card red) (not next model))
;;;       ((michaels-card red) (not prev model))).

(defun only-one (event-pair value)
        (list
        (list (cons (first event-pair) value) '(not next model))
        (list (cons (second event-pair) value) '(not prev model)))
)


;;; One is called in CreateModels. It accepts a list of events, e.g.,
;;;      (johns-card  michaels-card)
;;; and the value to be assigned to those events, e.g.,
;;;      (Red).
;;; It constructs models by assigning the value to each of the events.
;;; events at random, e.g.,
;;;      (johns-card red).
;;; It returns a list of models, e.g.,
;;;      (((johns-card red)) ((michaels-card red)))

(defun one (event-list value)
        (do ((events (reverse event-list) (rest events))
                (new-list nil))
        ((null events) new-list)
        (push (list (cons (first events) value)) new-list)
))


;;; Both is called in CreateModels. It accepts a list of two events,
;;; e.g.,
;;;      (johns-card michaels-card)
;;; and the value to be assigned to those events, e.g.,
;;;      (not Red).
;;; It constructs one model which assigns the value to the two events.
;;; For example, given the input described above, the function returns
;;;      (((johns-card red) (michaels-card red)))

(defun both (event-pair value)
  (list (list (cons (first event-pair) value) (cons (second event-pair) value)))
)
```

```
;;; Or-x is called in RepresentContext. It accepts two sets of models,
;;; e.g.,
;;;     (((Johns-card red)))
;;; and
;;;     (((Michaels-card black)))
;;; which are to be linked by the connective 'or'. It returns a new set of
;;; models containing the first set of models and the second set of models.
;;; For example, given the above input, the function returns
;;;     (((Michaels-card black))
;;;      ((Johns-card red)))

(defun or-x (models-1 models-2)
  (append models-2 models-1))



;;; Ori-x is called in RepresentContext. It accepts two sets of models,
;;; e.g.,
;;;     (((Johns-card red)))
;;; and
;;;     (((Michaels-card black)))
;;; which are to be linked by the connective 'ori' (or inclusive). It
;;; returns a new set of models containing the first set of models, the
;;; second set of models and a conjunction of the first and second set. For
;;; example, given the above input, the function returns
;;;     (((Michaels-card black))
;;;      ((Johns-card red))
;;;      ((Michaels-card black) (Johns-card red)))

(defun ori-x (models-1 models-2)
        (let ((new-list (append models-2 models-1)))
        (append new-list (and-x models-1 models-2))
))



;;; Ore-x is called in RepresentContext. It accepts two sets of models,
;;; e.g.,
;;;     (((Johns-card red)) ((Michaels-card red)))
;;; and
;;;     (((Michaels-card black)))
;;; which are to be linked by the connective 'ore'. It combines the two sets
;;; of models into one. For example, given the above input, the function
;;; returns
;;;     (((Michaels-card black)) ((Johns-card red)) ((Michaels-card red)))
;;;
;;; The function also adds footnotes to the models but only where the models
;;; are very simple. If each set of models contains just one model and each
;;; model contains just one token, then the function adds footnotes to the
```

```
;;; models and returns, for example,
;;;    (((Johns-card red) (not next model))
;;;     ((Michaels-card red) (not prev model))).
;;; It is assumed that people have difficulty using footnotes with more
;;; complex models or sets of models and so the program does not represent
;;; these.

(defun ore-x (models-1 models-2)
  (cond ((and (eq (length models-1) 1) (eq (length (first models-1)) 1)
              (eq (length models-2) 1) (eq (length (first models-2)) 1))
         (list (list (first (first models-2)) '(not next model))
                     (list (first (first models-1)) '(not prev model))))
        (t (append models-2 models-1)))
)



;;; And-x is called in RepresentContext. It accepts two sets of models
;;; which are to be linked by the connective 'and', e.g.,
;;;     (((Johns-card red)))
;;; and
;;;     (((Michaels-card black))).
;;; It takes one model at a time from the first set of models and it calls
;;; combine to combine this model with each of the models in the second set.
;;; It returns a new combined set of models, e.g.,
;;;     (((Michaels-card black) (Johns-card red))).

(defun and-x (models-1 models-2)
        (do* ((first-list models-1 (rest first-list))
                  (model-1 (first first-list) (first first-list))
                  (new-list nil (append (first new-list) (rest new-list))) )
        ((null first-list) (if (null models-1) models-2 new-list))
        (push (combine models-2 model-1) new-list)
))



;;; Combine is called in and-x. It accepts one set of models e.g.,
;;;     (((Johns-card red)) ((Michaels-card red)))
;;; This set of models represents two possibilities, John selecting a red
;;; card and Michael selecting a red card. It also accepts a single model,
;;; e.g.,
;;;     ((Johns-card red)(Michaels-card black))
;;; This model represents a single possibility in which John selects a red
;;; card and Michael selects a black card. The function takes each of the
;;; models in the set in turn and combines it with the single model. It
;;; removes footnotes from the first model in the set and from the single
;;; model. This is because it assumes that people have difficultly
;;; combining models that contain footnotes. It then calls combine-models
```

```
;;; to combine the two models. It puts the combined models in a list and
;;; returns a new set of models. For example, given the example above, the
;;; function returns a set which contains a single model, i.e.,
;;;      (((Johns-card red) (Michaels-card black)))
;;; which represents the possibility of Michael selecting a black card and
;;; John selecting a red card.

(defun combine (list-of-models 1-model)
  (do* ((condition-list list-of-models (rest condition-list))
        (condition (first condition-list) (first condition-list))
        (condition (remove-footnotes condition) (remove-footnotes condition))
        (1-model (remove-footnotes 1-model) 1-model)
        (new-condition (combine-models condition 1-model)
                       (combine-models condition 1-model))
        (new-condition-list (list new-condition) (push new-condition new-condition-list)))
       ((eq (length condition-list) 1) (remove nil new-condition-list))
))



;;; Remove-footnotes is called in Combine. It accepts a model and removes
;;; any footnotes from it.

(defun remove-footnotes (model)
  (let* ((model1 (remove '(not next model) model :test #'equal))
         (model1 (remove '(not prev model) model1 :test #'equal)))
    model1
))



;;; Combine-models is called in combine to combine two models. It accepts
;;; two models e.g.,
;;;      ((Johns-card red))
;;; which represents one event in which John selects a red card, and
;;;      ((Michaels-card black))
;;; which also represents a single event, Michael selecting a black card.
;;; It takes each event in the second model in turn. If the event is already
;;; contained in the first model with the same value, then it does nothing.
;;; If the event is contained in the first model but they both have
;;; different values, then the function returns nil because the two models
;;; cannot be combined. If the event is not contained in the first model,
;;; then it adds the event to the model. It returns the combined model.
;;; Given the example above, the function returns the following model,
;;;      (((Johns-card red) (Michaels-card black)))
;;; which represents the possibility that Michael selects a black card and
;;; John selects a red card.
```

```
(defun combine-models (model-1 model-2)
  (do* ((model2 model-2 (rest model2))
        (event (first model2) (first model2))
        (new-model (reverse model-1)))
       ((null model2) (reverse new-model))
    (cond ((compare event model-1) nil)
          ((find-event event model-1) (return nil))
          (t (push event new-model)))
))
```

;;; Remove-duplicates1 is called in RepresentContext to remove
;;; duplicates from a set of models. It takes one model at a time and it
;;; calls find-duplicate to see if there is a duplicate model in the
;;; remaining set of models. If no duplicate is found, the model is placed
;;; on a new list. After checking all the models, the program returns the
;;; new list.

```
(defun remove-duplicates1 (list-of-models)
  (do* ((models list-of-models (rest models))
        (model (first models) (first models))
        (new-models nil))
       ((null models) (reverse new-models))
    (cond ((not (find-duplicate model (rest models)))
           (push model new-models)))
))
```

;;; Find-duplicate is called in remove-duplicates. It accepts a single model
;;; and a list of models. It takes one model from the list at a time and it
;;; calls same-models to see if this model is equal to the single model.
;;; If a match is found, it returns true. Otherwise, it returns nil.

```
(defun find-duplicate (model list-of-models)
  (do* ((models list-of-models (rest models))
        (1st-model (first models) (first models)))
       ((null models) nil)
    (if (same-model model 1st-model)
      (return t)
      nil)
))
```

;;; Same-model is called in find-duplicates. It accepts two models. It
;;; takes the second model and it goes through one event at a time checking
;;; to see if the event is contained in the first model. If an event is
;;; found which is not contained in the first model, it returns nil.
;;; Otherwise, if all of the events in the second model are in the first and
;;; the two models contain the same number of events, then the function
;;; returns true.

```
(defun same-model (model-1 model-2)
  (do* ((events-2 model-2 (rest events-2))
        (event (first events-2) (first events-2)))
       ((null events-2) (if (eq (length model-1) (length model-2)) t))
    (if (compare event model-1) nil (return nil))
))
```

;;; Add-outcome is called in RepresentContext It accepts the set of
;;; models which have been constructed to represent the antecedent
;;; conditions and the consequent. It adds the consequent to each of the
;;; models and returns the new set of models.

```
(defun add-outcome (model-list outcome)
  (do* ((models (reverse model-list) (rest models))
        (model (first models) (first models))
        (new-list nil))
       ((null models) new-list)
    (push (reverse (cons outcome (reverse model))) new-list)
))
```

;;; The function Add-signs is called in RepresentContext. It accepts
;;; a list of facts and conditions respectively. They update the
;;; representation of positive events from (name value) to (name nil value)
;;; to indicate that the value is positive.The representation of negative
;;; events, i.e., (name not value) is left unchanged.

```
(defun Add-signs (list-of-conditions)
  (do* ((conditions (reverse list-of-conditions) (rest conditions))
        (condition (first conditions) (first conditions))
        (Possibility-Models nil) )
       ((null conditions) Possibility-Models)
    (push (enter-signs condition) Possibility-Models)
))
```

;;; Enter-signs is called in Add-signs and it accepts one condition.
;;; It takes one token at a time and if the token is not a footnote, if
;;; calls enter-sign to add insert the sign.

```
(defun enter-signs (one-condition)
  (do ((condition (reverse one-condition) (rest condition))
       (new-condition nil))
      ((null condition) new-condition)
    (if (GetFootnote (first condition))
      (push (first condition) new-condition)
      (push (enter-sign (first condition)) new-condition))
))
```

;;; Enter-sign is called in list-facts and add-signs. It takes the an event
;;; and inserts the sign, i.e., nil, if there is not one already.

```
(defun enter-sign (item)
  (if (eq (length item) 3) item (list (first item) nil (second item))))
```

;;; CONSTRUCT A MODEL TO REPRESENT THE FACTS

;;; The function RepresentFacts is called in GenerateCounterfactual. It
;;; accepts a list of facts and it updates the representation of positive
;;; events from (name value) to (name nil value) to indicate that the value
;;; is positive. The representation of negative events, i.e.,
;;; (name not value) is left unchanged.

```
(defun RepresentFacts (fact-list)
  (do* ((facts (reverse fact-list) (rest facts))
        (fact (first facts) (first facts))
        (Fact-Model nil) )
       ((null facts) (PrintModel Fact-Model 'f) Fact-Model)
    (push (enter-sign fact) Fact-Model)
))
```

;;; MUTATING AN ANTECEDENT EVENT

;;; MutateFacts is called in GenerateCounterfactual. It accepts the
;;; Fact-Model, e.g.,
;;; ((John nil Red) (Michael nil Black) (Outcome nil Lose)).
;;; It selects the most recent antecedent event to mutate. Because the
;;; last item in the factual model is the outcome, it selects the second
;;; last item. The function mutates this item by negating it
;;; (in NegateAntecedent). It returns the mutated event, e.g.,
;;; (Michael not Black).

```
(defun MutateFacts (Fact-Models)
  (let* ((AntecedentFacts (antecedents Fact-Models))
         (LastEvent (first (last AntecedentFacts))))
    (NegateAntecedent LastEvent)
))
```

;;; NegateAntecedent is called in MutateFacts. It accepts an event
;;; and negates its value. If the value of the item is affirmative it
;;; assigns 'not' to the value. If is negative then it removes the 'not'.
;;; Given the event (Michael nil Black), the function returns
;;; (Michael not Black).

```
(defun NegateAntecedent (event)
  (cond ((eq (second event) nil) (substitute 'not nil event))
        ((eq (second event) 'not) (substitute nil 'not event))
))
```

;;; INSPECTION OF THE CONTEXT OF POSSIBILITIES

;;; CheckContext is called in GenerateCounterfactual. It checks to see if the newly
;;; constructed putative counterfactual model leads to a new consequent. It does this
;;; by comparing it to the winning conditions represented in the Possibility-Models.
;;; It accepts the the Possibility-Models, e.g.,
;;; (((John nil Red)(Michael nil Red)(Outcome nil Win))
;;; ((John nil Black)(Michael nil Black)(Outcome nil Win))
;;; ((Outcome not win)))
;;; and the putative Counterfact-Model, e.g.,
;;; ((John nil Red) (Michael not Black) (Outcome nil Lose)).
;;; It calls ExamineContext to find the outcome of the Counterfact-Model. It then
;;; calls UpdateModel to update Counterfact-Model with the new outcome if
;;; there is one. It returns the upated model, e.g.,
;;; ((john nil red) (michael not black) (outcome nil win))

```
(defun CheckContext (Possibility-Models Counterfact-Model)
  (let* ((consequent
          (ExamineContext Possibility-Models Counterfact-Model))
         (Counterfact-Model (UpdateModel consequent Counterfact-Model)))
Counterfact-Model))
```

;;; ExamineContext is called in CheckContext and it takes the same input as that
;;; function, i.e., the Possibility-Models and the putative Counterfact-Model. It takes
;;; one Possibility-Model at a time and it calls MatchPossibility to see if all of
;;; the antecedent conditions are contained in the Counterfact-Model. If they
;;; are, it returns the consequent of that Possibility-Model. If the program
;;; arrives at the the consequent of that Possibility-Model. If the last condition
;;; without having found a match, and the last condition contains a consequent
;;; and no antecedents (indicating that all the other consequents have been
;;; exhaustively represented), then the function returns this consequent, e.g.,
;;; ((Outcome nil Lose))

```
(defun ExamineContext (Possibility-Models Counterfact-Model)
  (do* ((CFModels Possibility-Models (rest CFModels))
        (Possibility-Model (first CFModels) (first CFModels))
       (PrevModel nil)
       (NextModel (second CFModels) (second CFModels))
       (Footnote (FindFootnote Possibility-Model) (FindFootnote Possibility-Model))
       (Explicit-Model (FleshOut Possibility-Model Footnote PrevModel NextModel)
                       (FleshOut Possibility-Model Footnote PrevModel NextModel))
       (PrevModel Possibility-Model Possibility-Model)
         (antecedents (antecedents Explicit-Model) (antecedents Explicit-Model))
       (consequent (last Explicit-Model) (last Explicit-Model)) )
      ((null CFModels) nil)
    (cond ((MatchPossibility antecedents Counterfact-Model) (return consequent))
          ((null antecedents) (return consequent))
)))
```

```
;;; FindFootnote is called in ExamineContext. It accepts a model
;;; representing a set of antecedent conditions. It takes one token at a
;;; time and it calls GetFootnote to see if it is a footnote. If it is,
;;; then it returns the Footnote, otherwise it returns nil.

(defun FindFootnote (model)
  (do* ((tokens model (rest tokens))
        (token (first tokens) (first tokens))
        (Footnote (GetFootnote token) (GetFootnote token)))
       ((or (null tokens) Footnote) Footnote)
))



;;; GetFootnote is called in FindFootnote. It takes one token from a model
;;; and it checks to see if it is a footnote. If it is, it returns the
;;; token. Otherwise, it returns nil.

(defun GetFootnote (token)
         (if (member token '((not next model) (not prev model)) :test #'equal)
                                                                  token))



;;; FleshOut is called in ExamineContext. It accepts the
;;; Counterfactual-Model, a footnote if the Counterfactual-Model contains
;;; one, the previous Counterfactual-Model and the following one. If there
;;; is a footnote, then depending on its content, FleshOut negates the
;;; item contained either in the previous or the next model. It replaces the
;;; footnote with this negated item. It then returns the fully explicit model.

(defun FleshOut (Possibility-Model Footnote PrevModel NextModel)
  (let* ((explicit-item (cond ((equal Footnote '(not next model))
                               (negate-model (antecedents NextModel)))
                              ((equal Footnote '(not prev model))
                               (negate-model (antecedents PrevModel)))))
       (Explicit-Model (if explicit-item
                           (substitute explicit-item Footnote Possibility-Model)
                           Possibility-Model)))
    Explicit-Model)
)



;;; Negate-model is called in FleshOut. It accepts the antecedent conditions
;;; of a model. These antecedents will contain just one explicit token
;;; representing an event. It takes one antecedent at a time. If the
;;; token is a footnote it does nothing. If the token is explicit, then it calls
;;; negate-antecedent and then it returns the negated antecedent.
```

```
(defun negate-model (model)
  (do* ((antecedents model (rest antecedents))
        (antecedent (first antecedents) (first antecedents)))
       ((null antecedents) nil)
    (if (GetFootnote antecedent) nil (return (negate-antecedent antecedent)))
))
```

;;; Negate-antecedent is called in negate-model and RepresentContext.
;;; It takes one antecedent event. If its value is negative, it changes it
;;; to positive and if it is positive it changes it to negative. It returns
;;; the negated event.

```
(defun negate-antecedent (event)
  (cond ((eq (second event) nil) (substitute 'not nil event))
        ((eq (second event) 'not) (substitute nil 'not event))
))
```

;;; MatchPossibility is called in ExamineContext. It checks to see if
;;; all of the antecedents of the Possibility-Model are present in the
;;; Counterfact-Model. It accepts a list of the antecedents of a
;;; Possibility-Model, e.g.,
;;; ((john nil red)(michael nil red))
;;; and the Counterfact-Model, e.g.,
;;; ((john nil red)(michael not black)(outcome nil lose))
;;; It takes each of the antecedents in turn and compares them to the events
;;; in the Counterfact-Model. If a difference is found, it returns
;;; nil, otherwise it returns true.

```
(defun MatchPossibility (antecedent-list Counterfact-Model)
  (do* ((antecedents antecedent-list (rest antecedents))
        (antecedent (first antecedents) (first antecedents)))
        ((null antecedents) t)
    (cond ((not (compare antecedent Counterfact-Model)) (return nil)))
))
```

;;; Antecedents is called in MutateFacts and ExamineContext. It accepts
;;; a model containing a set of antecedent conditions and a consequent. It
;;; removes the consequent and it returns the antecedents.

```
(defun antecedents (model)
  (remove (first (last model)) model ))
```

;;; UpdateModel is called in CheckContext. It takes a model and a list
;;; of events with new values. It then calls replace1 which substitutes
;;; these new values for the old values in the model and it returns the
;;; new model.

```
(defun UpdateModel (event-list model)
  (do ((events (reverse event-list) (rest events))
       (new-model model (replace1 (first events) new-model)) )
      ((null events) new-model )
))
```

;;; Replace1 is called in GenerateCounterfactual and UpdateModel. It
;;; accepts an event, e.g.,
;;; (michael red)
;;; and a scenario, e.g.,
;;; ((john red) (michael black)).
;;; It updates the scenario by substituting the new event for the old event
;;; with the same name. It returns the new scenario, i.e.,
;;; ((john red) (michael red)).

```
(defun replace1 (new-event model)
        (substitute new-event (find-event new-event model) model))
```

;;; MATCHING FUNCTIONS

;;; Compare is called in combine-models, same-model and MatchPossibility.
;;; It takes an event, e.g.,
;;; (michael not black)
;;; and a scenario in the form of a list of events, e.g.,
;;; ((john nil red) (michael nil red))
;;; It calls find-event to see if the event is contained in the scenario
;;; and it tests to see if the values match. If so, it returns true.

```
(defun compare (event scenario)
        (let ((imagined-event (find-event event scenario)))
                (cond ((same-value (enter-sign event) (enter-sign imagined-event)) t)
)))
```

```
;;; Same-value is called in compare. It takes two tokens, e.g.,
;;;     (john nil red) and (john not black)
;;; and it checks to see if they are equivalent. It treats the values as
;;; binary, so it assumes that the value 'Red' is equal to the value
;;; 'not Black'. If the two tokens are equal it returns true, otherwise it
;;; returns nil.

(defun same-value (event1 event2)
  (cond ((equal event1 event2) t)
        ((and (eq (first event1) (first event2))
          (not (eq (third event1) (third event2)))
          (not (eq (second event1) (second event2))))
       t)
))



;;; Find-event is called in combine-models, compare and replace1. It
;;; accepts an event and a scenario and it calls same-event to see if the
;;; event is contained in the scenario, regardless of whether the values
;;; are the same. If a match is found, it returns the event contained in
;;; the scenario. E.g., given the event
;;;     (michael nil red)
;;; and the scenario
;;;     ((john nil red)(michael nil black)(outcome nil lose))
;;; the function returns
;;;     (michael nil black).

(defun find-event (event scenario-list)
        (do ((scenario scenario-list (rest scenario)))
                ((null scenario) nil)
        (cond ((same-event event (first scenario)) (return (first scenario))))
))



;;; Same-event is called in find-event. It accepts two events and checks to
;;; see if the first element in each list is the same (i.e., both items
;;; refer to the same event although they may have different values). If
;;; so, it returns true, otherwise it returns nil.

(defun same-event (event1 event2)
        (cond ((eq (first event1) (first event2)) t) ))
```

```
;;; PRINT MODELS
;;;
;;; PrintModel prints out models. It accepts a scenario as a list, e.g.,
;;;      ((john nil red)(michael nil black)(outcome nil lose)
;;; and the status of the scenario, i.e., f for factual, cf for
;;; counterfactual, sf for semifactual and p for possibility. It prints out
;;; the model in the following format:
;;;
;;; Factual:    John Red     Michael Black  Outcome Lose

(defun PrintModel (model status)
  (format t "~%~%")
  (cond ((eq status 'f)  (format t "~&~%Factual:        "))
        ((eq status 'cf) (format t "~&~%Counterfactual: "))
        ((eq status 'sf) (format t "~&~%Semifactual: "))
        ((eq status 'p)  (format t "~&Possibility:    ")))
  (dolist (event model)
    (format t "~S  " (first event) )
    (cond ((eq (length event) 2) (format t "~S       " (second event)))
          ((eq (second event) 'nil) (format t "~S       " (third event)))
          (t (format t "~S ~S   " (second event) (third event)))))  )
```

Appendix 2: TempContext program code

;;; A simulation of the context theory of the temporal order effect in
;;; counterfactual thinking.
;;; The aim of this program is to construct a counterfactual conditional from a factual
;;; scenario and a conditional. It simulates a theory of the processes underlying the
;;; temporal order effect.

;;; ***Representation of the Facts***
;;; The program accepts a factual scenario as a series of antecedent events and an
;;; outcome. Each event is assigned a name and a value. For example, an event might be
;;; named 'Johns-card' and might be assigned a value 'Red'. To denote that the value is
;;; negative, the model can be tagged with the word 'not' e.g., (John not Red). If the
;;; event is not tagged, the program assumes that the value is positive. The program also
;;; assumes that the events are presented in their temporal order of occurrence. For
;;; example, given the following facts:

;;; 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.

;;; The program accepts the following list of facts:

;;; ((Johns-card Red) (Michaels-card Black) (Outcome Lose))

;;; ***Representation of the Context of Possibilities***
;;; The program also accepts a set of conditions that determine the outcome. The
;;; conditions contain a list of events without their values, a set of antecedent
;;; conditions and the consequent. Take for example, the following conditional:

;;; If the two cards they pick are of the same colour (i.e., both red or both black),
;;; each individual wins £1,000. Otherwise, neither individual wins anything.

;;; The program accepts this conditional in the following format:

```
;;; (((Johns-card) (Michaels-card))    ..... the events
;;;  ((Both Red) or (Both Black))      ..... the antecedent conditions
;;;  ((Outcome Win)))                  ..... the outcome.
```

;;; From this conditional, the program constructs the following mental models:

```
;;;      Johns-card  Red      Michaels-card  Red      Outcome Win
;;;      Johns-card  Black    Michaels-card  Black    Outcome Win
```

;;; The program aims to minimise the amount of information that it represents explicitly.
;;; For this reason, it does not represent the losing conditions explicitly and it only
;;; represents some of the winning conditions explicitly. For example, take a case
;;; where the winning conditions are described as follows:
;;;
;;; If one player picks a card from a red suit, then each individual wins £1,000.

;;;
;;; The program constructs an explicit representation of one possibility in which one of
;;; the players picks a red card, e.g.,
;;;
;;; Johns-card Red Outcome Win
;;;
;;; It does not explicitly represent the second possibility in which Michael picks a red
;;; card. The representation of a disjunction also contains implicit information. Take
;;; for example the following exclusive disjunction:
;;;
;;; If one or the other but not both pick a card from a red suit.
;;;
;;; The program constructs a representation of two possibilities. In one, John picks a
;;; red card is represented explicitly but it is only implicit in this model that Michael
;;; picks a black card. In the second model, Michael picks a red card is explicit and
;;; John picks a black card is implicit, i.e.,
;;;
;;; Johns-card Red *(not next model)* Outcome Win
;;; Michaels-card Red *(not prev model)* Outcome Win
;;;

;;; TOP-LEVEL FUNCTION - GenerateCounterfactual

**;;; The top-level function is GenerateCounterfactual. This function goes
;;; through the steps involved in generating a counterfactual conditional.**

**;;; The function begins by calling RepresentFacts and RepresentContext
;;; to construct models of the facts (Fact-Model) and winning conditions
;;; (Possibility-Models) respectively. RepresentFacts calls PrintModel to
;;; print a model of the facts to the screen, e.g.,**

;;; Factual: Johns-card Red Michaels-card Black Outcome Lose

**;;; The function then calls MatchFirst to see if the first fact is represented
;;; explicitly in any of the models. If begins by looking for an event that has the
;;; same value. If possible, it selects a counterfactual model which contains just
;;; one event which is different from the factual model. If no model contains an
;;; explicit match for the first fact then MatchFirst tries to find a model which
;;; contains the negation of the first fact, i.e., John picks not red. Again, it looks
;;; for a minimally different counterfactual model.
;;;
;;; Once a model is selected the program calls MutateFacts to create a new
;;; counterfactual model. It does this by modifying the factual scenario to
;;; include the new counterfactual events.
;;;
;;; Finally GenerateCounterfactual prints the counterfactual model and it
;;; prints a conditional in the following form: If "the mutated event"
;;; then "the new outcome".**

```
(defun GenerateCounterfactual (facts conditional)

  (let* ((Possibility-Models (RepresentContext (first conditional)
                             (second conditional) (third conditional)))
         (Fact-Model (RepresentFacts facts))

        (Possibility-Model (MatchFirst Fact-Model Possibility-Models))
        (Counterfact-Model
         (if Possibility-Model (MutateFacts Possibility-Model Fact-Model)))

        (MutatedEvent (first Possibility-Model))
        (MutatedEvent2 (second Possibility-Model))
        (CounterfactConsequent (first (last Possibility-Model)))
        (FactualConsequent (first (last Fact-Model)))
        (CounterfactStatus (if (equal CounterfactConsequent FactualConsequent) 'sf 'cf))

)
    (dolist (Counterfactual Possibility-Models) (PrintModel Counterfactual 'p))
    (PrintModel Counterfact-Model CounterfactStatus)

    (if (eq (length Possibility-Model) 2)
      (format t "~&~%~%If it had been the case that ~S
                ~% then it would have been the case that ~S."
              (remove nil MutatedEvent)
              (remove nil CounterfactConsequent))
      (format t "~&~%~%If it had been the case that ~S and ~S
                ~% then it would have been the case that ~S."
              (remove nil MutatedEvent)
              (remove nil MutatedEvent2)
              (remove nil CounterfactConsequent)))
    ))
```

;;; CONSTRUCT MODELS TO REPRESENT THE CONTEXT

;;; RepresentContext is the top-level of a set of functions which
;;; constructs a set of models to represent the conditional. It is called
;;; in GenerateCounterfactual. It calls ModelContext to construct a
;;; a set of models to represent the antecedent conditions.
;;; RepresentContext then adds the consequent to each model and adds a
;;; sign to events with positive values. For example, given the event-list
;;; (johns-card michaels-card)
;;; the antecedent conditions
;;; ((both red) or (both black))
;;; and the outcome
;;; (outcome win)
;;; the function returns the following set of models:
;;;
;;; (((johns-card nil red) (michaels-card nil red) (outcome nil win))
;;; ((johns-card nil black) (michaels-card nil black) (outcome nil win)))

;;; Note, in the following functions:
;;; An event is represented by a single list, e.g.,
;;; (Johns-card red),
;;; a model contains a list of events, e.g.,
;;; ((Johns-card red) (Michaels-card red))
;;; and a set of models contains a list of models, e.g.,
;;; (((Johns-card red) (Michaels-card red))
;;; ((Johns-card Michaels-card red))).

(defun RepresentContext (events antecedent-conditions outcome)
 (add-signs (add-outcome (ModelContext events antecedent-conditions) outcome)))

;;; Model-Context is called in RepresentContext. From the conditional
;;; it accepts the list of events without their values, e.g.,
;;; (johns-card michaels-card)
;;; and the antecedent conditions. These conditions may directly specify
;;; the value to be assigned to a particular event, e.g., 'John picks a
;;; red card' and this will be represented as follows:
;;; ((johns-card red)).
;;; Alternatively the conditions may specify that both, one or only-one
;;; event/s have a particular value, e.g., 'both players pick a red card' is
;;; represented as follows:
;;; ((both red)).
;;; ModelContext also accepts four connectives: and, or, ori (or inclusive)
;;; and ore (or exclusive). The antecedent conditions may take the following
;;; format:
;;; (((one red) and (one black)) or (both black))
;;; or
;;; ((john red) or (michael red)).

```
;;; In order to interpret the conditional, ModelContext takes one item from
;;; the antecedent conditions at a time. If the item contains an embedded list, e.g.,
;;;    ((one red) and (one black))
;;; then the function calls itself again and puts the antecedent conditions equal to
;;; this item. If the item is a connective, then it is put on the stack. Otherwise the
;;; function calls CreateModels to generate models to represent the item. It assumes
;;; that people minimise the amount of information that they make explicit in their
;;; models because of working memory limitations (Johnson-Laird & Byrne, 1991).
;;; It puts these models onto a list called Possibility-Models.
;;;
;;; Next, the function looks to see if there is a connective on the stack. If there is
;;; and the connective has not just been put there, it takes the two sets of models
;;; in Possibility-Models and it joins them with the connective. It replaces the two
;;; sets of models with the new combined set of models. It then reads the next item
;;; on the antecedent condition list and repeats this process until the list is empty.
;;;
;;; For example, given the event-list
;;;       (johns-card michaels-card)
;;; and the antecedent conditions
;;;       ((both red) or (both black))
;;; the program constructs the following set of models:
;;;       (((johns-card red) (michaels-card red))
;;;        ((johns-card black) (michaels-card black))).

(defun ModelContext (events antecedent-conditions)
 (do* ((conditions antecedent-conditions (rest conditions))
       (item (first conditions) (first conditions))
       (stack nil)
       (events events (if (and (eq (first stack) 'and) (eq (first item) 'one))
                          (remove (caaaar Possibility-Models) events)
                          events))
       (Possibility-Models nil))
      ((null conditions) (remove-duplicates1 (first Possibility-Models)))

   (cond ((member item '(or and ori ore)) (push item stack))
         ((eq (type-of (first item)) 'cons)
          (push (ModelContext events item) Possibility-Models))
         (t (push (CreateModels events item) Possibility-Models)))

   (cond ((and stack (not (member item '(or ori ore and))))
     (push
      (funcall
       (cond ((eq (first stack) 'and) 'and-x)
             ((eq (first stack) 'ore) 'ore-x)
             ((eq (first stack) 'or) 'or-x)
             ((eq (first stack) 'ori) 'ori-x))
       (pop Possibility-Models)(pop Possibility-Models))
      Possibility-Models)
     (pop stack))) ))
```

```
;;; CreateModels is called in ModelContext. It takes the list of
;;; events before they have been assigned values, e.g.,
;;;      (johns-card michaels-card)
;;; and one item from the antecedent conditions, e.g.,
;;;      (both red).
;;; If the item contains one, only-one or both, then it calls the
;;; corresponding function with that name. Otherwise, if the item refers to
;;; a specific event, e.g.,
;;;      (johns-card red)
;;; then the function returns the item embedded in two more lists, i.e.,
;;;      (((johns-card red))).

(defun CreateModels (events item)
  (cond ((member (first item) '(one only-one both))
         (funcall (first item) events (rest item)))
        (t (list (list item)))   ))


;;; Only-one is called in CreateModels. It accepts a list of two events,
;;; e.g.,
;;;      (johns-card  michaels-card)
;;; and the value to be assigned to those events, e.g.,
;;;      (Red).
;;; It constructs two models; each one contains an explicit item in which
;;; the value is assigned to one of the events. Each model also contains a
;;; footnote to note that the other model is false. For example, given
;;; the input described above, the function returns
;;;      (((johns-card red) (not next model))
;;;       ((michaels-card red) (not prev model))).

(defun only-one (event-pair value)
          (list
          (list (cons (first event-pair) value) '(not next model))
          (list (cons (second event-pair) value) '(not prev model)))  )


;;; One is called in CreateModels. It accepts a list of events, e.g.,
;;;      (johns-card  michaels-card)
;;; and the value to be assigned to those events, e.g.,
;;;      (Red).
;;; It constructs one explicit model by assigning the value to one of the
;;; events at random, e.g.,
;;;      (johns-card red).
;;; The remaining possibilities, e.g., (michaels-card red) are implicit.

(defun one (event-list value)
  (do* ((random-no (random (length event-list)) (- random-no 1))
        (events event-list (rest events)))
    ((eq random-no 0) (list (list (cons (first events) value))))    ))
```

```
;;; Both is called in CreateModels. It accepts a list of two events,
;;; e.g.,
;;;      (johns-card michaels-card)
;;; and the value to be assigned to those events, e.g.,
;;;      (not Red).
;;; It constructs one model which assigns the value to the two events.
;;; For example, given the input described above, the function returns
;;;      (((johns-card red) (michaels-card red)))

(defun both (event-pair value)
  (list (list (cons (first event-pair) value) (cons (second event-pair) value)))
)



;;; Or-x is called in ModelContext. It accepts two sets of models,
;;; e.g.,
;;;     (((Johns-card red)))
;;; and
;;;     (((Michaels-card black)))
;;; which are to be linked by the connective 'or'. It returns a new set of
;;; models containing the first set of models and the second set of models.
;;; For example, given the above input, the function returns
;;;     (((Michaels-card black))
;;;      ((Johns-card red)))

(defun or-x (models-1 models-2)
  (append models-2 models-1))



;;; Ori-x is called in ModelContext. It accepts two sets of models,
;;; e.g.,
;;;     (((Johns-card red)))
;;; and
;;;     (((Michaels-card black)))
;;; which are to be linked by the connective 'ori' (or inclusive). It
;;; returns a new set of models containing the first set of models, the
;;; second set of models and a conjunction of the first and second set. For
;;; example, given the above input, the function returns
;;;     (((Michaels-card black))
;;;      ((Johns-card red))
;;;      ((Michaels-card black) (Johns-card red)))

(defun ori-x (models-1 models-2)
        (let ((new-list (append models-2 models-1)))
        (append new-list (and-x models-1 models-2))
))
```

```
;;; Ore-x is called in ModelContext. It accepts two sets of models,
;;; e.g.,
;;;     (((Johns-card red)) ((Michaels-card red)))
;;; and
;;;     (((Michaels-card black)))
;;; which are to be linked by the connective 'ore'. It combines the two sets
;;; of models into one. For example, given the above input, the function
;;; returns
;;;     (((Michaels-card black)) ((Johns-card red)) ((Michaels-card red)))
;;;
;;; The function also adds footnotes to the models but only where the models
;;; are very simple. If each set of models contains just one model and each
;;; model contains just one token, then the function adds footnotes to the
;;; models and returns, for example,
;;;     (((Johns-card red) (not next model))
;;;      ((Michaels-card red) (not prev model))).
;;; It is assumed that people have difficulty using footnotes with more
;;; complex models or sets of models and so the program does not represent
;;; these.

(defun ore-x (models-1 models-2)
  (cond ((and (eq (length models-1) 1) (eq (length (first models-1)) 1)
              (eq (length models-2) 1) (eq (length (first models-2)) 1))
         (list (list (first (first models-2)) '(not next model))
                     (list (first (first models-1)) '(not prev model))))
        (t (append models-2 models-1)))
)



;;; And-x is called in ModelContext. It accepts two sets of models
;;; which are to be linked by the connective 'and', e.g.,
;;;     (((Johns-card red)))
;;; and
;;;     (((Michaels-card black))).
;;; It takes one model at a time from the first set of models and it calls
;;; combine to combine this model with each of the models in the second set.
;;; It returns a new combined set of models, e.g.,
;;;     (((Michaels-card black) (Johns-card red))).

(defun and-x (models-1 models-2)
        (do* ((first-list models-1 (rest first-list))
                    (model-1 (first first-list) (first first-list))
                    (new-list nil (append (first new-list) (rest new-list))) )
        ((null first-list) (if (null models-1) models-2 new-list))
        (push (combine models-2 model-1) new-list)
))
```

;;; Combine is called in and-x. It accepts one set of models e.g.,
;;; (((Johns-card red)) ((Michaels-card red)))
;;; This set of models represents two possibilities, John selecting a red
;;; card and Michael selecting a red card. It also accepts a single model,
;;; e.g.,
;;; ((Johns-card red)(Michaels-card black))
;;; This model represents a single possibility in which John selects a red
;;; card and Michael selects a black card. The function takes each of the
;;; models in the set in turn and combines it with the single model. It
;;; removes footnotes from the first model in the set and from the single
;;; model. This is because it assumes that people have difficultly
;;; combining models that contain footnotes. It then calls combine-models
;;; to combine the two models. It puts the combined models in a list and
;;; returns a new set of models. For example, given the example above, the
;;; function returns a set which contains a single model, i.e.,
;;; (((Johns-card red) (Michaels-card black)))
;;; which represents the possibility of Michael selecting a black card and
;;; John selecting a red card.

```
(defun combine (list-of-models 1-model)
 (do* ((condition-list list-of-models (rest condition-list))
       (condition (first condition-list) (first condition-list))
       (condition (remove-footnotes condition) (remove-footnotes condition))
       (1-model (remove-footnotes 1-model) 1-model)
       (new-condition (combine-models condition 1-model)
                      (combine-models condition 1-model))
       (new-condition-list (list new-condition) (push new-condition new-condition-list)))
      ((eq (length condition-list) 1) (remove nil new-condition-list))
))
```

;;; Remove-footnotes is called in Combine. It accepts a model and removes
;;; any footnotes from it.

```
(defun remove-footnotes (model)
 (let* ((model1 (remove '(not next model) model :test #'equal))
        (model1 (remove '(not prev model) model1 :test #'equal)))
   model1
))
```

```
;;; Combine-models is called in combine to combine two models. It accepts
;;; two models e.g.,
;;;      ((Johns-card red))
;;; which represents one event in which John selects a red card, and
;;;      ((Michaels-card black))
;;; which also represents a single event, Michael selecting a black card.
;;; It takes each event in the second model in turn. If the event is already
;;; contained in the first model with the same value, then it does nothing.
;;; If the event is contained in the first model but they both have
;;; different values, then the function returns nil because the two models
;;; cannot be combined. If the event is not contained in the first model,
;;; then it adds the event to the model. It returns the combined model.
;;; Given the example above, the function returns the following model,
;;;      (((Johns-card red) (Michaels-card black)))
;;; which represents the possibility that Michael selects a black card and
;;; John selects a red card.

(defun combine-models (model-1 model-2)
  (do* ((model2 model-2 (rest model2))
        (event (first model2) (first model2))
        (new-model (reverse model-1)))
       ((null model2) (reverse new-model))
    (cond ((compare event model-1) nil)
          ((find-event event model-1) (return nil))
          (t (push event new-model)))
))



;;; Remove-duplicates1 is called in ModelContext to remove
;;; duplicates from a set of models. It takes one model at a time and it
;;; calls find-duplicate to see if there is a duplicate model in the
;;; remaining set of models. If no duplicate is found, the model is placed
;;; on a new list. After checking all the models, the program returns the
;;; new list.

(defun remove-duplicates1 (list-of-models)
  (do* ((models list-of-models (rest models))
        (model (first models) (first models))
        (new-models nil))
       ((null models) (reverse new-models))
    (cond ((not (find-duplicate model (rest models)))
           (push model new-models)))
))
```

;;; Find-duplicate is called in remove-duplicates. It accepts a single model
;;; and a list of models. It takes one model from the list at a time and it
;;; calls same-models to see if this model is equal to the single model.
;;; If a match is found, it returns true. Otherwise, it returns nil.

```
(defun find-duplicate (model list-of-models)
 (do* ((models list-of-models (rest models))
       (1st-model (first models) (first models)))
      ((null models) nil)
   (if (same-model model 1st-model)
     (return t)
     nil)
))
```

;;; Same-model is called in find-duplicates. It accepts two models. It
;;; takes the second model and it goes through one event at a time checking
;;; to see if the event is contained in the first model. If an event is
;;; found which is not contained in the first model, it returns nil.
;;; Otherwise, if all of the events in the second model are in the first and
;;; the two models contain the same number of events, then the function
;;; returns true.

```
(defun same-model (model-1 model-2)
 (do* ((events-2 model-2 (rest events-2))
       (event (first events-2) (first events-2)))
      ((null events-2) (if (eq (length model-1) (length model-2)) t))
   (if (compare event model-1) nil (return nil))
))
```

;;; Add-outcome is called in RepresentContext. It accepts the set of
;;; models which have been constructed to represent the antecedent
;;; conditions and the consequent. It adds the consequent to each of the
;;; models and returns the new set of models.

```
(defun add-outcome (model-list outcome)
   (do* ((models (reverse model-list) (rest models))
         (model (first models) (first models))
         (new-list nil))
        ((null models) new-list)
     (push (reverse (cons outcome (reverse model))) new-list)
))
```

```
;;; The function Add-signs is called in RepresentContext. It accepts
;;; a list of facts and conditions respectively. They update the
;;; representation of positive events from (name value) to (name nil value)
;;; to indicate that the value is positive.The representation of negative
;;; events, i.e., (name not value) is left unchanged.

(defun Add-signs (list-of-conditions)
 (do* ((conditions (reverse list-of-conditions) (rest conditions))
       (condition (first conditions) (first conditions))
       (Possibility-Models nil) )
      ((null conditions) Possibility-Models)
   (push (enter-signs condition) Possibility-Models)   ))


;;; Enter-signs is called in Add-signs and it accepts one condition.
;;; It takes one token at a time and if the token is not a footnote, if
;;; calls enter-sign to add insert the sign.

(defun enter-signs (one-condition)
 (do ((condition (reverse one-condition) (rest condition))
      (new-condition nil))
     ((null condition) new-condition)
   (if (GetFootnote (first condition))
     (push (first condition) new-condition)
     (push (enter-sign (first condition)) new-condition))    ))


;;; Enter-sign is called in list-facts and add-signs. It takes the an event
;;; and inserts the sign, i.e., nil, if there is not one already.

(defun enter-sign (item)
 (if (eq (length item) 3) item (list (first item) nil (second item))))



;;; CONSTRUCT A MODEL TO REPRESENT THE FACTS

;;; The function RepresentFacts is called in GenerateCounterfactual. It
;;; accepts a list of facts and it updates the representation of positive
;;; events from (name value) to (name nil value) to indicate that the value
;;; is positive. The representation of negative events, i.e.,
;;; (name not value) is left unchanged.

(defun RepresentFacts (fact-list)
 (do* ((facts (reverse fact-list) (rest facts))
       (fact (first facts) (first facts))
       (Fact-Model nil) )
      ((null facts) (PrintModel Fact-Model 'f) Fact-Model)
   (push (enter-sign fact) Fact-Model)    ))
```

;;; MATCHING THE FACTS AND THE COUNTERFACTUAL CONTEXT

;;; MatchFirst is called in GenerateCounterfactual. It accepts the Fact-Model
;;; and the Possibility-Models. It checks to see if the first event is
;;; represented explicitly in any of the models. First, it calls Findmatch
;;; to look for an event with the same value. If possible, it selects a
;;; counterfactual model which contains just one event which is different
;;; from the factual model. If no model contains an explicit match for the
;;; first fact then MatchFirst calls FindAlternative to look for a model
;;; which contains the negation of the first fact, e.g., John picks not red.
;;; Again, it looks for a minimally different counterfactual model.

```
(defun MatchFirst (Fact-Model Possibility-Models)
 (let* ((Possibility-Model (FindMatch Fact-Model Possibility-Models))
      (Possibility-Model (if Possibility-Model
                  Possibility-Model
                  (FindAlternative Fact-Model Possibility-Models))))
  Possibility-Model
))
```

;;; FindMatch is called in MatchFirst It takes the Fact-Model,
;;; e.g.,
;;; ((johns-card nil red) (michaels-card nil black) (outcome nil lose))
;;; and the Possibility-Models, e.g.,
;;; (((johns-card nil red) (michaels-card nil red) (outcome nil win))
;;;   ((johns-card nil black) (michaels-card nil black) (outcome nil win))).
;;; It takes the first fact and it looks in each of the Possibility-Models
;;; to see if it contains an event with the same value. In the above
;;; example, there is a match for first fact (johns-card nil red) in the
;;; first model.
;;;
;;; If a match is found, the function calls FindFootnote to see if the
;;; matched model contains a footnote. If so, it fleshes out the
;;; model to make this item explicit. The function then calls CompareModels
;;; to compare the explicit model to the facts. This returns a list of the
;;; facts which are different from the antecedents in the model. If no
;;; difference is found, it means that the model is factual and the
;;; function continues to search for a semifactual model. If one difference
;;; is found, the function returns this difference along with the consequent
;;; of the model. If two of the facts are different from the model, the
;;; function stores the difference and the consequent in a list called
;;; counterfactual and it continues to look for a model which is more like
;;; the facts. If none is found it returns this counterfactual.
;;;   Alternatively, if no match at all for the first fact is found the
;;; function returns nil.

```
(defun FindMatch (Fact-Model Possibility-Models)
  (do* ((conditions (push 'x Possibility-Models) (rest conditions))
        (Possibility-Model nil (first conditions))
        (NextModel nil (second conditions))
        (1st-fact nil (first Fact-Model))
        (match nil (compare 1st-fact Possibility-Model))

        (Footnote nil (if match (FindFootnote (antecedents Possibility-Model))))
        (Explicit-Model nil (if Footnote
                                (FleshOut Possibility-Model Footnote PrevModel NextModel)
                                Possibility-Model))

        (difference nil (if match (CompareModels Fact-Model Explicit-Model)))
        (PrevModel nil Possibility-Model)
        (counterfactual nil))
       ((null conditions) (first counterfactual))
    (cond ((eq difference nil) nil)
          ((> (length difference) 1)
                (push (append difference (last Possibility-Model)) counterfactual))
          ((eq (length difference) 1)
                (return (append difference (last Possibility-Model)))))    ))
```

;;; FindAlternative is called in MatchFirst It is identical to FindMatch
;;; except that instead of looking for an explicit match in the
;;; Possibility-Models, this function also looks for an explicit alternative.

```
(defun FindAlternative (Fact-Model Possibility-Model)
  (do* ((conditions (push 'x Possibility-Model) (rest conditions))
        (Possibility-Model nil (first conditions))
        (NextModel nil (second conditions))
        (1st-fact nil (first Fact-Model))
        (match nil (find-event 1st-fact Possibility-Model))

        (Footnote nil (if match (FindFootnote (antecedents Possibility-Model))))
        (Explicit-Model nil (if Footnote
                                (FleshOut Possibility-Model Footnote PrevModel NextModel)
                                Possibility-Model))

        (difference nil (if match (CompareModels Fact-Model Explicit-Model)))
        (PrevModel nil Possibility-Model)
        (counterfactual nil))
       ((null conditions) (first counterfactual))
    (cond ((eq difference nil) nil)
          ((> (length difference) 1)  (push (append difference (last Possibility-Model))
                                            counterfactual))
          ((eq (length difference) 1) (return (append difference (last Possibility-Model)))))
))
```

```
;;; FindFootnote is called in FindMatch and FindAlternative. It accepts a
;;; model representing a set of antecedent conditions. It takes one token
;;; at a time and it calls GetFootnote to see if it is a footnote. If it is,
;;; then it returns the Footnote, otherwise it returns nil.

(defun FindFootnote (model)
 (do* ((tokens model (rest tokens))
       (token (first tokens) (first tokens))
       (Footnote (GetFootnote token) (GetFootnote token)))
      ((or (null tokens) Footnote) Footnote)
))



;;; GetFootnote is called in FindFootnote. It takes one token from a model
;;; and it checks to see if it is a footnote. If it is, it returns the
;;; token. Otherwise, it returns nil.

(defun GetFootnote (token)
        (if (member token '((not next model) (not prev model)) :test #'equal)
                                                              token))


;;; Antecedents is called in FindMatch, FindAlternative and CompareModels.
;;; It accepts a model containing a set of antecedent conditions and a
;;; consequent. It removes the consequent and it returns the antecedents.

(defun antecedents (model)
        (remove (first (last model)) model ))



;;; FleshOut is called in FindMatch and FindAlternative. It accepts the
;;; Counterfactual-Model, a footnote if the Counterfactual-Model contains
;;; one, the previous Counterfactual-Model and the following one. If there
;;; is a footnote, then depending on its content, FleshOut negates the
;;; item contained either in the previous or the next model. It replaces
;;; the footnote with this negated item. It then returns the fully explicit
;;; model.

(defun FleshOut (Possibility-Model Footnote PrevModel NextModel)
 (let* ((explicit-item (cond ((equal Footnote '(not next model))
                              (negate-model (antecedents NextModel)))
                             ((equal Footnote '(not prev model))
                              (negate-model (antecedents PrevModel)))))
     (Explicit-Model (if explicit-item
                         (substitute explicit-item Footnote Possibility-Model)
                         Possibility-Model)))
  Explicit-Model) )
```

;;; Negate-model is called in FleshOut. It accepts the antecedent
;;; conditions of a model. These antecedents will contain just one
;;; explicit token representing an event. It takes one antecedent at a
;;; time. If the token is a footnote it does nothing. If the token is
;;; explicit, then it calls negate-antecedent and then it returns the
;;; negated antecedent.

```
(defun negate-model (model)
 (do* ((antecedents model (rest antecedents))
       (antecedent (first antecedents) (first antecedents)))
      ((null antecedents) nil)
   (if (GetFootnote antecedent) nil (return (negate-antecedent antecedent)))
))
```

;;; Negate-antecedent is called in negate-model. It takes one antecedent
;;; event. If its value is negative, it changes it to positive and if it
;;; is positive it changes it to negative. It returns the negated event.

```
(defun negate-antecedent (event)
 (cond ((eq (second event) nil) (substitute 'not nil event))
       ((eq (second event) 'not) (substitute nil 'not event))
))
```

;;; CompareModels is called in FindMatch and FindAlternatives. It accepts
;;; the Fact-Model, e.g.,
;;; ((johns-card nil black) (michaels-card nil black) (outcome nil lose))
;;; and an explicit Possibility-Model, e.g.,
;;; ((johns-card nil red) (michaels-card not red) (outcome nil win)).
;;;
;;; It takes the antecedents of the Possibility-Model and compares them
;;; one at a time to the facts. If an antecedent is contained in the
;;; Fact-Model, it does nothing. It places the antecedents which did not
;;; actually occur in a list called difference and after all of the
;;; antecedents have been examined it returns this list.

```
(defun CompareModels (Fact-Model counterfactual)
 (do* ((cf-antecedents (antecedents counterfactual) (rest cf-antecedents))
       (difference nil))
      ((null cf-antecedents) (reverse difference))
   (if (compare (first cf-antecedents) Fact-Model)
     nil
     (push (first cf-antecedents) difference))
))
```

```
;;; MutateFacts is called in GenerateCounterfactual. It takes the
;;; Fact-Model and a list of events with new-values. It then substitutes
;;; these new values for the old values in the model and it returns the
;;; new model.

(defun MutateFacts (event-list Fact-Model)
  (do* ((events (reverse event-list) (rest events))
        (event (first events) (first events))
        (new-model (if (find-event event Fact-Model) (replace1 event Fact-Model)
                                                     (push event Fact-Model))
                   (if (find-event event new-model) (replace1 event new-model)
                                                    (push event new-model))))
       ((eq (length events) 1) new-model)

))



;;; Replace1 is called in MutateFacts. It accepts an event, e.g.,
;;;      (michael red)
;;; and a scenario, e.g.,
;;;      ((john red) (michael black)).
;;; It updates the scenario by substituting the new event for the old event
;;; with the same name. It returns the new scenario, i.e.,
;;;      ((john red) (michael red)).

(defun replace1 (new-event model)
        (substitute new-event (find-event new-event model) model))
```

```
;;; MATCHING FUNCTIONS

;;; Compare is called in combine-models, same-model, FindMatch and
;;; CompareModels. It takes an event, e.g.,
;;;             (michael not black)
;;; and a scenario in the form of a list of events, e.g.,
;;;             ((john nil red) (michael nil red))
;;; It calls find-event to see if the event is contained in the scenario
;;; and it tests to see if the values match. If so, it returns true.

(defun compare (event scenario)
        (let ((imagined-event (find-event event scenario)))
                (cond ((same-value (enter-sign event) (enter-sign imagined-event)) t)
)))



;;; Same-value is called in compare. It takes two tokens, e.g.,
;;;     (john nil red) and (john not black)
;;; and it checks to see if they are equivalent. It treats the values as
;;; binary, so it assumes that the value 'Red' is equal to the value
;;; 'not Black'. If the two tokens are equal it returns true, otherwise it
;;; returns nil.

(defun same-value (event1 event2)
 (cond ((equal event1 event2) t)
        ((and (eq (first event1) (first event2))
          (not (eq (third event1) (third event2)))
          (not (eq (second event1) (second event2))))
      t)  ))



;;; Find-event is called in combine-models, FindAlternative, MutateFacts,
;;; compare and replace1. It accepts an event and a scenario and it calls
;;; same-event to see if the event is contained in the scenario, regardless
;;; of whether the values are the same. If a match is found, it returns
;;; the event contained in the scenario. E.g., given the event
;;;       (michael nil red)
;;; and the scenario
;;;       ((john nil red)(michael nil black)(outcome nil lose))
;;; the function returns
;;;       (michael nil black).

(defun find-event (event scenario-list)
        (do ((scenario scenario-list (rest scenario)))
                ((null scenario) nil)
        (cond ((same-event event (first scenario)) (return (first scenario))))
))
```

```
;;; Same-event is called in find-event. It accepts two events and checks to
;;; see if the first element in each list is the same (i.e., both items
;;; refer to the same event although they may have different values). If
;;; so, it returns true, otherwise it returns nil.

(defun same-event (event1 event2)
        (cond ((eq (first event1) (first event2)) t) ))



;;; PRINT MODELS
;;;
;;; PrintModel prints out models. It accepts a scenario as a list, e.g.,
;;;      ((john nil red)(michael nil black)(outcome nil lose)
;;; and the status of the scenario, i.e., f for factual, cf for
;;; counterfactual, sf for semifactual and p for possibility. It prints out
;;; the model in the following format:
;;;
;;; Factual:    John Red     Michael Black  Outcome Lose

(defun PrintModel (model status)
 (format t "~%~%")
 (cond ((eq status 'f)  (format t "~&~%Factual:        "))
       ((eq status 'cf) (format t "~&~%Counterfactual: "))
       ((eq status 'sf) (format t "~&~%Semifactual: "))
       ((eq status 'p)  (format t "~&Possibility:    ")))
 (dolist (event model)
  (format t "~S  " (first event) )
  (cond ((eq (length event) 2) (format t "~S       " (second event)))
        ((eq (second event) 'nil) (format t "~S       " (third event)))
        (t (format t "~S ~S   " (second event) (third event)))))  )
```

Appendix 3: Materials used in the temporal order effect experiments

Instructions (page 1)

Thank you for agreeing to take part in this experiment. On the following page you will find a short story. Please take your time and read it carefully. The story is followed by some questions which ask for your views on it. There are no right or wrong answers. Please answer the questions in the order in which they are given, and do not change an answer once you have written it. You may refer back to the story as often as you wish. On the final page, I have outlined the aim of the experiment. Please do not read it until you have completed all the questions.

Before you turn this page, please record your gender and age in the space provided below:

Gender: Male_____ Female_____

Age: _____

The scenarios used in the three experiments (page 2)

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 their own deck.

Winning conditions
Experiment 1
Conjunctive (one black and one red)
If the two cards they pick are of different colours (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 (one red and one black)
If the two cards they pick are of different colours (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.

Experiment 3
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 colours (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.

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

Questions (page 2)

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?

De-briefing paragraph (page 3)

The aim of this study is to examine what people imagine when they think about how things might have turned out differently and how these thoughts may influence emotions and social judgements such as guilt and blame. This experiment examines how the order of events may influence these judgements. Thank you again for taking part in this experiment. If you have any comments to make, these are welcome.

Appendix 4: Materials used in the reasons for acting experiments

Experiment 4

Instructions (page 1)

Thank you for your participation in this experiment. On the following page you will find a scenario. Please take your time and read it carefully. The scenario is followed by some questions which ask for your views on it. There are no right or wrong answers. Please answer the questions in the order in which they are given, and do not change an answer once you have written it. You may refer back to the story as often as you wish.

On the final page, I have outlined the aim of the experiment. Please do not read this until you have completed all questions.

Before you turn this page, please record your gender and age in the space provided below:

Gender: Male _____ Female _____

Age: _____

Scenario (page 2) **Each of the three versions of the scenario contained one of the sentences in italics.**

Tom has been working in a bank in the city centre for the past few years. He lives with his wife on the outskirts of town. On Thursday, he left the office at the usual time. He was only a short distance down the road, when he drove over some glass and got a flat tyre. After changing it, he headed off again. Because of this delay however, he ran into the rush hour traffic which he usually managed to avoid. This meant that his journey was considerably slower than usual. Tom passes a local gym on his way home every day. It is normally closed but because of his delays that day, when he got there it had already opened for the evening.

[Control: It is quite a large gym.]

[Own-sake reason: He really likes to go to the gym.]

[Plan-based reason: He is trying hard to lose weight.]

He decided to stop off for a quick workout. Thirty minutes later he was finished and he set off again. When he arrived home soon after, he found his wife lying on the floor. He realised that she had had a heart attack and that she was dying. He tried to help her, but his efforts were in vain.

Questions (page 3)

Q.1. As commonly happens in such situations, Tom often thought, "if only . . . .", during the days and weeks that followed the accident. Please list in the order of importance, four ways in which he may have completed this thought so that the outcome would be different.

1.

2.

3.

4.

Q.2. On the following scale, please rate the extent to which you think Tom felt to blame for his wife's death?

1	2	3	4	5	6	7	8	9
Not at all to Blame								Completely to Blame

Q.3. On the following scale, please rate how much Tom regretted going to the gym.

1	2	3	4	5	6	7	8	9
No Regret								Strong Regret

Debriefing paragraph (page 4)

The study aims to examine what people imagine when they think about how things might have turned out differently and how these thoughts may influence may influence emotions such as blame.

If you have any comments to make, these are welcome. Thank you again for your participation.

Experiment 5

Instructions (page 1)

Thank you for your participation in this experiment. On the following pages you will find six scenarios. Please take your time and read them carefully. The scenarios are followed by some questions which ask for your views on them. There are no right or wrong answers. Please read the scenarios and answer the questions in the order in which they are given, and do not change an answer once you have written it. You may refer back to the stories as often as you wish.

On the final page, I have outlined the aim of the experiment. Please do not read this until you have completed all questions.

Before you turn this page, please record your gender and age in the space provided below:

Gender: Male _____ Female _____

Age: _____

Scenarios (pages 2 – 7) **The six versions of the scenario contained either an action or an event along with a reason, a cause or a filler (control), as presented in the sentences in italics.**

1. Tom has been working in a bank in the city centre for the past few years. He lives with his wife on the outskirts of town. On Thursday, he left the office a bit later than usual. The traffic was quite heavy so his journey was slow. On his way home, Tom normally drives down Church Street.

   [*Reason for Action:* However on this day, he wanted to avoid the heavy traffic that was in the inside lane.]
   [*Cause of Action:* However on this day, he automatically swerved the car to avoid a dog.]
   [*Control:* ..... which is a busy road. However, on this day.....]
   [*Action:* He pulled into the outside lane and as a result, he missed the turn.]

   [*Reason for Event:* However later that evening, there was going to be a bike race on Church Street. As a result.....]
   [*Cause of Event:* However on this day, a tree had fallen across Church Street and the traffic could not pass. As a result.....]
   [*Control:* ..... which is a busy road. However on this day, ...]
   [*Event:* The traffic was diverted onto a different route.]

   Tom turned onto Hill Street. As he approached a crossroads the lights changed to red and he stopped the car. As he began to cross after the lights changed, a truck charged through the lights at high speed and rammed Tom's car from the right. Tom was seriously injured.

2. Joe's son, David, attends a school in the suburbs of town. Every day at 2.30 p.m., Joe collects him from school. He normally leaves the house around 2 o'clock so that he arrives in time. However, on Thursday before he left .....
   [*Action:* he let the hamster out of its cage into the living room.]
   [*Reason for Action:* He wanted to give her some exercise.]
   [*Cause of Action:* He accidentally let her escape when he opened the door to feed her.]
   [*Control:* The hamster runs very fast and....]

   [*Event:* the dog escaped from the garden out onto the street.]
   [*Reason for Event:* The dog wanted to chase a cat she had seen.]
   [*Cause of Event:* There was a hole in the fence which allowed the dog to escape.]
   [*Control:* The dog runs very fast and...]

   It took longer than usual to get her back into the cage/ garden. As a result, he was late leaving the house. The traffic got stuck behind a learner driver so his journey was slower than usual. Because of the delay, Joe's neighbour who waited 15 minutes for Joe to show up, took the child home in his car. On the way home, the neighbour's car was struck by a drunk driver. Joe's son, the only seriously hurt victim, received severe leg injuries.

3. Eugene and Tina were a young married couple who lived in the country. Both were partially paralysed and confined to wheelchairs. They met four years before when Tina was a counsellor with the Irish Paraplegic Association, had fallen in love, and were married one year later. On this particular evening, Eugene had phoned to request a taxi to take them into town. However.....

   [*Action:*. the taxi-driver collected two other people who were waiting on the footpath.]

   [*Reason for Action:* One of them was injured and he wanted to take him home.]

   [*Cause of Action:* He collected them by mistake because they were waiting close to Eugene and Tina's house.]

   [*Control:* They were going on a short journey.]

   [*Event:* The taxi driver was delayed at a road block on the way to Eugene and Tina's house.]

   [*Reason for Event:* The police were checking tax discs.]

   [*Cause of Event:* There had been an accident on the road.]

   [*Control:* After a short while he was able to pass.]

   As a result, he didn't arrive at Eugene and Tina's house at the expected time. After waiting a while for the taxi, Eugene and Tina decided to take Tina's car which was equipped with special hand controls. In order to get into town from their house, they had to travel across Rupert river. A severe storm the night before had weakened the structure of the bridge. About 2 minutes before Eugene and Tina reached it, a section of the bridge collapsed. In the dark, Eugene and Tina drove off the bridge and plummeted into the river below. Both of them were badly injured.

4. Ned works as a safety inspector in a refrigerator manufacturing company. He is in his mid-60s and he has been working there for over twenty years. Ned's job involves carrying out inspections of the buildings on the plant. His rota is set by management at the beginning of each week. On Friday morning, he was supposed to inspect buildings 1 and 2, before going to the shipping department to oversee the loading of the trucks. However, he didn't check building 2 that morning as planned...

   [*Action:* Instead he went to inspect building 3]

   [*Reason for Action:* He wanted to check the wiring there.]

   [*Cause of Action:* He misread the rota and went there by mistake.]

   [*Control:* Building 3 is a large building.]

   [*Event:* because the building was closed.]

   [*Reason for Event:* This was to allow for repairs to the roof.]

   [*Cause of Event:* This was because of a water leakage.]

   [*Control:* Building 2 is a large building.]

   Instead, he decided to inspect Building 2 that afternoon. He finished his lunch a few minutes early and set off. Building 2 was located at the far side of the plant. As he walked past the loading dock on his way to building 2, a vehicle exploded, severely burning its driver and Ned.

5. Sam is an avid tennis player. Recently, he took part in an important tournament in France. On Sunday he was due to take part in the semi-finals. Normally, Sam does some training the day before an important match. However, on Saturday .....

   [*Action:* he decided to rest instead.]
   [*Reason for Action:* He wanted to be fresh for the following day.]
   [*Cause of Action:* After the long journey, he was unable to stay awake.]
   [*Control:* He went to his room.]

   [*Event:* the courts were closed.]
   [*Reason for Event:* This was to save the lawn for the upcoming matches.]
   [*Cause of Event:* Heavy rain the night before meant that the courts were water-logged.]
   [*Control:* He went to his room.]

   As a result, he didn't do any training that day. The day of the match was very hot. Although Sam had often played in hot conditions, he really didn't like it. That morning a draw was held to see which of the other three semi-finalists Sam would play. After the draw, Same made a choice of two rackets, one made of synthetic material and one made of metal. He finally decided to use the metal racket because it is most suited to his opponent's manner of play. That afternoon, Sam played the match and lost. He was very disappointed.

6. Susan is a second year science student in university. She obtained good marks in her first year exams. Susan has an important exam on Friday in her chemistry course. She spent about her usual amount of time studying her class notes and assigned readings.
   Susan planned to study on the Tuesday night before the exam, however
   [*Action:* she went to her neighbour's party instead.]
   [*Reason for Action*:: She wanted to meet some old friends from school.]
   [*Cause of Action:* She couldn't concentrate on her books with the loud music in the background.]
   [*Control:* She enjoyed the party very much.]

   [*Event:*. that evening the library closed early.]
   [*Reason for Event:* This was to allow for a reorganisation of the book shelving.]
   [*Cause of Event:* This was due to an expected shortage of library staff.]
   [*Control:* Once the library closed, she headed home.]
   As a result, she didn't study that night. On Thursday, Susan's mother was ill. As a result, Susan spent a lot of time on Thursday evening looking after her. On Friday morning, Susan was very nervous and she hoped that her anxiety would not interfere with her concentration. Susan sat the exam later that day and passed, but her marks were considerably lower than usual.

Questions (page 2 - 7) **This set of questions was presented after each scenario**

Q.1. As commonly happens in such situations, *the protagonist* often thought, "if only . . . .", during the days and weeks that followed. Please list in the order of importance, four ways in which *he/she* may have completed this thought so that the outcome would be different.

1.

2.

3.

4.

Q.2. On the following scale, please rate the extent to which you think *the protagonist* felt to blame for the outcome.

1	2	3	4	5	6	7	8	9
Not at all to Blame								Completely to Blame

Q.3. On the following scale, please rate the extent to which you think *the protagonist* felt regret following the outcome?

1	2	3	4	5	6	7	8	9
No Regret								Strong Regret

Debriefing paragraph (page 3)

The aim of this study is to imagine what people imagine when they think about how things might have happened differently and how these thoughts may influence emotions and social judgements such as blame and regret.

If you have any comments to make, these are welcome. Thank you again for your participation.

Experiment 6

Instructions (page 1)

Thank you for your participation in this experiment. On the following page you will find a scenario. Please take your time and read it carefully. The scenario is followed by some questions which ask for your views on it. There are no right or wrong answers. Please answer the questions in the order in which they are given, and do not change an answer once you have written it. You may refer back to the story as often as you wish.

On the final page, I have outlined the aim of the experiment. Please do not read this until you have completed all questions.

Before you turn this page, please record your gender and age in the space provided below:

Gender: Male _____ Female _____

Age: _____

Scenario (page 2) **Each of the three versions of the scenario contained one of the sentences in italics.**

Bernard is a famous violinist. At the moment, he is staying in Vienna where he has been invited to perform in an important open-air festival of classical music. In the course of this festival, Bernard intends to perform Beethoven's violin concerto. He knows the concerto well and his previous performances of it have been successful.

Bernard will be accompanied by the well-known Geneva Chamber Orchestra. Since the accompanying orchestra plays a prominent role in the soloists performance, Bernard and his agent had negotiations with several orchestras prior to the festival. The English Polyphonic Orchestra had been a real alternative, however, finally they decided on the Geneva Chamber Orchestra.

The evening before the concert, there is a ball organised by the festival committee.

[Obligation: The ball is being run to help cover the festival expenses and therefore Bernard is obliged to attend.]

[Goal-based reason: Bernard will have a unique opportunity to meet a world-renowned violinist at the ball and this is his goal for attending.]

[Control: The ball is being held in the Vienna City Ballroom.]

Bernard goes to the ball and as a result he has a late night, something he doesn't usually do before a concert.

The day of the concert Bernard thinks carefully about which violin he is going to use. He has the choice between a Stradivari and an Amati. He finally decides on the Amati because although he doesn't use it very often, he thinks that the timber of this violin is best suited to Beethoven's concerto.

At the concert that evening, Bernard's performance of the concerto was poor and the audience did not respond well. He was bitterly disappointed.

Questions (page3)

Q.1. Bernard often thought, "if only . . . ." in the days that followed the concert. Please list in the order of importance, four ways in which he may have completed this thought so that the outcome would be different.

1.

2.

3.

4.

Q.2. On the following scale, please rate the extent to which you think Bernard felt to blame for his poor performance?

1	2	3	4	5	6	7	8	9
Not at all to Blame								Completely to Blame

Q.3. On the following scale, please rate how much you think Bernard regretted going to the ball.

1	2	3	4	5	6	7	8	9
No Regret								Strong Regret

Q.4. On the following scale, please rate the extent to which you think Bernard had a choice in going to the ball.

1	2	3	4	5	6	7	8	9
No Choice								Completely Free Choice

Q.5. On the following scale, please rate the extent to which you think Bernard felt compelled to go the ball.

1	2	3	4	5	6	7	8	9
Not at all Compelled								Strongly Compelled

Debriefing paragraph (page 4)

The study aims to examine what people imagine when they think about how things might have turned out differently and how these thoughts may influence emotions such as blame. This study looks at the way that people think about their own decisions and is designed to compare actions which they are obliged to do with actions which they have a purpose for doing.

If you have any comments to make, these are welcome. Thank you again for your participation.

Appendix 5: Materials used in the mental representations experiments

Experiment 7

Instructions

Thank you for agreeing to take part in this experiment.

This program is going to present pairs of sentences on the screen.
After each pair of sentences it will ask you whether you can draw a conclusion from them.
Please read the sentences carefully and type your answer on the screen.
When you have finished typing your response, please press the return key.

Before we start the experiment you will be presented with four practice questions.
Please take your time and make sure that you understand these questions before starting the experiment.
If you have any questions please ask the experimenter.

Press any key followed by return to continue.

Practice problems

If the box is blue then it is empty.
The box is blue.

Kim owns a car or Kim owns a bike or both.
Kim owns a car.

Tom has an apple or Jim has an orange but not both.
Tom has an apple.

If and only if there is a circle on the board then there is a triangle on the board.
There is not a circle on the board.

Anne is wearing a white jumper or Jo is wearing a black skirt but not both.
Jo is wearing a black skirt.

Experimental problems

Inclusive disjunction: fillers
Patrick is tidying the desk or Kevin sitting on the chair or both.
Patrick is tidying the desk.

Robert is making the bed or Lucy is searching in the cupboard or both.
Lucy is searching in the cupboard.

Walter is emptying the dishwasher or Walter is listening to the stereo or both.
Walter is emptying the dishwasher.

Marie is looking at the fishtank or Charlie is looking at the fishtank or both.
Charlie is looking at the fishtank.

Biconditional: two-persons, two-action
If and only if Brian is not standing by the fireplace then Joanne is looking in the mirror.
Joanne is not looking in the mirror.

If and only if Rachel is not climbing up the stairs then David is cooking at the stove.
David is cooking at the stove.

If and only if Alan is not looking under the bed then Cathy is washing at the sink. Alan is not looking under the bed.

If and only if Karen is not sitting in front of the TV then Louise is sleeping in the bed.
Karen is sitting in front of the TV.

Biconditional: one person, two-actions
If and only if Rita is not looking into the wardrobe then Rita is sweeping under the table.
Rita is not sweeping under the table.

If and only if Mary is not reading beside the lamp then Mary is swimming in the pool.
Mary is swimming in the pool.

If and only if Barry is not working at the bench then Barry is kneeling on the rug. Barry is not working at the bench.

If and only if Sarah is not sitting in the arm-chair then Sarah is opening the front door.
Sarah is sitting in the arm-chair.

Biconditional: two persons, one-action (non-simultaneous)
If and only if Graham is not standing on the scales then Carol is standing on the scales.
Carol is not standing on the scales.

If and only if Michael is not writing at the desk then Alex is writing at the desk.
Alex is writing at the desk.

If and only if Eric is not sitting in the bath then Daniel is sitting in the bath.
Eric is not sitting in the bath.

If and only if Martin is not standing on the stool then Susan is standing on the stool.
Martin is standing on the stool.

Biconditional: two persons, one-action (simultaneous)
If and only if Ruth is not eating at the table then Alice is eating at the table.
Alice is not eating at the table.

If and only if Linda is not looking out the window then Richard is looking out the window.
Richard is looking out the window.

If and only if Peter is not sitting on the sofa then Andrew is sitting on the sofa.
Peter is not sitting on the sofa.

If and only if Mark is not leaning on the counter then Lisa is leaning on the counter.
Mark is leaning on the counter.

Exclusive disjunction: two-persons, two-action
Brian is standing by the fireplace or Joanne is looking in the mirror but not both.
Brian is standing by the fireplace.

Rachel is climbing up the stairs or David is cooking at the stove but not both.
Rachel is not climbing up the stairs.

Alan is looking under the bed or Cathy is washing at the sink but not both.
Cathy is washing at the sink.

Karen is sitting in front of the TV or Louise is sleeping in the bed but not both.
Louise is not sleeping in the bed.

Exclusive disjunction: one person, two-actions
Rita is looking into the wardrobe or Rita is sweeping under the table but not both.
Rita is looking into the wardrobe.

Mary is reading beside the lamp or Mary is swimming in the pool but not both.) Mary is not reading beside the lamp.

Barry is working at the bench or Barry is kneeling on the rug but not both.) Barry is kneeling on the rug.

Sarah is sitting in the arm-chair or Sarah is opening the front door but not both.
Sarah is not opening the front door.

Exclusive disjunction: two persons, one-action (non-simultaneous)
Graham is standing on the scales or Carol is standing on the scales but not both.
Graham is standing on the scales.

Michael is writing at the desk or Alex is writing at the desk but not both.
Michael is not writing at the desk.

Eric is sitting in the bath or Daniel is sitting in the bath but not both.
Daniel is sitting in the bath.

Martin is standing on the stool or Susan is standing on the stool but not both.
Susan is not standing on the stool.

Exclusive disjunction: two persons, one-action (simultaneous)
Ruth is eating at the table or Alice is eating at the table but not both.
Ruth is eating at the table.

Linda is looking out the window or Richard is looking out the window but not both.
Linda is not looking out the window.

Peter is sitting on the sofa or Andrew is sitting on the sofa but not both. Andrew is sitting on the sofa.

Mark is leaning on the counter or Lisa is leaning on the counter but not both.
Lisa is not leaning on the counter.

Experiment 8: One version of the problems used in Experiment 8

Instructions

Thank you for taking part in this experiment.

During the experiment, you will be presented with sentences on the screen. When you have read each sentence you should press the spacebar to continue. When you have finished reading two sentences, you will be presented with a question. Please read the sentences and the question carefully. You should answer the question by pressing the 'Yes', 'No' or 'Cannot tell' keys on the keyboard.

Before you start the experiment, you will be presented with 5 practice questions. Please take our time and make sure you understand these questions by starting the experiment. If you have any questions, please ask the experimenter.

Press the space bar to continue.

Practice problems

Either Barbara is washing the dishes or she is sweeping the floor.
Barbara is washing the dishes.
Is she sweeping the floor?

All of the teachers are athletes.
David is a teacher.
Is David an athlete?

Mary is talking to Barry or she is talking to Susan or both.
Mary is talking to Barry.
Is she talking to Susan?

Either Peter is playing football or otherwise Andrew is playing basketball.
Peter is not playing football.
Is Andrew playing basketball?

Some of the nurses went to see the film.
Deirdre is a nurse.
Did she go to see the film?

Experimental problems (one version)

4 persons condition

Either Lynn is sitting on the stool and Chris is reading beside the lamp or otherwise Karl is walking towards the bookshelf and Bob is looking at the picture.
Lynn is sitting on the stool.
Is Chris reading beside the lamp?

Either Neil is sitting at the table and Max is sewing on the machine or otherwise June is walking around the bed and Ben is gathering up the toys.
Neil is sitting at the table.
Is June walking around the bed?

Either Ken is standing by the window and Jane is painting on the canvas or otherwise Tom is lying on the sofa and Niamh is looking through the paper.
Ken is not standing by the window.
Is Jane painting on the canvas?

Either Kate is sitting beside the clock and Tim is chatting on the phone or otherwise Sam is running on the treadmill and Maeve is listening to the walkman.
Kate is not sitting beside the clock.
Is Sam running on the treadmill?

2 persons, 1 per model

Either Rose sitting on the sofa and she is blowing out the candles or otherwise Jim is standing beside the piano and he is drinking through the straw.
Rose is sitting on the sofa.
Is she blowing out the candles?

Either Tim is kneeling on the carpet and he is winding up the toy or otherwise Greg is running up the stairs and he is holding onto the bags.
Tim is kneeling on the carpet.
Is Greg running up the stairs?

Either Liz is sitting on the chair and she is playing on the piano or otherwise Bob is standing by the fireplace and he is juggling with the balls.
Liz is not sitting on the chair.
Is she playing on the piano?

Either Chris is leaning on the table and he is writing with the pencil or otherwise Ben is coming through the door and he is pushing along the pram.
Chris is not leaning on the table.
Is Ben coming through the door?

2 persons, 2 per model, (non-simultaneous)
Either Steve is walking past the door and Sean is looking at the notice or otherwise Steve is leaning against the wall and Sean is writing with the marker.
Steve is walking past the door.
Is Sean looking at the notice?

Either Anne is leaning on the counter and Paul is writing in the notebook or otherwise Anne is walking around the table and Paul is searching for the paper.
Anne is leaning against the counter.
Is she walking around the table?

Either Ruth is climbing up the ladder and Kay is looking in the window or otherwise Ruth is resting on the sofa and Kay is glancing through the magazine.
Ruth is not climbing up the ladder.
Is Kay looking in the window

Either John is lying on the bed and Pam is talking on the phone or otherwise John is relaxing in the bath and Pam is listening to the radio.
John is not lying on the bed.
Is he relaxing in the bath?

2 persons, 2 per model, (simultaneous)
Either Bill is kneeling on the floor and Max is standing on the stool or otherwise Bill is hammering in the nail and Max is screwing in the bulb.
Bill is kneeling on the floor.
Is Max standing on the stool?

Either Niamh is leaning out the window and Pat is walking around the table or otherwise Niamh is lifting up the flower-pots and Pat is tidying up the dishes.
Niamh is leaning out the window.
Is she lifting the flower-pots?

Either Jack is relaxing in the armchair and Sue is walking up the stairs or otherwise Jack is counting up the coins and Sue is talking on the phone.
Jack is not relaxing in the armchair.
Is Sue walking up the stairs?

Either Maeve is standing at the doorway and Kim is kneeling beside the box or otherwise Maeve is drinking from the mug and Kim is blowing up the balloon.
Maeve is not standing in the doorway.
Is she drinking from the mug?

1 person

Either Jo is sitting in the chair and she is typing on the keyboard or otherwise she is crawling under the table and she is looking for the ring.
Jo is sitting in the chair.
Is she typing on the keyboard?

Either Dave is sitting at the table and he is sorting through the mail or otherwise he is kneeling on the rug and he is playing with the cards.
Dave is sitting at the table.
Is he kneeling on the rug?

Either Jill is climbing up the stairs and she is holding onto the box or otherwise she is standing in the corner and she is washing at the sink.
Jill is not climbing up the stairs.
Is she standing in the corner?

Either Jane is kneeling by the fire and she is looking at the TV or otherwise she is standing at the window and she is peering into the garden.
Jane is not kneeling by the fire.
Is she looking at the TV?