# THE RELATIONSHIP BETWEEN FORMAL AND INFORMAL REASONING

A Thesis Submitted to the College of

Graduate Studies and Research

in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy

in the Department of Psychology

University of Saskatchewan

Saskatoon

by

Eyvind Ohm

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Psychology
Arts Building
University of Saskatchewan
9 Campus Drive
Saskatoon, Saskatchewan
S7K 5A5
Canada

Abstract

In traditional tasks of formal reasoning, participants are asked to evaluate the validity of logical arguments. While this research tradition has contributed in many ways to our understanding of human reasoning, the extent to which this body of research generalizes to everyday, or informal, reasoning is unclear (e.g., Evans & Thompson, 2004; Galotti, 1989). The main goal of this dissertation was to illustrate the benefits of applying an informal approach to the study of conditional reasoning. In six experiments, everyday conditionals in the form of inducements (promises and threats) and advice (tips and warnings) were investigated. The results support three main conclusions. First, people recruit a substantial amount of background knowledge when interpreting and reasoning with these conditionals. Specifically, inducements were found to be different from advice on several pragmatic variables (Experiment 1); these variables also predicted differences in inference patterns (Experiment 2). Second, these studies provide further support for a probabilistic interpretation of conditionals (e.g., Evans & Over, 2004; Oaksford & Chater, 2001). Thus, in Experiments 3-5, estimates of different conditional probabilities predicted a number of judgments people make about inducements and advice. A particularly interesting finding was that the effectiveness of these conditionals in changing behaviour did not seem to depend on how likely they were perceived to be true. Finally, Experiment 6 adopted a decision-theoretic analysis (e.g., Over, Manktelow, & Hadjichristidis, 2004), showing that the effectiveness and quality of inducements and advice were tied to perceptions of subjective utility and preferences among possible outcomes. This dissertation concludes with a theoretical discussion of the nature of the relationship between formal and informal reasoning.

Acknowledgements

First and foremost, I wish to thank my supervisor, Dr. Valerie Thompson, for her support, patience and friendship. It would be impossible to overestimate her contribution to my progress over the last 4 years and to the work reported in this dissertation. I would also like to express my gratitude to Dr. Jamie Campbell for his thoughtful comments on my work and constant reassurances. In addition, I thank Drs. Jim Cheesman, Scott Bell, and Ron Borowsky for their contributions at various stages of my stay here, as well as my external examiner, Dr. Ken Manktelow.

Thanks should also be given to former and present members of the Cognitive Science Lab at the University of Saskatchewan. In particular, I wish to thank Raymond Gunter, Kealee Playford, and Jody Shynkaruk for their technical and practical assistance. For all the good times, I must also acknowledge Dr. Tom Phenix, David Lane, and, more recently, Dr. Orlando Espino.

Finally, I would like to thank my parents, Anders and Eva, for their continued support and understanding.

Table of Contents

List of Tables

*Experiment 5*



*Experiment 6*



*General Discussion*

List of Figures

List of Appendices

# THE RELATIONSHIP BETWEEN FORMAL AND INFORMAL REASONING

The psychology of deductive reasoning has been a major paradigm in cognitive psychology over the last 40 years or so (for a historical review, see Evans, 2002). Research in this paradigm has contributed significantly to our understanding of human reasoning, by documenting a multitude of empirical findings and generating theoretical debates about a number of profound issues (e.g., Evans, 2005a; Evans, Newstead, & Byrne, 1993; Johnson-Laird, 1999; Manktelow, 1999; Stanovich, 1999). However, this research endeavour has mainly investigated *formal* reasoning, a rather restrictive type of reasoning that involves drawing necessary conclusions from premises that can be assumed true. Whether, or to what extent, this body of research can inform us about the reasoning people typically engage in during their everyday lives is unclear (e.g., Evans & Thompson, 2004; Galotti, 1989). In everyday or *informal* reasoning, people instead draw inferences from uncertain premises and with varying degrees of confidence (e.g., Over & Evans, 2003). In this introductory chapter, I discuss the relationship between formal and informal reasoning, in the process arguing that the dominant deductive paradigm is in fact ill-equipped to explore informal reasoning. Instead, to capture the complexities of everyday reasoning, researchers will need to move beyond the restrictions of the deductive paradigm and make use of a more informal approach. Subsequently, this dissertation presents a series of experiments, which aims to illustrate the benefits of applying precisely such an approach.

## *Contrasting formal and informal reasoning*

Within psychology, formal reasoning can be classified into two general domains: deductive reasoning and statistical inference (Evans & Thompson, 2004). Research in the

latter domain explores how people judge the probability or frequency of uncertain events (for comprehensive reviews of this research tradition, see Gilovich, Griffin, & Kahneman, 2002; Koehler & Harvey, 2004). For the purpose of this dissertation, however, the discussion of formal reasoning will be restricted to the domain of deductive reasoning.[1] With this restriction in mind, I define formal reasoning broadly as the processes involved in the evaluation of logical arguments. Historically, the study of formal reasoning grew out of a tradition of *logicism*, a doctrine where logic was believed to provide the basis for rational human thought (Evans, 2002). Given the assumed importance of logic to human reasoning, the initial research agenda in the field was geared towards assessing people's logical competence on deductive reasoning problems.

In such deductive problems, participants are given premises and asked to determine the validity of conclusions (or, less frequently, to generate valid conclusions). As an example, consider a syllogism with these two premises: "Some A are B" and "No B are C". The participant may be asked whether we can conclude from these premises that "Some A are not C". In the instructions to these tasks, participants are told to assume the truth of the premises and to endorse only conclusions that *necessarily* follow from the premises. In other words, a conclusion that is (even highly) probable, but not necessary, is invalid according to the principle of logical necessity. Reasoning performance is, furthermore, assessed by the application of a normative standard that classifies responses as either correct or fallacious. The appropriate normative system for deductive reasoning

[1] Deductive reasoning is often contrasted to inductive reasoning, which involves making generalized conclusions from particular observations or instances. The study of inductive reasoning covers many topics (Sloman & Lagnado, 2005), such as categorization, hypothesis testing, analogical thinking, and cue learning; indeed, the domain of statistical inference is a form of inductive reasoning. In addition to this diversity, no over-arching normative theory is applicable to inductive reasoning (Manktelow, 1999); for these reasons, I do not view the general area of induction as a domain of formal reasoning. The issue of inductive reasoning is, however, considered again in the General Discussion.

has traditionally been provided by formal logic (i.e., the propositional and predicate calculi). The use of logic guarantees valid inferences, that is, the endorsement of conclusions that must be true given the truth of the premises.

A crucial feature of these problems is that the premises provide all the information that the participant should consider when reasoning (Galotti, 1989). In other words, these problems are well-defined, as no additional information or background knowledge need be recruited to arrive at the correct answer (Evans & Thompson, 2004). Indeed, the correct solution to these problems often requires participants to avoid adding such extraneous information. However, this requirement is typically not specified in the task instructions, and the tendency to recruit background knowledge and prior belief is thought to be a common source of error in deductive reasoning (e.g., Henle, 1962; Stanovich, 1999).

In contrast, most informal reasoning problems are ill-defined, and their solutions have little, if anything, to do with logic. Informal reasoning can be broadly defined as the types of reasoning people engage in during their everyday lives, and I will in this dissertation use the terms informal reasoning and everyday reasoning interchangeably. Unlike formal reasoning, all information relevant to the evaluation of informal arguments is unlikely to be explicitly provided. Rather, successful everyday reasoning requires people to go beyond what they are told, by searching memory or consulting outside resources to find relevant information (Galotti, 1989). Furthermore, although informal arguments may contain premises and conclusions, these components are typically not clearly defined (Shaw, 1996). Indeed, some premises may not be stated explicitly, but

must somehow be inferred by the reasoner. For reasons such as these, informal arguments will rarely be deductively valid.

Informal reasoning also differs from formal reasoning in other important respects. For example, few premises in everyday life can simply be assumed true, a requirement in traditional deductive reasoning tasks. Instead, people reason on the basis of beliefs they cannot be completely certain about and from statements they cannot completely trust (Over & Evans, 2003). Furthermore, rather than reasoning on the basis of logical necessity (where conclusions either follow or do not), people express *degrees* of confidence in the inferences they make. Everyday reasoning is also considered to be defeasible or non-monotonic, such that a conclusion can be withdrawn in the light of new, potentially contradictory, evidence (e.g., Evans, 2005a; Oaksford & Chater, 2001). A final attribute of informal arguments is that they are usually uttered to achieve some goal of personal relevance (e.g., Sperber & Wilson, 1986) to the speaker. For instance, many informal arguments are uttered to convince another person of the utility of some opinion or to persuade a person of the benefits of undertaking some action. Later in this introduction, I will discuss research that has a direct bearing on these aspects of informal reasoning. However, first I will give a brief assessment of the formal deductive paradigm (for a comprehensive review, see Evans, 2002).

*Formal reasoning research: Contributions and limitations*

Although an in-depth analysis of formal reasoning research is beyond the scope of this introductory chapter, I will make some general comments about the contribution of this research to our understanding of human reasoning. As previously mentioned, the original research agenda in the psychology of reasoning was directed towards assessing

the extent to which people could reason logically, that is, consistent with the doctrines of formal logic. On the basis of many studies using abstract problems (i.e., containing artificial materials devoid of meaning), it can be concluded that performance of adults untutored in logic is often quite poor (e.g., Evans et al., 1993). On the one hand, performance is generally above chance levels, providing evidence for an "irreducible minimal deductive competence" (Evans, 2002, p. 982). On the other hand, errors in logical reasoning are widespread. These errors or fallacies involve both drawing invalid inferences (e.g., accepting conclusions that could, but need not, be true given the premises) and failing to endorse valid inferences.

To illustrate, consider the case of conditional reasoning, which involves drawing inferences from statements of the form "if *p*, then *q*". In the conditional inference task, four different arguments can be constructed by pairing a conditional statement with a categorical premise corresponding to either the presence or absence of the antecedent *p*, or the presence or absence of the consequent *q* (for a review of conditional reasoning, see Evans et al., 1993; Manktelow, 1999). While participants almost universally accept the valid Modus Ponens inference (i.e., "*p*, therefore *q*"), many participants also conclude that *p* follows from the affirmation of the consequent (i.e., "*q*, therefore *p*"), an inference that is regarded as invalid in standard logic (note, however, that this inference *is* valid under a biconditional interpretation, which reads "if and only if *p*, then *q*"). Likewise, participants often fail to endorse the valid Modus Tollens inference, whereby the absence of *p* can be inferred from the absence of *q* (i.e., "~*q*, therefore ~*p*").

The logical errors to which participants are prone tend furthermore to be systematic (i.e., non-random) in nature, a finding that has fueled an interest into the study

of reasoning biases (e.g., Evans, 1989). For instance, a major bias in the conditional reasoning literature involves *negative conclusion bias*, which refers to a greater willingness to accept inferences leading to negative, as compared to positive, conclusions (e.g., Evans et al., 1993). In the case of Modus Tollens, fewer participants accept the conclusion when a negation is inserted into the antecedent of the conditional (i.e., "if not-*p*, then *q*"), which has the result of making the conclusion positive (i.e., "not *q*, therefore not not-*p*, thus *p*"). This finding is interpreted as a bias because performance is influenced by a feature that is irrelevant to the logic of the task (i.e., the valence of the conclusion).

In addition to these empirical observations, research in formal reasoning has generated theoretical debates about a number of profound issues. At the forefront of these issues is the question of how to account for people's deductive competence. The two leading theoretical approaches addressing this issue are the mental rules/logic theory (e.g., Braine, 1978; Braine & O'Brien, 1991, O'Brien, 2004) and the mental model theory (e.g., Johnson-Laird, 2001; Johnson-Laird & Byrne, 1991, 2002). The former approach characterizes human reasoning as the application of a set of inference rules stored in a mental logic, or natural deduction system; difficulty in reasoning is determined by the number of inferential steps involved (e.g., there is a direct rule for the Modus Ponens inference, while Modus Tollens is solved indirectly with the use of Reductio ad Absurdum). The latter approach, in contrast, proposes that people reason by manipulating mental models constructed on the basis of the premises (a mental model is a representation of a possible state of affairs in which the premises are true). According to this theory, reasoning difficulty is determined by the number of mental models that need to be represented or "fleshed out".

While the mental logic and mental model theories have (often vehemently) contested the nature of deductive competence, other theories in the field have emerged that instead attempt to explain reasoning performance by appealing to a variety of non-logical processes. For example, some theories posit that reasoning is mediated by pragmatic or linguistic factors (e.g., Cheng & Holyoak, 1985; Cosmides, 1989; Fiddick, Cosmides, & Tooby, 2000; Polk & Newell, 1995; Sperber, Cara, & Girotto, 1995). Other theories go even further by claiming that reasoning performance should be interpreted within a probabilistic or decision-theoretic, rather than deductive, framework (e.g., Chater & Oaksford, 1999; Evans & Over, 2004; Manktelow & Over, 1991, 1995; Oaksford & Chater, 2001, 2002, 2003).

Another important theoretical issue generated by the study of formal reasoning relates to the implications this research may have for the notion of human rationality. That is, if logic provides the basis for rational human thought, then evidence showing that people often do rather poorly on these logical problems would seem to cast doubt on the assumption that humans are rational. A variety of responses have been made in an attempt to resolve this issue (e.g., Chase, Hertwig, & Gigerenzer, 1998; Chater & Oaksford, 2000; Cohen, 1981; Evans, 2002; Evans & Over, 1996, 1997; Gigerenzer & Selten, 2001; Henle, 1962; Hilton, 1995; Oaksford & Chater, 2001; Stanovich, 1999; Stanovich & West, 1998, 1999, 2000). Common to most of these explanations is the view that errors in deductive tasks do not reflect irrationality, but are instead due to factors extrinsic to logical competence. For example, normative violations have been attributed to incorrect norm application on the part of researchers, alternative task construal on the

part of participants, various performance factors (e.g., memory and other processing limitations), and the dominant influence of pragmatic and interpretive processes.

Thus, it seems fair to conclude that the study of formal reasoning, originally geared towards assessing people's logical competence, has made both empirical and theoretical contributions to our understanding of human reasoning. However, commitment to this original research agenda has diminished sharply over the years, and has been replaced by an interest in how reasoning is affected by the use of realistic (as opposed to abstract) problem materials (Evans, 2002). This body of research has demonstrated the pervasive influence of pragmatic factors, and, in particular, the importance of prior knowledge and belief on reasoning performance. For example, when the prior believability of conclusions is manipulated, a robust belief bias effect is found in the syllogistic reasoning literature (e.g., Evans, Barston, & Pollard, 1983). Specifically, participants endorse more arguments whose conclusions are consistent with their beliefs than arguments with unbelievable conclusions, irrespective of their logical validity (although this effect of believability is stronger for invalid arguments).

The deduction paradigm classifies such pragmatic factors as non-logical, as they are irrelevant to the formal task of determining logical validity. These factors are, however, clearly relevant to the evaluation of informal arguments. The belief bias effect, for instance, could be seen as consistent with the (arguably reasonable) tendency for people to evaluate the plausibility or believability of everyday utterances, such that participants should be reluctant to accept conclusions that are in conflict with their currently held beliefs, even when such conclusions are strictly valid (e.g., Evans, Over, & Manktelow, 1993). Indeed, the generalization of informal reasoning strategies to formal

tasks is often used as an explanation for the logical errors people make in the laboratory (e.g., Evans & Thompson, 2004; Oaksford & Chater, 2001). By documenting the pervasive influence of pragmatic factors, therefore, one could argue that the study of formal reasoning has the potential to uncover important insights into the nature of informal reasoning. However, due to the fundamental differences between the two types of reasoning, I argue that this potential is, in fact, rather limited.

Specifically, although formal reasoning research has demonstrated the importance of pragmatic factors in human reasoning, the methodology of this approach is in truth ill-equipped to explore such factors (e.g., Evans & Thompson, 2004). For example, by instructing participants to assume the truth of the premises, they are in effect asked to abandon their customary strategy of assigning degrees of certainty to utterances. Similarly, the task of determining whether inferences necessarily follow does not reflect the tendency for people to express degrees of belief or confidence in their conclusions. For reasons such as these, it is unlikely that a formal approach using traditional instructions will succeed in capturing the essential properties of informal reasoning.

Furthermore, much of the evidence illustrating the pragmatic nature of human reasoning has been obtained in a rather indirect manner. That is, the instructions of the standard methodology serve to suppress pragmatic factors, and their influence is typically identified by the observation of errors in logical reasoning (Evans & Thompson, 2004). If what we really are interested in is the influence of pragmatic factors, it would seem to make more sense to give participants tasks that permit a more direct exploration of such factors, and that do not afford logic such a crucial role. Moreover, studies should cease to

use instructions that traditionally define the influence of prior belief, knowledge, and other pragmatic factors as errors or biases (Evans, 2002).

The aforementioned discrepancies between formal and informal reasoning should, therefore, give rise to a considerable degree of concern to researchers in the field. Specifically, most of the reasoning people do in life, including such activities as planning, evaluating arguments, and choosing between alternative options, is informal in nature (e.g., Galotti, 1989; Perkins, 1985). Despite the primacy of informal reasoning, most reasoning studies in cognitive psychology explore performance on a small number of formal tasks that are designed to examine a particular and restrictive type of inferential process (i.e., that of deriving or evaluating a conclusion from simple premises); furthermore, there are strong reasons to doubt that the findings from these studies will generalize to informal settings. In addition, while the theoretical focus of the field has undergone substantial change over the years (i.e., from assessing logical competence to explaining reasoning performance), the methodology used in reasoning studies has not witnessed a corresponding change (e.g., Evans, 2002; Evans & Thompson, 2004).

To address these concerns, and also to improve the external validity of reasoning research, I suggest that a move away from the formal approach is needed. In the following sections, I review research that has, to varying degrees, liberated itself from the restrictions of the formal approach, and in so doing, contributed to our understanding of informal or everyday reasoning. Note that this review will mainly focus on research conducted in the area of conditional reasoning. I first discuss research that, although largely formal in nature, is relevant to the uncertain and defeasible aspects of informal reasoning. Next, insights from several studies that introduce novel tasks and ask non-

deductive questions are reviewed. Finally, I examine the field of argumentation, an area that has attracted little interest from most reasoning researchers but aptly illustrates the benefits of an informal approach to the study of human reasoning.

*The uncertainty of informal reasoning*

In tasks of formal reasoning, participants are instructed to draw necessary conclusions from premises that can be assumed true. As previously noted, these instructions are inconsistent with the uncertain nature of informal reasoning. That is, people in everyday life reason on the basis of uncertain premises and express degrees of uncertainty in the inferences they make (e.g., Over & Evans, 2003). In this section, I discuss research that has a bearing on the uncertain aspect of informal reasoning. While the two elements of uncertainty discussed here (i.e., reasoning from uncertain premises vs. expressing uncertainty in the conclusion) are clearly separate, I will group these two issues together, as many of the relevant studies speak to both.

Specifically, I review research that, either directly or indirectly, has introduced some degree of uncertainty into the premises. While this research has made use of traditional deductive (i.e., formal) tasks, some of these studies have given participants non-standard (i.e., more pragmatic) instructions. Thus, participants may not be told to assume the truth of the premises, and rather than basing their inferences on the principle of logical necessity, they may simply be asked to evaluate the conclusions by whatever criteria they feel are relevant. Furthermore, while some studies provide only two, or at most three, response options (e.g., yes, no, maybe), others allow participants a much better opportunity to express uncertainty in their conclusions, usually by providing a rating scale (e.g., Cummins, 1995; Cummins, Lubart, Alksnis, & Rist, 1991; De Neys,

Schaeken, & d'Ydewalle, 2003; George, 1995, 1997; Liu, Lo, & Wu, 1996; Verbrugge, Dieussart, Schaeken, & van Belle, 2004).

I first discuss some studies that have indirectly introduced uncertainty into the premises. They have done so by using content that, because of participants' background knowledge, conveys varying degrees of uncertainty. To illustrate, a number of studies in the conditional reasoning literature have shown that the perceived strength or certainty of the relationship between the antecedent and consequent terms can affect reasoning performance (e.g., Cummins, 1995; Cummins et al., 1991; De Neys et al., 2003; George, 1995; Markovits, 1984; Liu et al., 1996; Politzer & Bourmaud, 2002; Quinn & Markovits, 1998; Thompson, 1994, 1995, 1996).

For example, Cummins et al. (1991) investigated reasoning with causal statements (i.e., if cause *p*, then effect *q*). These statements differed in terms of the number of 1) possible causes, other than *p*, that could produce the effect (i.e., alternative causes) and 2) possible events that could prevent the effect from occurring despite the presence of the cause (i.e., disabling conditions). In this study, acceptance ratings to conclusions were found to be lower for conditionals with many alternative causes and disabling conditions. One interpretation of this finding is that the presence of such factors reduces the perceived certainty of the link between the cause and the effect. Thus, in the statement "If I eat candy often, then I have cavities", there are other reasons besides eating candy that may result in cavities (e.g., drinking soft drinks), and also several ways of preventing cavities even if candy is eaten (e.g., brushing one's teeth). More precisely, the presence of alternative causes and disabling conditions casts doubt on the perceived necessity and sufficiency of the cause in bringing about the effect; in turn, perceptions of necessity and

sufficiency serve to modify the degree of belief in the conclusions (e.g., Cummins, 1995; Politzer & Bourmaud, 2002; Thompson, 1994, 1995).

Another indirect way of introducing uncertainty into the premises involves giving participants statements uttered by speakers in a conversational context. Examining a variety of such speech acts, Newstead, Ellis, Evans, and Dennis (1997), for example, observed consistent differences in reasoning performance between conditionals with different content. A particularly interesting finding was that participants accepted more inferences for conditionals uttered as inducements (i.e., promises and threats) than for conditional advice (i.e., tips and warnings). In trying to explain this difference, Newstead et al. argued that inducements express a more certain link between the antecedent and consequent than advice does. For example, the occurrence of the consequent will more likely result from the satisfaction of the antecedent in a promise (e.g., "If you wash the car, I will give you $10") than in a tip (e.g., "If you show up early for work, you will impress the boss").

Uncertainty can also be introduced in speech acts by manipulating various attributes of the speaker. Indeed, Evans and Twyman-Musgrove (1998) argued that an important reason for the greater certainty associated with conditional inducements, as compared to conditional advice, is that speakers of the former, but not the latter type, usually possess considerable control over the outcome of the consequent. Consistent with this hypothesis, these researchers found that acceptance rates for inducements and advice were consistently lower when the degree of speaker's control was reduced. A similar result was reported by Stevenson and Over (2001), who varied the authority of the speaker of conditional statements such as "If Bill has typhoid, he will make a good

recovery". In this study, conclusions were rated as less certain when the statement was uttered by a novice (e.g., a medical student) than by an expert (e.g., a professor of medicine).

In contrast to these indirect approaches, a few other studies have introduced uncertainty in a more direct manner. Specifically, these studies have modified premises by the use of explicit qualifiers, such as "probably" or "sometimes" (e.g., George, 1997, 1999; Stevenson & Over, 1995). For instance, George (1997) manipulated the certainty of conditional premises by qualifying the consequent term, as in "If Pierre is in the kitchen, then it is very probable that Marie is in the garden". Two main findings have emerged from this type of manipulation (see George, 1999). First, when evaluating arguments based on such uncertain premises, very few participants are willing to attribute absolute certainty to the conclusions, but instead choose to express some degree of uncertainty. Furthermore, the plausibility of conclusions is often linked to the uncertainty asserted in the premises, such that the degree of uncertainty in the premises is fairly directly transmitted, or scaled, to the conclusion.

In general, these studies exploring uncertain reasoning have contributed to our understanding of informal reasoning by demonstrating that people readily take into account the plausibility of utterances presented to them. That is, people are sensitive to the degree of uncertainty conveyed, either directly or indirectly, by the premises, and readily use this degree of uncertainty when making inferences. At the very least, the results of these studies suggest that to understand everyday reasoning, researchers should dispense with deductive instructions where participants are asked to assume the truth of the premises and to reason on the basis of what necessarily follows.

*The defeasibility of informal reasoning*

Another aspect of informal reasoning that is poorly captured by the traditional deductive paradigm involves the defeasible nature of everyday inferences. That is, the inferences people make in the real world are considered to be provisional, and can be strengthened, weakened, or even withdrawn in the light of new evidence (e.g., Evans, 2002; Oaksford & Chater, 2001). In formal deductive tasks, participants are given a fixed number of premises (usually two), and on this basis, asked to generate or evaluate a conclusion; if a conclusion necessarily follows from the premises, then no subsequent information can invalidate it (Johnson-Laird, 1999). These tasks are, therefore, unable to explore how people accommodate additional information that may be relevant to the conclusion. In particular, these tasks do not speak to the issue of how people resolve the conflicts arising when new information contradicts a conclusion already endorsed. In this section, I discuss research that has a bearing on the defeasibility of everyday reasoning. As the formal tasks in their original form are unable to address this issue, the research described here necessarily introduces various modifications of the traditional deductive paradigm.

In standard tasks of conditional reasoning, participants are given two premises (i.e., the conditional and categorical premises). One approach that could be regarded as having a bearing on the defeasible nature of reasoning modifies these tasks by the presentation of one or several premises in addition to the standard pair (Evans, 2002). The use of such additional premises has been found to decrease endorsement rates of conditional inferences (e.g., Bonnefon & Hilton, 2004; Byrne, 1989; Byrne, Espino, & Santamaria, 1999; Chan & Chua, 1994; Hilton, Jaspars, & Clarke, 1990; Rumain,

Connell, & Braine, 1983) or, in studies that allow participants to express degrees of uncertainty in their answers, to reduce the certainty of these conclusions (e.g., De Neys et al., 2003; Manktelow & Fairley, 2000; Politzer & Bourmaud, 2002; Stevenson & Over, 1995). For example, Byrne (1989) observed that endorsement of the MP inference was greatly suppressed when a second conditional premise (i.e., if *r*, then *q*) served as an additional requirement for *q* to occur (e.g., "if she has an essay to write, then she will study late in the library; if the library stays open, then she will study late in the library; she has an essay to write; will she study late in the library?").

Subsequent research with this approach has demonstrated that reasoners are sensitive to both the quality and quantity of additional information. Chan and Chua (1994), for example, showed that the decrease in acceptance rates for MP depends upon the perceived importance or salience of the additional antecedent. Thus, given the major premise "If Steven is invited, then he will attend the dance party", the suppression of MP was greater when the additional antecedent referred to a necessary requirement for the consequent (e.g., "If Steven completes his report tonight, he will attend the dance party") than when it specified a less important condition (e.g., "If Steven knows the host well, he will attend the dance party"). In other words, the more necessary the additional requirement was perceived to be for the occurrence of the consequent, the more reluctant reasoners were to accept conclusions from the original premise. Not only do participants consider the type of additional information, they also take into account the amount of such information. Specifically, De Neys et al. (2003) recently manipulated the number of disabling conditions explicitly presented (between zero and four), and found that MP acceptance decreased in a linear function with every additional disabler.

The aforementioned studies all involve the presentation of additional information prior to conclusion evaluation. In these studies, reasoners are in effect required to integrate additional information with the original premise set before committing to a conclusion. A more direct test of defeasible reasoning is provided by studies where additional information is presented *after* an initial conclusion has been endorsed (Evans, 2002). A few recent studies have taken this approach, by exploring how people revise their beliefs when new information explicitly contradicts a previously derived conclusion.

For example, Elio and Pelletier (1997) gave participants complete standard conditional arguments, containing both premises and conclusions (e.g., MP: if *p* then *q*; *p*; therefore, *q*). This information was said to be well-established knowledge at time 1. Subsequently, they were given a statement contradicting the earlier conclusion (i.e., ~*q*); this information had come to light at a later time 2, but was also said to be well-established knowledge and should be considered true. Their task was to indicate how they would reconcile this contradiction, by revising some part of the information given at time 1. A majority of participants preferred to disbelieve the conditional premise (i.e., the hypothesis) rather than to abandon their belief in the minor premise (i.e., the observation). Subsequent research using this methodology (Dieussaert, Schaeken, De Neys, & d'Ydewalle, 2000; Politzer & Carles, 2001) has shown that when given a greater opportunity to express degrees of uncertainty in their responses, participants prefer to doubt the truth of the conditional premise rather than to categorically deny it. In addition, the willingness to revise the conditional premise depends on prior belief, such that reasoners are less likely to revise a conditional premise when they think it is highly plausible.

Finally, a particularly interesting study pertaining to the defeasibility of reasoning involves the use of "combined" arguments (George, 1999). In this study, a separate conditional argument served as the additional information. Specifically, participants were given two MP arguments, containing different antecedents but the same consequent (allowing for an evaluation of a common conclusion). Furthermore, different modals were used to assert degrees of uncertainty into the conditional premises; the two modal terms were either in the same direction (i.e., converging arguments) or in the opposite direction (i.e., diverging arguments). An example of a diverging argument is as follows: "If she meets Nicolas, it is very improbable she will go to the swimming pool; if she meets Sophie, it is very probable she will go to the swimming pool; she meets Nicolas; she meets Sophie".

How did participants resolve the conflict inherent in such diverging arguments? George (1999) found a strong tendency to choose a conclusion intermediate between the modal terms in the premises (i.e., in the above example, "the odds are the same that she will go to the swimming pool or not"). In other words, they adopted a compromise strategy when faced with contradictory information. A different picture emerged, however, when the content of the premises provided cues about the relative importance of the two arguments. In this case, responses shifted towards the modal of the dominant argument. Thus, when given the premises "If his mother is very ill, it is very probable that he will leave to visit his family" and "If he passes his exam, it is very improbable that he will leave to visit his family", the former argument was perceived to be more important than the latter; consequently, participants shifted their conclusions towards "it is very probable that he will leave to visit his family".

In general, these studies exploring the defeasibility of reasoning have contributed to our knowledge of informal reasoning by demonstrating that reasoners readily accommodate new information when making inferences. Thus, the introduction of additional information, especially when it contradicts currently held beliefs, results in either lower inference rates or a reduction in the degree of certainty people express in their conclusions. While this research has introduced various modifications of the traditional deductive paradigm, the tasks are still essentially formal in nature (i.e., they contain simple premises and conclusions from textbook logic). The reliance on these tasks may restrict the ability to explore the complexities of informal reasoning, which may better be achieved by asking different types of questions and by developing entirely new tasks. In the following section, I describe research that differs more radically from the formal approach.

*Asking non-deductive questions*

Most of what we know about human reasoning has come from studies employing a small number of formal reasoning tasks, in which participants are instructed to engage in deductive reasoning (e.g., draw necessary conclusions, determine whether logical possibilities are true or false). However, the inferential processes examined in such tasks do not seem to fully capture the complexities of everyday reasoning, and likely only constitute a subset of the processes involved in understanding and evaluating informal arguments (Evans & Thompson, 2004). As these formal tasks are unlikely to reveal the essential characteristics of how people reason in informal settings, a move away from the traditional paradigm is required. In this section, I discuss research that has contributed to

our understanding of informal reasoning by virtue of asking questions different from those used in traditional deductive reasoning tasks.

An example of such work is provided by studies exploring invited inferences. In these studies, the focus is not on whether participants endorse the traditional inferences of standard logic, but rather whether they judge certain situations or outcomes to be pragmatically implied (Geis & Zwicky, 1971). For instance, some studies have shown that conditional statements often are perceived to have implications about the truth status of their propositions. Bonnefon and Hilton (2004), for example, found that a consequential conditional such as "If Cedric takes up his new job, his life will improve" invites the belief that *p* is true (i.e., that Cedric will take up his new job). Similarly, Thompson and Byrne (2002) observed that counterfactual conditionals like "If Sarah had gone to Moose Jaw, then Tom would have gone to Medicine Hat" imply the falsity of both *p* and *q*. Collectively, these and other related studies (e.g., Byrne & Egan, 2004; Feeney, Scrafton, Duckworth, & Handley, 2004; Fillenbaum, 1974, 1975, 1976; Thompson, Evans, & Handley, 2005) illustrate that reasoners often attribute intentions on behalf of the speaker, by making judgments about what can pragmatically be implied (as opposed to logically inferred) from everyday statements.

Another illustration of the utility of asking non-traditional questions has recently been provided by Evans, Handley, and Over (2003). Instead of examining inference or truth-table patterns, these researchers asked participants to judge the *probability* of conditional statements. Their motivation for doing so was in part due to evidence emerging over the last 10 years supporting a probabilistic account of human reasoning. That is, several authors have proposed that everyday reasoning is uncertain and that

performance on deductive tasks therefore should be interpreted within a probabilistic framework (e.g., Chater & Oaksford, 1999; Kirby, 1994; Liu et al., 1996; Manktelow, Sutherland, & Over, 1995; Oaksford & Chater, 2001, 2003; Oaksford, Chater, & Grainger, 1999; Oaksford, Chater, & Larkin, 2000; Oberauer & Wilhelm, 2003; Politzer & Bourmaud, 2002). The standard deductive paradigm is not, however, particularly well-suited to explore such probabilistic effects in reasoning, as people are generally instructed to make binary decisions about validity or truth/falsity. The methodology of asking participants directly for probability estimates has provided important new insights into the question of how people interpret conditional statements (for a review, see Evans & Over, 2004). For instance, a close relationship has been documented between the judged probability of a conditional statement and the conditional probability of its consequent given its antecedent, that is, P($q$/$p$).

Other studies have investigated informal reasoning by asking questions that differ even more radically from those used in traditional deductive tasks. For instance, participants have been asked to make judgments about the acceptability and assertability (e.g., Edgington, 1995) of conditional statements. Exploring conditionals phrased as inducements (e.g., promises) and deterrents (e.g., threats), Fillenbaum (1975, 1977) documented the important role of the relationship between the values expressed in the antecedent and consequent terms. For example, when there was a disproportion between the action being induced and the inducement being offered (e.g., "If you break your mother's arm, I'll give you $50"), participants judged such promises to be strange or extraordinary. Similarly, Over, Manktelow, and Hadjichristidis (2004) recently asked participants to judge the "goodness" of conditional obligations such as "If you go for a

run, then you must take a shower". These judgments were found to be related to performance on another non-deductive task, where participants were presented with the four possible outcomes of the conditional (i.e., *pq*, *p*~*q*, ~*pq*, and ~*p*~*q*) and asked to arrange them in order of preference. Specifically, these conditionals were judged to be good rules when the *pq* possibility (i.e., going for a run and taking a shower) was preferred to the *p*~*q* possibility (i.e., going for a run but not taking a shower).

An especially promising approach to the study of informal reasoning involves asking questions relevant to the social and communicative functions of everyday arguments. Such statements are typically uttered with a specific communicative purpose in mind, whereby the speaker is trying to influence the opinions and/or behaviour of other people (e.g., Evans, 2005b; Fillenbaum, 1978, 1986; Rips, 1998). As a traditional analysis based solely on the types of inferences people endorse is insufficient to capture the social and communicative aspects of informal reasoning, researchers have recently begun to ask questions that are better suited for this purpose. These questions aim to address such issues as the goals of a statement, the types of arguments generated for and against a position, and the emotional reactions that arise in various circumstances (e.g., Beller, 2002; Fiddick, 2004; Thompson et al., 2005; Verbrugge et al., 2004).

An example of the benefits of this approach is provided by Thompson et al. (2005), who investigated a class of conditionals they termed persuasions and dissuasions. These statements are attempts to influence or change the opinions of other people. Thus, in the persuasion "If the railway line is closed, then the taxpayers will save millions of dollars", the speaker utters *q* in an attempt to convince the listener that *p* is desirable and, consequently, that *p* should happen. In addition to the traditional conditional inference

task, Thompson et al. used a variety of methods to explore how these statements are interpreted. For example, to assess whether participants grasped the speaker's goal when uttering a persuasion, they were asked to indicate what they thought the speaker's position on *p* was. Participants were found to have little difficulty in understanding the speaker's goal, responding that the speaker's position on the issue was that *p* should happen.

An interesting finding of this study was that performance on these tasks often depended on whether participants were asked to base their responses on their own perspective or on what they believed the speaker intended to communicate. These discrepancies suggest that persuasions and dissuasions are interpreted on two distinct levels: on the one hand, people have little difficulty understanding the purpose of these statements, but on the other, may or may not actually be persuaded by the argument (Evans, 2005b; Thompson et al., 2005). For instance, performance on the standard inference task indicated that reasoners draw more inferences when reasoning from the speaker's perspective, suggesting that they are more likely to question the premises when reasoning from a personal perspective (especially when the premises are inconsistent with their own beliefs). A similar pattern was observed on another (non-deductive) task, in which participants were asked to list arguments both supporting and contradicting the position advocated in the statement. Here, results indicated that both the number and types of arguments generated depended on whether reasoners approached the task from their own or the speaker's perspective.

Collectively, the findings of these studies contribute to an understanding of informal reasoning by examining reasoning processes other than those typically gauged

in formal deductive tasks. That is, instead of asking whether certain conclusions follow, or whether various possibilities are true or false, these studies have begun the task of explaining people's ordinary reasoning competence directly, without the limitations imposed by the traditional formal paradigm. To this end, a number of issues have been addressed, ranging from an analysis of the underlying pragmatics of everyday reasoning to the processes by which people judge questions of likelihood and assertability. These studies provide but a starting point; the future should witness a veritable revolution in the types of questions we ask participants in reasoning studies and the types of tasks we engage them in.

*Informal argumentation*

Finally, I will in this section describe research that explores informal reasoning directly by completely abandoning the traditional deductive framework. Broadly speaking, this approach involves the study of argumentation, a field that has until recently attracted little interest from reasoning researchers (Neuman & Weizman, 2003; Oaksford & Hahn, 2004; Rips, 2002). The methodology used to examine the processes involved in argumentation differs quite radically from the formal approach, in part because the quality, or soundness, of arguments is not captured by any formal criterion (Neuman, Glassner, & Weinstock, 2004). In other words, participants in these studies are not given explicit premises from which they are asked to generate or evaluate a conclusion. Instead, they are presented with an argument (for instance, from a newspaper editorial or in the form of a dialogue between two people), and then may be asked to evaluate its quality (e.g., Baron, 1995; Rips, 2002; Stanovich & West, 1997) or to generate possible objections to the argument (e.g., Edwards & Smith, 1996; Neuman, 2003; Shaw, 1996;

Toplak & Stanovich, 2003). Alternatively, they may be asked to provide an argument of their own, by justifying their position on a topic (e.g., Brem & Rips, 2000; Kuhn, 1992; Perkins, 1985). I will here attempt to illustrate the potential contributions of this approach to the understanding of informal reasoning.

A number of interesting findings have emerged from studies that explore how participants evaluate weak or unsound arguments. For instance, Neuman (2003) gave participants several types of informal reasoning fallacies, such as the argument from ignorance. This fallacy involves concluding that a proposition is true (or false) because it has not been proven false (or true), as in "No one has proved that God does not exist; therefore we can conclude that God exists". As Oaksford and Hahn (2004) point out, the problem with this form of argumentation is that absence of proof is generally considered insufficient to establish the truth/falsity of a proposition; if it were, all sorts of nonsense conclusions would be warranted (e.g., replace "God" with "flying pigs" in the previous example).

A majority of the participants in Neuman's (2003) study indicated that there was a problem with this type of argument (see also Neuman & Weizman, 2003; Weinstock, Neuman, & Tabak, 2004). However, far fewer participants were able to articulate the nature of the problem or to provide appropriate counter-arguments, suggesting that their understanding of this fallacy was more restricted. In addition, the ability to resist these informal reasoning fallacies was correlated with prior belief. For example, participants who expressed a belief in God were less likely to object to the argument from ignorance concluding God's existence, compared to those who did not believe in God. In fact, the difficulty in divorcing one's belief from the process of argument evaluation is a common

finding in studies investigating informal reasoning (e.g., Edwards & Smith, 1996; Klaczynski & Gordon, 1996; Klaczynski & Robinson, 2000; Lord, Ross, & Lepper, 1979; Nickerson, 1998; Stanovich & West, 1997). Thus, when asked to rate the quality of arguments or the strength of evidence, most people give lower ratings to arguments and evidence incompatible with their beliefs and personal convictions. These findings, of course, resemble the robust belief bias effect in deductive reasoning (e.g., Evans et al., 1983), and therefore suggest a convergence between the processes involved in formal and informal reasoning.

The pervasive influence of prior belief in informal reasoning is not limited to argument evaluation, but has also been demonstrated in studies where participants are asked to justify their position on real-world issues or to generate possible objections to arguments (e.g., Edwards & Smith, 1996; Perkins, 1985; Shaw, 1996; Toplak & Stanovich, 2003). These studies have shown that most people are reluctant to provide objections to their own position and are instead motivated to disconfirm *incompatible* arguments. These findings reflect a general tendency for people to be partial to evidence and arguments supporting the side they favour (Nickerson, 1998). Furthermore, Baron (1995) reported that one-sided arguments were rated as stronger than two-sided arguments, suggesting that the ubiquitous nature of this *my-side bias* in reasoning may in part be due to a (mis)perception of what makes an argument strong. For our purposes, it is interesting to note that these findings concerning the role of prior belief would likely not have been documented with the use of a formal approach, where participants are generally not asked to justify their answers or to produce counter-arguments to conclusions.

Besides the inability to detach prior belief when evaluating and responding to arguments, what other difficulties have been uncovered in studies investigating informal reasoning? One line of inquiry has demonstrated that people are not very good at providing evidence when justifying their positions. Kuhn (1992), for example, asked participants for their causal explanations of various social issues (e.g., what causes prisoners to return to crime?) and also for evidence that would support their explanations. Kuhn reported that many participants did not provide "genuine" evidence, such as covariation data (where variation in the cause corresponds to variation in the outcome); instead, they often offered "pseudoevidence", for instance in the form of a description or story of how the cause could lead to the effect. This finding may in fact indicate a more general difficulty in understanding the conceptual distinction between explanations and evidence (e.g., Brem & Rips, 2000; Kuhn, 1989, 1992, 2001).

Another area of research has focused on identifying factors that can predict informal reasoning (see also van Gelder, Bissett, & Cumming, 2004, for a recent attempt to improve informal reasoning skills by computer-based practice). In an early study, Perkins (1985) found that education resulted in only modest gains in various skills involved in argument construction. Cognitive ability, as measured by IQ tests, was found to be a better predictor, a result since replicated by the individual differences approach of Stanovich and West (1997, 1998). However, this research has failed to provide a consistent picture, as other studies have found no association between cognitive ability and degree of my-side bias (e.g., Klaczynski & Gordon, 1996; Klaczynski & Robinson, 2000; Toplak & Stanovich, 2003). A final predictor of informal reasoning involves participants' thinking dispositions (or cognitive styles); for example, the ability to

evaluate arguments independently of prior belief has been found to positively correlate with those thinking dispositions that reflect a greater degree of open-mindedness, skepticism, and cognitive flexibility (Stanovich & West, 1997, 1998).

In summary, the studies reviewed in this section have informed us about some of the difficulties people have with informal arguments, and some of the factors that predict skill in informal reasoning. What we know very little about, however, is whether performance on these informal tasks is correlated with performance on formal reasoning tasks. The most informative study in this regard is that of Stanovich and West (1998), who explored individual differences across a number of reasoning tasks. While their main concern was to establish a link between reasoning performance and cognitive ability/thinking dispositions, they did report significant correlations between formal reasoning (e.g., syllogistic reasoning) and an informal task. This finding suggests that there is some overlap between the two domains (but see Neuman, 2003, and Ricco, 2003, for less encouraging results); however, much more work is clearly needed to resolve this issue.

*Conclusions*

In this introductory chapter, I have attempted to explore the extent to which research in formal reasoning has contributed to an understanding of everyday informal reasoning. As previously discussed, we do not really know whether the empirical and theoretical advances made in the psychology of reasoning generalize to informal settings, or whether these contributions instead are limited to explaining performance on a small number of formal deductive tasks (Evans & Thompson, 2004). Based on the studies here reviewed, there seems to be an inverse relationship between the contributions of these

studies to informal reasoning and their degree of compliance with the traditional deductive paradigm. That is, by inheriting the restrictions of the formal approach, many studies relying on the use of deductive tasks and traditional instructions can, at best, make only indirect contributions to an understanding of the subtleties and complexities of informal reasoning. Other studies, which differ more radically from the deductive paradigm, have a much greater potential to uncover insights into informal reasoning processes.

Good examples of this potential are provided by studies asking participants different types of questions than those traditionally used in formal reasoning tasks. Thus, researchers have recently begun to address issues such as the pragmatic implications of everyday utterances, participants' understanding of the purpose or goals of statements, and the degree to which speakers are perceived to be successful in achieving these goals. Other studies have developed new tasks that aim to explore different types of inferential processes than those deductive tasks are designed to measure. Examples of this approach include participants' ability to justify their position on a topic, to detect weaknesses in other people's argumentation, and to generate appropriate counter-arguments.

In the future, more contributions will surely follow from studies applying an informal approach, whether by abandoning traditional instructions, asking different questions, or developing new tasks. As we continue to learn more about people's ordinary reasoning competence, we will be in a better position to ascertain the nature of the relationship between formal and informal reasoning. On the one hand, I have argued that the fundamental differences between these two types of reasoning suggest that the study of formal reasoning will be unable to tell us much about informal reasoning. On the

other hand, the research described in this introduction also points to some degree of convergence between the two domains (e.g., the difficulty in evaluating arguments independently from prior belief, correlations with measures of cognitive ability/styles). Such findings give cause for optimism, as they suggest that the contributions of formal reasoning research in fact are relevant to reasoning in everyday life, and also provide support for the continued existence of the formal paradigm. This issue will be revisited in the General Discussion.

*The present experiments*

The overall goal of the current series of experiments was to illustrate the benefits of applying an informal approach to the study of human reasoning, and more specifically, the area of conditional reasoning. Two types of realistic conditionals were chosen for this purpose: conditional inducements (i.e., promises and threats) and conditional advice (i.e., tips and warnings). These conditionals seem to be very rich in pragmatic information and are normally uttered with some social or communicative purpose in mind; in particular, they are attempts to influence or change the behaviour of other people. As the traditional deductive paradigm is ill-equipped to explore such pragmatic factors, I argue that a proper understanding of how people interpret and reason with these statements requires the use of a more informal approach.

The first section includes two experiments that attempt to explore why conditional inducements and advice are interpreted differently, and also why they elicit different reasoning patterns. [2] Participants in Experiment 1 were asked to rate these statements on a number of pragmatic variables, while participants in Experiment 2 performed a

[2] These two experiments have previously been published in Ohm and Thompson (2004); the same applies to Experiment 3, which is published in Ohm and Thompson (2005). Since Experiments 1-3 are taken verbatim from these sources, please note that there is some redundancy within the text.

standard deductive reasoning task. The results from these studies provide support for the idea that people recruit a substantial amount of background knowledge when encountering these conditionals. The next unit presents Experiment 3, in which we apply a recent probabilistic approach to human reasoning (e.g., Evans & Over, 2004; Oaksford & Chater, 2001); the results suggest that estimates of different conditional probabilities can predict a number of judgments people make about conditional inducements and advice.

In the following section, two follow-up studies (Experiments 4 and 5) attempt to resolve a counter-intuitive finding observed in Experiment 3. To foreshadow this issue, the degree to which inducements and advice were perceived to be effective in changing behaviour did not depend on how likely they were to be true; furthermore, these two measures were affected by different variables. Consequently, the main purpose of Experiments 4 and 5 was to explore this independence hypothesis; the results suggest that the dissociation between probable truth and behavioural effectiveness is a robust finding. A final study, Experiment 6, is then presented. This experiment applies a decision-theoretic analysis (e.g., Manktelow & Over, 1991, 1995; Over, et al., 2004) of inducements and advice and explores how perceptions of utilities and preferences can be tied to judgments people make about these conditionals. Finally, the General Discussion addresses a number of theoretical issues, and in particular, how we should view the relationship between formal and informal reasoning.

# EXPERIMENTS 1 & 2

## EVERYDAY REASONING WITH INDUCEMENTS AND ADVICE

Conditional reasoning is a common form of reasoning that entails drawing inferences about statements of the form *if p then q*, where *p* and *q* refer to the antecedent and consequent terms, respectively. Most of what we know about conditional reasoning comes from a small number of deductive reasoning tasks. In these tasks, participants are asked to assume the truth of the premises and reason only on the basis of the information explicitly provided. However, there is a growing consensus that this deductive paradigm is too restrictive to fully capture the complexities of everyday reasoning (see Evans, 2002). Instead, a complete understanding of people's reasoning processes will require a multi-dimensional approach that goes beyond the limitations imposed by the deductive paradigm (Beller, 2002). The goal of this paper is to apply such an approach to the study of conditional relations.

An important step towards this goal has been the trend to use familiar, pragmatically rich materials in deductive reasoning studies. Much of the early research on conditional reasoning used abstract or artificial problem materials, in which the terms represented by *p* and *q* were not connected in any meaningful way (Fillenbaum, 1986). Abstract tasks were thought to afford a pure means of investigating reasoning performance, avoiding "contamination" by prior knowledge, experience, or beliefs (e.g., Manktelow, 1999). However, the reasoning people engage in during their everyday lives is not devoid of meaning, and the processes by which people interpret, contextualise, and modify given information are integral to making inferences in most real-world situations.

These interpretive processes must be understood before a complete model of reasoning can be achieved (Thompson, 1997, 2000).

In the past two decades, substantial effects of content and context on reasoning have been documented using thematic or realistic material. For example, conditional reasoning is influenced by the believability of conditional premises (e.g., George, 1995, 1997; Politzer & Bourmaud, 2002; Stevenson & Over, 1995; Thompson, 1996), the reasoner's perceptions of necessary and sufficient relations between *p* and *q* (e.g., Cummins, Lubart, Alksnis, & Rist, 1991; Hilton, Jaspars, & Clarke, 1990; Thompson, 1994, 1995, 2000), and the pragmatics of language use and social discourse (e.g., Fillenbaum, 1978, 1986; Manktelow & Over, 1991, 1995; Politzer, 1986; Rumain, Connell, & Braine, 1983; Sperber, Cara, & Girotto, 1995).

In the present paper, we investigate two types of realistic conditionals, namely conditional inducements and advice. Because these statements are rich in pragmatic information, they appear well suited to a multi-dimensional investigation of interpretational processes in reasoning. Conditional inducements are speech acts uttered in an attempt to make another person do, or refrain from doing, some action (Fillenbaum, 1986). Thus, a promise (e.g., "If you wash the car, I will give you $10") encourages an action *p* by offering *q* as a reward, while a threat (e.g., "If you continue drinking, I will file for a divorce") aims to deter behaviour by pointing to potential punishment. Conditional advice, on the other hand, involves recommendations about future behaviour. A tip (e.g., "If you show up early for work, you will impress your boss") highlights potentially good outcomes for the addressee resulting from the satisfaction of *p*, while a

warning (e.g., "If you tease the dog, it will bite you") outlines negative consequences that will likely ensue unless one refrains from doing *p*.

Conditional inducements and advice are types of deontic conditionals, in that they describe actions that people may, should, or must take (e.g., Manktelow & Over, 1995). For example, if *p* is fulfilled in a promise, then the speaker, to be fair, is obligated to reciprocate by giving *q*. Similarly, a tip describes behaviour that should be performed in order to obtain *q*. Deontic conditionals require reasoners to make use of a considerable amount of background knowledge and beliefs. A proper understanding of a promise, for instance, requires a consideration of the communicative purpose associated with this utterance, as it constitutes an attempt to control or manipulate the behaviour of other people (Fillenbaum, 1978, 1986).

Although they are similar in many respects, conditionals in the form of inducements and advice seem to differ for pragmatic reasons. A promise, for instance, takes the general form "If you do *p* for me, I will do *q* for you". To be interpreted as a true promise, we must assume that *q* is useful or desirable for the addressee (or else there would be no incentive to do *p*) and also that *p* is useful or desirable for the speaker (or else *q* would not be offered as an inducement). Tips, in contrast, have the form "If you do *p*, then *q* will follow". Again, we must assume that *q* is desirable for the addressee; however, in this case, the speaker is unlikely to find *p* useful. A similar pragmatic approach can be used to differentiate threats and warnings. Specifically, the addressee wants the absence of *q* in both types of statements. However, the speaker only wants the absence of *p* in a threat, and is neutral as to the outcome of *p* in a warning.

Consistent with the assumption that inducements and advice are pragmatically different, Newstead, Ellis, Evans, and Dennis (1997) recently reported differences in reasoning performance between these statements. In one task (Experiment 4), they asked participants to evaluate the validity of four inferences. These inferences are called Modus Ponens (MP: '*p*, therefore *q*'), Denying the Antecedent (DA: '~*p*, therefore ~*q*), Affirming the Consequent (AC: '*q*, therefore *p*'), and Modus Tollens (MT: '~*q*, therefore ~*p*'). Newstead et al. found that reasoners endorsed more inferences of all kinds for conditional inducements (i.e., promises and threats) than for conditional advice (i.e., tips and warnings).

The finding that more inferences (especially DA and AC) are endorsed for inducements than for advice is consistent with earlier findings that conditional inducements seem to invite a biconditional interpretation (Fillenbaum, 1975, 1976, 1978; Geis & Zwicky, 1971; Light, Girotto, & Legrenzi, 1990; Markovits & Lesage, 1990; Politzer & Nguyen-Xuan, 1992). For example, Fillenbaum (1975) found that participants are very willing to accept the inverse of a conditional statement (i.e., *if ~p, then ~q*) as following from conditional promises and threats. Although technically a fallacy, this inference is pragmatically sound. That is, the listener interprets the speaker's intention as implying that *q* will only result if *p* is performed. Indeed, a promise will lose its force as an inducement if the reward is given regardless of whether *p* is fulfilled (e.g., Fillenbaum, 1976). Similarly, a threat can only be an effective deterrent if *q* is contingent upon the addressee's behaviour, whereby refraining from *p* should result in ~*q*. Advice, on the other hand, seems less likely to be interpreted in this way. Thus, given the tip "If you

show up early for work, you will impress your boss", one can easily imagine ways other than *p* to bring about *q* (e.g., impressing one's boss by working hard).

Newstead et al. (1997) proposed that a crucial difference between inducements and advice is the degree of certainty that they express. Specifically, they posited that promises and threats express a more certain, less probabilistic relationship between antecedent and consequent than do tips and warnings. For example, consider the "car wash" promise. If the antecedent of this promise has been satisfied (i.e., the car has been washed), the probability that the consequent (i.e., the reward) will occur would seem quite high. In contrast, the tip "If you show up early for work, you will impress your boss" suggests a weaker contingency between *p* and *q*: performing the antecedent (i.e., showing up early for work) may or may not result in the occurrence of the consequent (i.e., impressing your boss). In the current paper, we tested this hypothesis by asking participants how likely *q* is to come about given the occurrence of *p* both for advice and inducement conditionals.

Newstead et al. (1997) also speculated about the reason that inducements should suggest a stronger sense of certainty than advice. Specifically, they proposed that this sense of certainty is tied to the degree to which the speaker is perceived to have control over the occurrence of the consequent. Moreover, this degree of control is perceived to be greater for inducements than advice because the speaker making a promise or a threat can usually determine whether or not the consequent will take place. For instance, in the case of the "car wash" example, the speaker is able to determine the outcome of this exchange because he/she has the power to enforce or withhold the reward. For warnings and tips,

however, the speaker can better be regarded as offering a best guess, and will have little direct control over the outcome of *q*.

Evans and Twyman-Musgrove (1998) tested this hypothesis by constructing inducement and advice conditionals that were both high and low in speaker's control. These researchers reported a strong influence of speaker's control on reasoning: the frequency with which MP, DA, AC, and MT were endorsed was higher when the speaker of the conditional statement was in control of the consequent than when he/she was not. In addition, once statements had been equated in terms of the degree of speaker's control, the difference in endorsement rates between advice and inducements was considerably attenuated. Evans and Twyman-Musgrove concluded on the basis of these results that degree of speaker's control accounts for much of the difference in reasoning between inducements and advice.

However, inducements and advice are information-rich statements. Thus, one might expect that there could be other aspects, apart from speaker's control, that are crucial to their interpretation. Some factors, such as perceived necessity and sufficiency, have been identified in the literature as strong predictors of conditional reasoning (e.g., Cummins et al., 1991; Thompson, 1994, 1995, 2000). Perceived sufficiency is defined as the degree to which the occurrence of *p* guarantees the occurrence of *q*; perceived necessity refers to the degree to which the absence of *p* guarantees the absence of *q* (Thompson, 2000). Much research has shown that statements that are interpreted to be both sufficient and necessary invite a biconditional interpretation (e.g., Thompson, 1994, 1995, 2000); thus, a possible explanation for reasoners' tendency to give inducements a biconditional interpretation is that they are interpreted as necessary and sufficient

relations. For example, in the case of an inducement, the fulfillment of *p* would ordinarily guarantee the promised or threatened *q*; whereas for advice, *q* may or may not come about given the occurrence of *p*. The stimuli used by Evans and Twyman-Musgrove (1998) were not rated for necessity and sufficiency; the possibility therefore remains that the reported effect of speaker's control can instead be accounted for by these two factors. Alternatively, it is possible that speaker's control is one variable that contributes to perceived necessity and sufficiency, such that a high degree of control results in a sufficient and necessary relationship between *p* and *q*. One goal of the present studies was to investigate the relationship between perceived necessity/sufficiency and control, as well as the relationship between these variables and inference patterns.

In addition, we investigated a number of other variables that could potentially explain why the link between *p* and *q* appears to be stronger for conditional inducements than conditional advice. Fillenbaum (e.g., 1976, 1978) proposed that inducements differ from advice because an obligation arises for the speaker in the former, but not the latter. Specifically, the speaker of a promise has an obligation to give the reward, *q*, if the addressee performs *p*; likewise, an obligation arises for the speaker of a threat to withhold the punishment if the addressee refrains from doing *p* (Beller, 2002). In contrast, when giving advice, the speaker is describing a potentially fruitful course of action, and would not normally be seen as obligated to make *q* or ~*q* happen. To test this hypothesis, we asked participants whether they perceived the speaker to have an obligation to ensure the occurrence or non-occurrence of the consequent, predicting that inducements would incur a greater degree of obligation than advice.

Finally, the speaker of an inducement can also be regarded as having more at stake than the speaker giving advice. For both promises and threats, the speaker has indicated that he or she is prepared to act directly in the situation. In a promise, for example, the nature of the inducement compels the speaker to give something that would not otherwise be offered. In addition, the speaker will be very interested to know what the addressee ends up doing, as the speaker is unlikely to part with the reward unless *p* is performed. In contrast, there seems to be little at stake personally for the speaker giving advice. For example, in the "work" scenario the speaker should be less affected if showing up early fails to impress the boss; at most, there might be a slight loss of credibility for the speaker, but there is little other direct consequence.

In sum, we have posited several variables that may explain why conditional inducements and advice are interpreted differently, and why they produce different inference patterns on conditional reasoning tasks. However, a complete understanding of these statements may require an analysis that goes beyond the traditional deductive paradigm. Beller (2002), for example, argued that a focus on the inferential use of inducements is by itself unlikely to be adequate in fully capturing the meaning of these speech acts. Beller instead advocates a multi-level approach, in which motivational, linguistic, deontic, pragmatic, and emotional considerations are integrated.

We propose that adopting a similar approach may shed some light on the differences between conditional inducements and advice. In particular, we propose that an understanding of these statements requires an analysis of their purpose or perlocutionary force (e.g., McCawley, 1993). Both inducements and advice have as a subtext the desire to change the behaviour of the addressee (e.g., Fillenbaum, 1986). In

the tip “If you show up early for work, you will impress your boss”, for example, the goal of the statement is to increase the probability that the addressee shows up early for work. Similarly, the goal of the promise “If you wash the car, I will give you $10”, is to persuade the addressee to wash the car. The consequent, in both instances, describes the reason why the behaviour should occur (i.e., to impress the boss or to earn $10). This rationale also applies to statements with negative outcomes (i.e., warnings and threats), with the exception that the goal of these statements is to *decrease* the probability of the antecedent occurring, and the consequent describes the reason why the antecedent should *not* occur.

Thus, both advice and inducements can be interpreted as speech acts whose goal is to change the behaviour of the addressee. In this regard, however, inducements seem to be more ambitious speech acts than advice. Their manifest goal is to induce a behavioural change (Fillenbaum, 1986), and the speaker is taking a personal involvement to increase the chances of such behaviour change. Advice, in contrast, is more likely to be perceived as a recommendation about future behaviour, and the speaker will not be too concerned whether the addressee follows the advice or not. We would thus expect inducements to be perceived as more effective in changing behaviour (i.e., in bringing about or deterring the action described in the antecedent, *p*) than advice. To explore this hypothesis, we asked participants how likely the addressee is to perform *p* both in the absence and presence of the speaker’s utterance. The difference between these two scores provides a measure of the statement’s perceived effectiveness. For statements with positive outcomes (i.e., tips and promises), this difference score provides an estimate of the *increased* probability that *p* will be performed as a result of the speaker’s utterance. Conversely, for statements with

negative outcomes (i.e., warnings and threats), the difference score reflects a *decrease* in the probability of *p* occurring as a result of the speaker's utterance. We predicted that this difference score would be greater for inducements than advice, confirming the hypothesis that inducements indeed are perceived to be more effective statements in terms of changing behaviour.

Although the main focus of this paper was to explore differences between inducements and advice, we were also interested in the relationship between statements expressing positive and negative outcomes. That is, are there differences between a promise and a threat, or between a tip and a warning? Newstead et al. (1997) chose to highlight the differences between inducements and advice; however, a closer inspection of their data also reveals a number of differences *within* these two classes of conditionals. For instance, more inferences were overall endorsed for negative statements (i.e., warnings and threats) than for positive statements (i.e., tips and promises). Another focus of our study, therefore, was to explore the relationship between negatively and positively phrased inducements and advice on the dimensions we have identified.

In sum, the goal of this paper is to extend the work of Newstead et al. (1997) and Evans and Twyman-Musgrove (1998) by offering a detailed analysis of how inducements and advice are interpreted. In Experiment 1, participants were asked to rate conditional inducements and advice with respect to our six interpretive variables (i.e., speaker's control, stake, obligation, probability of *q*, sufficiency and necessity). In addition, we gathered ratings designed to reflect reasoners' sensitivity to the purposive difference (i.e., behavioural effectiveness) between inducements and advice, and examined the relationship between our measure of perceived behavioural effectiveness and our six

interpretive variables. In Experiment 2, participants performed a traditional arguments task, and we explored the relationship between our interpretive variables and performance on this task.

## Experiment 1

Inducements and advice were hypothesised to differ with respect to six interpretive variables. We predicted that inducements would be rated higher on the *probability of q given p* (reflecting a stronger link between *p* and *q*) and on the degree of *speaker's control* over the consequent than advice (Newstead et al., 1997). Relative to advice, inducements were also predicted to express a *necessary* and *sufficient* relationship, and to be rated higher on the variables of *stake* and *obligation* to the speaker. In addition, we predicted that inducements would be perceived to be more effective in changing the behaviour of the addressee (Fillenbaum, 1978, 1986). Finally, we tested the prediction that a statement's perceived effectiveness in changing behaviour would be related to the interpretive variables that we propose differentiate inducements and advice. To this end, we conducted multiple regression analyses with perceived behaviour change as the dependent measure, and our six interpretive variables as predictors.

### *Method*

*Participants*. A total of 109 University of Saskatchewan undergraduate students took part in this experiment. They participated in partial fulfillment of a course requirement for an introductory psychology course. [3]

*Materials*. A total of 40 conditional statements were used in this experiment. They were constructed to be similar to the four types of realistic conditionals used by Newstead

[3] Ethics support for this series of experiments was obtained on November 8th, 2001 (see Appendix A)

et al. (1997): tips, warnings, promises, and threats. There were 10 statements of each type. A brief paragraph preceded each statement and provided the conditional with a social context. A complete list of the statements and the accompanying contexts is provided in Appendix B. As an example, one tip read:

> Paula has recently found a new job. While talking to her friend, she is told:
>
> "If you show up early for work, you will impress your boss".

Following each statement, eight questions were presented, requiring participants to indicate their opinion on a 7-point scale (except for a yes-no option for question 4; see below). The questions were as follows:

1. Probability of *q*: how likely will *q* happen given *p*? (e.g., in general, how likely will you impress your boss if you show up early for work?).
2. Degree of speaker's control over *q* (e.g., how much control does the friend have over whether Paula will impress her boss?).
3. Stake to speaker (e.g., how much is at stake for the friend?).
4. Obligation (e.g., is the friend obligated to ensure that Paula's boss is impressed if she shows up early for work?).
5. Sufficiency (e.g., if Paula shows up early for work, how likely is it that her boss will be impressed?).
6. Necessity (e.g., if Paula does not show up early for work, how likely is it that her boss will be impressed?).
7. Probability of *p* (e.g., how likely is Paula to show up early for work?).
8. Probability of *p* without speaker's utterance (e.g., how likely would Paula show up early for work even without her friend's advice?).

*Procedure*. The participants were tested in large groups. Written instructions were provided, which informed the participants that the purpose of the study was to investigate how people interpret statements depicting various social situations. For each statement, all eight questions were presented on a single page.

Each participant was given a booklet of 20 statements, with every booklet containing five statements from each type (i.e., tip, warning, promise, threat). These booklets were composed as follows: the 10 statements within each type were randomly divided into two groups, allowing for 16 different combinations of the four types of statements. Each participant was assigned randomly to one of these combinations. The order of statements within the booklets was different for each participant, and was subject to the constraint that no more than two instances from the same category occurred in succession. The participants were asked to complete the questions at their own pace. On average, it took 30-40 minutes to complete the task.

Because the booklets were randomly distributed among large groups of participants, we did not obtain an equal number of responses for each of the 40 statements. On average, we collected approximately 53 responses per statement, with a minimum of 43 and a maximum of 63.

*Results and Discussion*

Data from 2 participants were discarded, as these participants only answered questions for a small number of statements. Table 1 shows the means and standard deviations on all variables except obligation (question 4, which is reported separately below). For the necessity variable, the scores were reversed, so that high scores would reflect a high degree of necessity. The variable labelled “change” was intended to be a

Table 1

*Mean ratings (standard deviations) on variables in Experiment 1 as a function of statement type*

| Type | Prob(*q*) | Control | Stake | Sufficiency | Necessity | Change |
|---|---|---|---|---|---|---|
| *Advice* | | | | | | |
| Tip | 5.15 | 3.00 | 2.62 | 4.95 | 3.90 | 1.83 |
| | (.69) | (1.03) | (1.00) | (.73) | (.56) | (.95) |
| Warning | 5.17 | 2.87 | 3.07 | 5.04 | 4.45 | .96 |
| | (.82) | (1.05) | (.99) | (.79) | (.74) | (.76) |
| Mean | 5.16 | 2.93 | 2.84 | 5.00 | 4.17 | 1.39 |
| *Inducements* | | | | | | |
| Promise | 5.55 | 5.78 | 3.44 | 5.98 | 4.87 | 2.55 |
| | (.80) | (.86) | (1.19) | (.68) | (.74) | (1.04) |
| Threat | 5.26 | 6.15 | 3.92 | 5.65 | 4.62 | 1.68 |
| | (.88) | (.87) | (.96) | (.80) | (.77) | (1.01) |
| Mean | 5.40 | 5.97 | 3.68 | 5.82 | 4.75 | 2.12 |

measure of a statement's effectiveness in changing behaviour. This variable was computed by subtracting the probability of *p* without the speaker's utterance (question 8) from the probability of *p* (question 7). This resulted in negative scores for threats and warnings (i.e., because the goal of these statements is to decrease the probability of *p* occurring); to make these statements comparable to positive ones, these scores were also reversed so that higher values represent greater effectiveness.

2x2 within-subject ANOVAs were performed on each of the dependent variables presented in Table 1. For each analysis, the factors were class of statement (inducements vs. advice) and valence of statement (positive vs. negative). To abbreviate reporting of the data, we will group the analyses by the independent variables, rather than by the dependent variable. That is, we first report the effects of statement class on all of the dependent measures; we will then do the same for the valence factor. For these analyses, we only report effects that were significant at $p < .05$.

*Inducements vs. advice*. Ratings were higher for inducements than for advice on all variables in Table 1, smallest $F(1, 106) = 15.82$. Thus, the link between *p* and *q* was stronger for inducements (evidenced by the higher probability of *q* given *p*), and inducements also differed from advice in that the speaker was perceived to have a higher degree of control over the outcome. These findings are consistent with the claims made by Newstead et al. (1997) and Evans and Twyman-Musgrove (1998). However, consistent with our hypotheses, inducements and advice differed along other dimensions as well. Specifically, the speaker of an inducement was regarded as having more at stake than the speaker giving advice, an inducement was perceived to express a more sufficient

and necessary relationship than advice, and inducements were also perceived to be more effective statements in terms of changing behaviour.

For the obligation variable, participants simply answered yes or no. We computed a 2x2 chi-square with statement class and absence/presence of obligation as classification variables. Participants were more likely to regard the speaker as obligated to ensure the outcome for inducements (63% of the overall responses) than for advice (18% overall), $\chi^2(1) = 447.26$. Thus, the speaker's obligation was perceived to be higher for inducements than for advice: in the case of promises, the speaker was perceived to have an obligation to ensure the occurrence of *q*, whereas for threats, the obligation was to refrain from *q*. Taken together, these results confirm our hypothesis that inducements and advice differ along a number of dimensions, including, but not limited to, the speaker's degree of control over the consequent.

*Positive vs. negative statements*. As previously discussed, a secondary aim of this paper was to explore the relationship between conditionals with positive outcomes (i.e., tips and promises) and negative outcomes (i.e., warnings and threats). For both inducements and advice, participants perceived that positive statements were more effective in changing behaviour than negative statements, $F(1, 106) = 162.59$. This result suggests that it may be easier to induce than to deter behaviour; that is, people may be more willing to comply with the speaker's goal when positive outcomes are spelled out (as in tips and promises) than when refraining from some behaviour only results in the absence of negative outcomes (as in warnings and threats). Another possible explanation for why negative statements are perceived to be less effective than positive statements in changing behaviour is that the listener may be more likely to doubt that the speaker of a

threat will in fact enforce the punishment than to doubt that the speaker of a promise will give the reward. That is, while the speaker of a promise is responsible for carrying through with *q* once the addressee has delivered *p*, the speaker of a threat is better regarded as *permitted* to punish the addressee who decides to perform *p* (e.g., Conison, 1997; Peetz, 1977).

There was also an asymmetry in the degree to which participants perceived an obligation to accrue for positive statements (53% overall) and negative statements (28% overall), $\chi^2(1) = 142.86$. Speakers of positive statements, therefore, were perceived to be obligated to ensure the outcome, but speakers of negative statements were perceived to be less obligated in ensuring its absence. However, this difference was entirely due to the pattern for inducements: only 37% of the responses for threats indicated that the speaker had an obligation, whereas this was true for 90% of the responses for promises, $\chi^2(1) = 316.85$. This difference between promises and threats is at odds with the results of Beller (2002), who reported that speakers of both threats and promises are obligated to reciprocate when the addressee cooperates with the speaker's wishes. Nonetheless, we still find a greater sense of obligation for the speaker of a threat than the speaker of a warning, confirming our main prediction that a greater sense of obligation arises for inducements than for advice.

There was not a consistent pattern for the remaining variables. For both inducements and advice, the speaker was perceived to have more at stake for negative statements than for positive statements, $F(1, 106) = 49.66$. Participants perceived positive statements to be more sufficient than negative statements, $F(1, 106) = 4.19$, but this trend was only present for inducements [interaction: $F(1, 106) = 16.42$]. Participants further

perceived positive statements as being less necessary than negative statements, $F(1, 106) = 6.31$, but this trend was only present for advice [interaction: $F(1, 106) = 75.66$]. Finally, the main effect of valence was not significant on the variables of probability of *q* and speaker's control, although the interactions with statement class were again significant [interaction: $F(1, 106) = 8.06$ and 14.43, respectively].

*Correlations between variables*. Although we found evidence that conditional inducements and advice differ on a number of variables, some of these variables were also correlated. The correlations between our interpretive variables are presented in Table 2. The responses to the obligation question were coded in such a way that a high correlation means that a high score on another variable (e.g., speaker's control) is associated with a tendency to attribute an obligation to the speaker. The correlations between control, stake, and obligation were fairly high (all $p < .01$, one-tailed). Thus, a speaker who is perceived to have control over the outcome is also likely to be perceived as having much at stake and to be obligated for the occurrence of *q*. Both control and obligation were further correlated with sufficiency, necessity, and perceived behaviour change, although the corresponding correlations were not significant for stake. Thus, as hypothesised, statements in which the speaker is perceived to have a lot of control are also statements that are perceived to express necessary and sufficient relations; furthermore, all three variables are related to the behaviour change variable. Finally, the probability of *q* was strongly correlated with sufficiency, and moderately correlated to speaker's control and the change variable. These findings show that the variables relevant to the difference between inducements and advice are correlated both with each other, as well as the degree to which a statement is perceived to be effective in changing the

Table 2

*Intercorrelations between variables in Experiment 1 collapsed across statement type*

| Variables | Prob(*q*) | Control | Stake | Obligation | Sufficiency | Necessity | Change |
|---|---|---|---|---|---|---|---|
| Prob(*q*) | --- | .30* | -.22 | .16 | .80** | .22 | .31* |
| Control | | --- | .47** | .68** | .65** | .49** | .48** |
| Stake | | | --- | .46** | .04 | .24 | -.01 |
| Obligation | | | | --- | .52** | .52** | .54** |
| Sufficiency | | | | | --- | .46** | .50** |
| Necessity | | | | | | --- | .35* |
| Change | | | | | | | --- |

Note. * $p <$ .05; ** $p <$ .01. All tests are one-tailed.

addressee's behaviour.

*Changing behaviour*. We have argued that a crucial variable that differentiates conditional inducements from conditional advice is the purpose of these statements, namely the degree to which they are perceived to be effective in changing the behaviour of the addressee. From the correlational analyses, we saw that a number of variables are correlated quite highly with behaviour change (especially control, obligation, and sufficiency). Therefore, in order to determine the factors that contribute to perceived behavioural effectiveness, we conducted two multiple regression analyses using behaviour change as the dependent variable, and entering the remaining variables in separate blocks. For the first analysis, sufficiency and necessity were entered at step 1, and accounted for approximately 23% of the variance [Adjusted $R^2 = .23$, $F(2, 37) = 6.89$, $p < .05$]. Entering control at step 2 allowed us to determine whether control would add to the prediction of change scores after the effects of sufficiency and necessity were partialled out. However, control did not contribute to a significant amount of the residual variance [Adjusted $R^2 = .24$, $F(1, 36) = 1.38$, $p > .05$]. Obligation, stake, and probability of *q* were then entered at step 3, again yielding a non-significant increment in variance accounted for [Adjusted $R^2 = .33$, $F(3, 33) = 2.54$, $p > .05$]. Hence, among these variables, prediction is not enhanced after first taking into account sufficiency and necessity.

A second sequential regression analysis was performed entering control at step 1, accounting for approximately 21% of the variance in change scores [Adjusted $R^2 = .21$, $F(1, 38) = 11.08$, $p < .05$]. Adding sufficiency and necessity at step 2 did not add to the prediction after control was taken into account [Adjusted $R^2 = .24$, $F(2, 36) = 1.86$, $p >$

.05]. It thus appears that necessity/sufficiency and control explain redundant portions of the variance in change scores.

What can we conclude from these analyses concerning perceived behaviour change? It appears that a number of factors contribute to the prediction of behaviour change for conditional inducements and advice. Thus, the degree of speaker's control over the outcome predicted how effective a statement is perceived to be in changing the addressee's behaviour. However, our analyses further established that conditional relationships that are perceived to be both necessary and sufficient are also perceived to be more effective than are non-necessary, non-sufficient relations. Moreover, degree of control accounted for the same variance in predicting behaviour change as perceived necessity and sufficiency. The remaining variables (i.e., probability of *q*, obligation, stake) did not contribute to the prediction of change scores after taking into account necessity/sufficiency and control.

We therefore postulate that (a) the speaker's control over the outcome *q* determines the degree to which the conditional is interpreted to express a necessary and sufficient relationship, and (b) the perceived probability that the addressee's behaviour will change varies according to whether the statement is perceived to be necessary and sufficient. In other words, we hypothesize that a statement will be perceived to be effective in changing behaviour when the behaviour described in the antecedent is sufficient to bring about the consequent, and when the behaviour described in the antecedent is the only way to realise the outcome described in the consequent (i.e., is necessary to bring about the consequent). Moreover, both of these conditions are more likely to hold when the speaker is perceived to have control over the consequent, that is,

when the speaker is perceived to have the power to bring about or to withhold the consequent, depending on the behaviour of the addressee.

## Experiment 2

The goal of Experiment 2 was to determine how well our six interpretive variables predict inference patterns on a traditional conditional inference task. From previous research, we expected that the willingness to accept the MP and MT inferences would depend on the perceived sufficiency of the conditional, whereas the DA and AC inferences would be highly correlated with necessity ratings (e.g., Cummins et al., 1991; Thompson, 1994, 1995). As in Experiment 1, we were interested in whether or not speaker's control would predict unique variance after the effects of sufficiency and necessity are accounted for, and what, if any, role our remaining interpretive variables would play in predicting conditional inferences.

### *Method*

*Participants*. A total of 32 undergraduate students (mean age: 24.4) from the University of Saskatchewan participated in partial fulfillment of a requirement for an introductory psychology course. None had previously completed a course in formal logic.

*Materials*. Of the 40 conditional statements used in Experiment 1, 20 were selected for this experiment. The selection was based on two criteria. First, seven independent raters categorised the 40 conditionals from Experiment 1 according to which type of statement they best represented (tip, warning, promise, threat, or don't know). Only statements that received a high level of category agreement were retained. Second, the ratings for each statement on the variables in Experiment 1 were compared to the mean values for its type of conditional. If a rating was deviating considerably from the

mean on one or more variables, this statement was a candidate for removal. For example, the warning "If you wander away from me, you will get lost", uttered by a mother to her daughter in a shopping mall, was rated high on speaker's control and stake. This statement was therefore not a good item for the warning category, and was not retained. After these selection procedures, 20 statements were included in Experiment 2, with five statements from each of the four types.[4]

The statements were again embedded in a social context. Each statement was then followed by the four inferences: Modus Ponens (MP), Denying the Antecedent (DA), Affirming the Consequent (AC), and Modus Tollens (MT). As an example, the inferences for the tip "If you show up early for work, you will impress your boss" were as follows:

Suppose Paula shows up early for work. Can we conclude that she will impress her boss? (MP)

Suppose Paula doesn't show up early for work. Can we conclude that she won't impress her boss? (DA)

Suppose that the boss was impressed by Paula. Can we conclude that she showed up early for work? (AC)

Suppose that the boss was not impressed by Paula. Can we conclude that she did not show up early for work? (MT)

*Procedure*. The participants were either tested individually or in small groups of 3 to 6 people. They were asked to read each of the 20 statements and answer the questions based on the inferences that followed logically from the information

[4] We performed the analyses of Experiment 1 again with only the 20 statements selected for Experiment 2. Overall, the results were very similar, with only the following exceptions. The differences between inducements and advice were larger for most variables, although inducements were no longer rated higher than advice on the probability of *q*. The outcome of the correlational and regression analyses was identical, save that some of the coefficients were higher.

provided. They were to answer "yes" if the conclusion necessarily followed, "no" if it did not follow, and "maybe" if it may or may not have followed.

The statements were presented in a booklet with four statements to a page. A computer program was used to generate the booklets, such that the order of the statements was randomized for each participant, and the order of the four inferences was randomized for each statement. Participants were told to work at their own pace, and most sessions were completed in about 20 - 30 minutes.

*Results and Discussion*

One participant only answered one question for each conditional statement, and was dropped from the analysis. An inference was scored as accepted if the participant responded "yes" to whether the conclusion followed. The percentage of accepted inferences for each of the four statement types is presented in Table 3 (standard deviations in parentheses).

*Rates of acceptance*. For the first set of analyses, we computed 2x2 within-subject ANOVAs on each inference type, with class of statement (inducement vs. advice) and valence of statement (positive vs. negative) as factors. Only effects significant at $p < .05$ are reported.

All four inferences were accepted more frequently for inducements than for advice, smallest $F(1, 30) = 29.65$. Only for MP and MT was the main effect of valence significant, smallest $F(1, 30) = 9.98$, whereby negative statements were more frequently endorsed than positive statements. However, for DA and AC there was a significant interaction, smallest $F(1, 30) = 13.54$. For these inferences, the difference in endorsement rates between inducements and advice was greater for positive statements than negative

Table 3

*Percentages (standard deviations) of accepted inferences in Experiment 2 as a function of inference type and statement type*

| | Inferences | | | | |
|---|---|---|---|---|---|
| Statement type | MP | DA | AC | MT | Mean |
| *Advice* | | | | | |
| Tip | 34 | 22 | 23 | 19 | 24 |
| | (38) | (28) | (32) | (30) | |
| Warning | 45 | 34 | 40 | 36 | 39 |
| | (38) | (25) | (31) | (34) | |
| Mean | 39 | 28 | 32 | 27 | 32 |
| *Inducements* | | | | | |
| Promise | 60 | 60 | 67 | 45 | 58 |
| | (35) | (28) | (34) | (33) | |
| Threat | 63 | 42 | 58 | 52 | 54 |
| | (35) | (33) | (35) | (36) | |
| Mean | 61 | 51 | 63 | 48 | 56 |
| Overall | 50 | 40 | 47 | 38 | |

statements.

Hence, we replicated the general pattern of results reported by Newstead et al. (1997). Specifically, participants endorsed more inferences of all kinds for conditional inducements than for conditional advice. In addition, for two inferences (i.e., MP and MT) we replicated the tendency for participants to endorse more inferences for negative statements than for positive statements. However, the rates of acceptance were overall lower in our study (an average of 44% collapsed across all inference types) than those obtained by Newstead et al. (an average of 66%). Two explanations are offered here as to why the overall endorsement rates were lower in the present study. First, Newstead et al. asked participants to make inferences about past behaviour (e.g., "Billy turned up the radio, therefore, his mother smacked him"). In contrast, in the present paper, the MP and DA inferences asked about future behaviour (e.g., "Can we conclude that the boss will be impressed?"). The focus on future behaviour may have introduced uncertainty and resulted in an overall suppression of inferences (e.g., Stevenson & Over, 1995). Second, Newstead et al. only used two response alternatives (i.e., "Does follow" and "Does not follow"), whereas three options were available for participants in this study (yes, no, and maybe). Hence, some participants who would endorse an inference with a two-response format may choose an indeterminate answer if it is available (e.g., George, 1997).

These findings are, nevertheless, consistent with the conclusions of Newstead et al. (1997) and others (e.g., Evans, 2002; Stanovich, 1999) that people recruit a substantial amount of information when asked to make seemingly simple inferences. Whereas acceptance of Modus Ponens can be close to 100% with abstract material (e.g., Evans, Newstead, & Byrne, 1993), the acceptance rate is much lower with real-life statements

(in our study, the rate was around 50% overall). In a tip, for instance, the relatively weak link between antecedent and consequent compels many people to reject the logically valid Modus Ponens inference, even with instructions to base their answers on what follows logically. Endorsement rates for all four inferences were higher for inducements than advice, suggesting that reasoners recruit knowledge about inducements that predisposes them to give a biconditional interpretation. Our next analyses tested the hypothesis that this knowledge is tied to some of the variables we examined in Experiment 1, such as knowledge about sufficiency and necessity as well as the perceived degree of speaker's control.

*Relationship between variables and inferences*. Table 4 shows the correlations between acceptance rates for the four inferences and our interpretive variables from Experiment 1. Note that these correlations are now based on the 20 statements selected for Experiment 2, and not the initial pool of 40 statements.

As seen from Table 4, the DA and AC inferences are highly correlated with necessity ratings. This finding is consistent with previous research documenting that the probability of endorsing DA and AC depends on the perceived necessity of the conditional (e.g., Cummins et al., 1991; Thompson, 1995). These studies also indicated that the probability of making the MP and MT inferences depends on the perceived sufficiency and not the perceived necessity of a conditional rule; in our data, however, MP and MT were correlated with both sufficiency and necessity. This finding is probably due to the fact that ratings of necessity and sufficiency were themselves correlated for this set of statements, allowing necessity to be correlated with accepting the MP and MT inferences. The table also reveals that the rates of acceptance for all four inferences were

Table 4

*Correlations between variables from Experiment 1 and inference rates in Experiment 2*

| Inferences | Prob(*q*) | Control | Stake | Obligation | Sufficiency | Necessity | Change |
|---|---|---|---|---|---|---|---|
| MP | .16 | .75** | .30 | .48* | .60** | .43* | .38* |
| DA | -.32 | .48* | .33 | .68** | .14 | .85** | .26 |
| AC | -.13 | .72** | .51* | .77** | .44* | .80** | .39* |
| MT | -.02 | .68** | .45* | .43* | .39* | .60** | .25 |

Note. * $p < .05$; ** $p < .01$. All tests are one-tailed.

highly correlated with control and obligation, whereas the correlations with the "stake" variable were lower. This finding confirms the trend from Experiment 1, which showed that the correlations between stake and the other variables were lower overall than for control and obligation. Finally, the likelihood of accepting these inferences was not strongly correlated with either the probability of *q* or the behaviour change variable.

Evans and Twyman-Musgrove's (1998) findings suggested that the degree of speaker's control is an important reason for the higher inference rates for inducements than advice. To test the hypothesis that the speaker's perceived control predicts unique variance in inference patterns, we conducted several regression analyses. In one set of analyses, rate of acceptance for the MP and MT inferences was the dependent variable, and perceived sufficiency, control, stake, obligation, and probability of *q* were entered as predictor variables. A comparable set of analyses explored acceptance rates for the AC and DA inferences, except that perceived necessity was entered in the regression equation in lieu of perceived sufficiency. As was the case in Experiment 1, each set of analyses proceeded in a sequential fashion, with necessity or sufficiency entered in one block and the remaining variables entered in another block.

For the MP and MT analyses, perceived sufficiency accounted for a significant portion of the variance for MP but not MT when entered in the first block [Adjusted $R^2$ = .32, $F(1, 18) = 9.91$, $p < .05$ and $R^2 = .10$, $F(1, 18) = 3.21$, $p > .05$, respectively]. When the remaining variables were entered, only control uniquely predicted residual variance [Adjusted $R^2 = .56$, $F(1, 17) = 10.69$, $p < .05$ and $R^2 = .40$, $F(1, 17) = 10.06$, $p < .05$, for MP and MT, respectively]. Hence, for both MP and MT, the speaker's perceived control predicted variance after accounting for the effects of perceived sufficiency, but the

remaining variables did not. In contrast, when control was entered first, perceived sufficiency did not explain any of the residual variance in either inference, $F(1,17) \leq 1.79$, $p > .05$.

The story for the DA and AC inferences was a little more complex. When perceived necessity was entered into the equation at step 1, it accounted for approximately 71% of the variance in the DA inference [Adjusted $R^2 = .71$, $F(1, 18) = 46.86$, $p < .05$] and 61% of the variance in the AC inference [Adjusted $R^2 = .61$, $F(1, 18) = 30.82$, $p < .05$]. For the DA inference, both probability of $q$ [Adjusted $R^2 = .80$, $F(1, 17) = 9.10$, $p < .05$] and obligation [Adjusted $R^2 = .83$, $F(1, 16) = 4.53$, $p < .05$] predicted a significant portion of the residual variance, whereas control did not. In contrast, for the AC inference, the speaker's perceived control did predict a significant portion of the residual variance [Adjusted $R^2 = .77$, $F(1, 17) = 13.58$, $p < .05$] while the other variables (probability of $q$, stake, and obligation) did not.

For both the AC and DA inferences, perceived necessity predicted a significant amount of variance, even after the control variable was partialled out. For the DA inference, control accounted for 18% of the variance when entered on the first step [Adjusted $R^2 = .18$, $F(1, 18) = 5.24$, $p < .05$], and perceived necessity predicted an additional 52% of the variance [Adjusted $R^2 = .70$, $F(1, 17) = 32.26$, $p < .05$]. For the AC inference, control accounted for 49% of the variance when entered first [Adjusted $R^2 = .49$, $F(1, 18) = 19.42$, $p < .05$] and necessity added another 28% of the variance when entered at step 2 [Adjusted $R^2 = .77$, $F(1, 17) = 22.89$, $p < .05$].

These findings corroborate the claims of Newstead et al. (1997) and Evans and Twyman-Musgrove (1998), and support the hypothesis that the speaker's perceived

degree of control mediates inferences about conditional inducements and advice. For three of four inferences, the speaker's perceived control predicted unique variance, even after the effects of necessity and sufficiency had been accounted for. It also appears that speaker's control is not redundant with perceived necessity, in that both variables accounted for a unique portion of the variance in the AC and DA inferences. In contrast, perceived sufficiency was essentially redundant with perceived control, suggesting that these may be overlapping constructs. Finally, although obligation and probability of *q* predicted unique variance on the DA inference, it appears that these variables (and also stake) are not as important as the speaker's degree of control in predicting inference patterns.

These data can be interpreted within the framework proposed by Thompson (2000). She argued that some conditional relations, namely deontic conditionals, are routinely interpreted on two levels. On one level, deontic conditionals describe an outcome that ought to happen; on the other hand, there is an implicit understanding that this outcome may not happen. For example, the conditional "If a person is drinking alcohol, she must be over 18 years old" describes a state of affairs that ought to hold, but that we know from experience may not hold. Thompson proposed that this interpretation leads people to form a model of the situation that includes the possibility of rule violations; and this representation, in turn, reduces the acceptability of certain inferences (e.g., Someone is drinking alcohol. Is he/she over 18 years old?).

We propose that advice and inducements may be interpreted in a similar manner, in that they describe a state of affairs that may or may not take place. For example, if one shows up early for work, one ought to impress the boss, although there is no guarantee

that this eventuality will indeed occur. Thus, the representation of these conditionals may include the possibility that the outcome described by the speaker may or may not actually occur. Moreover, the speaker's perceived degree of control over the outcome may be a critical factor in determining the perceived probability that the outcome does not occur, and as such, may determine the probability that counter-examples or rule violations are represented in reasoners' models of the situation.

However, our findings also suggest that reasoners may draw on other sources than speaker's perceived control to generate potential counter-examples to conditional inducements and advice. That is, much research has demonstrated that conditionals that are perceived to express necessary relationships are also those for which few counter-examples of the form *~p and q* are represented; the availability of these counter-examples further mediates the acceptability of the DA and AC inferences (e.g., Byrne, Espino, & Santamaria, 1999; Cummins et al., 1991; Markovits, 1984; Thompson, 1994, 1995, 2000). The finding that perceived necessity explained a non-redundant portion of the variance in the AC and DA inferences after taking into account the effects of speaker's control suggests that several sources may be used to generate counter-examples. One such source could be suspicions that a speaker of a threat may be unwilling to enforce the punishment, perhaps because of a reluctance to provoke conflict.

## General Discussion (Experiments 1 and 2)

People recruit a substantial amount of background knowledge when interpreting and reasoning with conditional inducements (i.e., promises and threats) and conditional advice (i.e., tips and warnings). We found that relative to advice, inducements suggest that the speaker has greater control over the consequent and more at stake in the outcome

of the events, and express a greater degree of obligation on behalf of the speaker. In addition, inducements have a stronger link between *p* and *q*, are more necessary and more sufficient, are perceived to be more effective in changing the addressee's behaviour, and elicit more inferences of all types on a conditional arguments task.

Of course, confidence in these conclusions hinges on the degree of confidence one has that the exemplars used to define our categories generalise to the larger population of inducement and advice conditionals. As is the case with much of the research in which performance across two or more categories of conditionals is compared (e.g., Cheng & Holyoak, 1985; Cosmides, 1989; Evans & Twyman-Musgrove, 1998; Gigerenzer & Hug, 1992; Liu, Lo, & Wu, 1996; Markovits & Savary, 1992; Newstead et al., 1997; Politzer & Bourmaud, 2002; Politzer & Nguyen-Xuan, 1992; Thompson, 1995, 2000), we were obligated to use different statements in each of our categories, allowing for the possibility that the differences that we observed may reflect item-specific differences rather than generalisable differences inherent to the categories that we have investigated. Ideally, to reduce the possibility that the differences observed reflect item-specific variation, one would like to control as many dimensions as possible when assigning statements to different categories. One contribution of our paper is to identify some of the dimensions that are relevant to the interpretation of advice and inducement conditionals, allowing subsequent researchers to exercise better control over item variability.

Despite the potential for item-specific effects, confidence in the generalisability of our findings is bolstered by the degree of convergence between our study and past research (e.g., Beller, 2002; Evans & Twyman-Musgrove, 1998; Fillenbaum, 1975, 1976,

1978; Newstead et al., 1997; Thompson, 2000). Hence, despite the fact that we developed an entirely novel item pool for this set of studies, we replicated broad inference patterns with advice and inducement conditionals. Also consistent with previous research, we have identified degree of obligation, speaker's control, and perceived necessity/sufficiency as important variables in the interpretation of these conditionals.

From our analyses, three of these variables (i.e., necessity, sufficiency, and speaker's control) emerged as important predictors of our outcome measures. In terms of behaviour change, these variables were largely redundant, leading us to hypothesise that the degree to which a speaker is perceived to have control over the situation determines a reasoner's interpretation of the relationship as necessary and/or sufficient, which in turn determines the degree to which the statement is perceived to be effective in changing behaviour. In terms of inference patterns, perceived necessity predicted variance over and above speaker's control, suggesting the presence of other variables (presumably in the form of counter-examples) that contributed to perceptions of necessity, and which mediate performance on this task.

*Behaviour change*

Beller (2002) argued that an analysis of conditional inducements requires an approach that goes beyond the traditional deductive paradigm, as this paradigm is ill equipped to reveal the complexities of such real-life statements. Hence, the essence of an inducement is not captured by the inferences people are willing to draw, or the logical possibilities perceived to be consistent with the statement. Rather, inducements must be understood by reference to their perlocutionary effects (e.g., Fillenbaum, 1986). Our findings demonstrate the benefits of adopting an approach that extends beyond the

questions usually asked in reasoning studies, as we show that participants are sensitive to a difference in behavioural effectiveness between conditional inducements and advice. Furthermore, participants' willingness to endorse inferences was not highly correlated with a statement's perceived effectiveness, suggesting that these two measures are capturing different aspects of the meaning of conditional inducements and advice.

How strong is the relationship between perceived behavioural effectiveness and actual behaviour change? Research from social psychology has documented that intentions to perform behaviours can predict actual behaviour with considerable accuracy (e.g., Ajzen, 1991). We are therefore confident that our measure of perceived effectiveness is a good substitute for actual behaviour change. The social psychology literature also emphasises the importance of perceived behavioural control in predicting both behavioural intentions and behaviour itself; our results concerning the degree of speaker's control provide an interesting convergence between these two lines of inquiry.

*Conditionals and conditional probability*

Recently, a number of authors have raised doubts about the appropriateness of formal logic as a framework for deductive reasoning, and instead favour a probabilistic approach (e.g., Evans, 2002; Evans, Handley, & Over, 2003; George, 1995, 1997; Liu et al., 1996; Oaksford & Chater, 2001, 2003; Oaksford, Chater, & Larkin, 2000; Oberauer & Wilhelm, 2003; Over & Evans, 2003; Politzer & Bourmaud, 2002; Stevenson & Over, 1995). Applied to conditional reasoning, this probabilistic approach proposes that the statement *if p then q* is represented as the conditional probability of *q* given *p*. Similarly, Newstead et al. (1997) suggested that a reason why inducements and advice are interpreted differently is that the probability of *q* given *p* is greater for inducements than

for advice. The findings of Experiment 1 support this hypothesis, in that scores on our measure of the probability of *q* were higher for inducements than advice. However, this variable was not as highly correlated with behaviour change as were some of the other variables, and overall, was not an important predictor of inference patterns in Experiment 2.

However, our findings are not necessarily inconsistent with the claim that conditional statements are represented as the conditional probability of *q* given *p*. This approach has hitherto mostly concerned indicative conditionals, which describe matters of fact. In contrast, deontic conditionals are not limited to a description of events; indeed, deontic statements can be true even when the probability of *q* given *p* is low. For example, a deontic rule such as "If you are within city limits, you should drive less than 50km/hr" can be true, even when compliance with the rule (and thus the probability of *q* given *p*) is low.

In addition, our measurement of conditional probability (i.e., asking how likely *q* is given *p*) was a relatively crude means to assess this probability. Other researchers have used much more sophisticated procedures, such as specifying the relative frequency of all logical possibilities (Evans et al., 2003; Oberauer & Wilhelm, 2003), or asking participants to provide probability estimates to these possibilities (Over & Evans, 2003). Finally, previous research has primarily focussed on establishing the conditions under which a conditional is perceived to be true, whereas we were mainly interested in the conditions under which a conditional is perceived to be effective. It is therefore possible that the probability of *q* given *p* may be important for one aspect of these conditionals but not the other.

*Relation to theories of conditional reasoning*

A number of theoretical approaches have been proposed to explain how people reason with conditional statements, and offer explanations for the content-based variability that is typical when reasoning with familiar materials. Although the present paper did not aim to test or compare different theories of conditional reasoning, our findings do contribute to the development of these models. In particular, many theoretical approaches lack an interpretive component sufficiently detailed to predict a priori the effects of the pragmatic and semantic factors that we have discussed here. One contribution of our paper, therefore, is to document some of the variables that need to be integrated into the theories in order to achieve a complete account of how people reason with information-rich statements such as inducements and advice.

For example, the mental models theory (e.g., Johnson-Laird & Byrne, 1991, 2002) proposes that reasoners construct mental models of the information conveyed by the premises. Semantic and pragmatic modulation can affect the construction of models, resulting in up to 10 different representations of factual and deontic conditionals. Although the mental logic theory proposes a different underlying mechanism in the form of inference rules (e.g., Braine & O'Brien, 1991), this theory also makes reference to pragmatic principles, such as conversational implicatures (Grice, 1975) and invited inferences (Geis & Zwicky, 1971). Our findings suggest that the speaker's degree of control over the situation needs to be added to the list of variables, such as perceived necessity and sufficiency (e.g., Thompson, 2000) and availability of counter-examples (e.g., Byrne et al., 1999), that contribute in a substantive manner to the interpretation and representation of conditional statements.

Whereas both the mental models and natural logic theories posit that the process of interpretation is separate from the underlying inferential mechanisms, others have made proposals in which these processes are more closely integrated. According to pragmatic schema theory (Cheng & Holyoak, 1985; Cheng, Holyoak, Nisbett, & Oliver, 1986), people reason using abstract knowledge structures that define broad classes of situations, such as permissions and obligations. In this view, one might propose that reasoners recruit similar schemas to reason with situations defined as advice and inducements (see Markovits & Lesage, 1990). A similar approach was adopted by Cosmides (1989), who argued that humans possess Darwinian algorithms specialized for reasoning about adaptive problem domains. Her social exchange scenarios (i.e., If you take a benefit, then you pay a cost) are similar to our conditional promises, while our conditional warnings overlap with the precaution rules specified by hazard management theory (e.g., Fiddick, Cosmides, & Tooby, 2000).

Although these two approaches seem to offer much promise in explaining our findings, they too suffer from the fact that the interpretive component of these theories is under-specified (Thompson, 1995). That is, they do not explain the circumstances under which additional information may be added to, deleted from, or overlooked in the problem space, and thereby modify the output of the pragmatic schema or Darwinian algorithm. For example, there is often as much variability within a schematic category as between categories (Thompson, 1994), a finding that demands the addition of an interpretive component similar to that proposed for the mental models and natural logic theories.

Finally, Manktelow and Over (e.g., 1991, 1995) argue that an analysis of subjective expected utility (SEU) is required for a full understanding of deontic conditionals, whereby reasoning is explained by reference to goals and subjective preferences. To see how this framework can be applied to inducements and advice, consider the promise "If you wash the car ($p$), I will give you \$10 ($q$)". In this view, reasoning with deontic conditionals is mediated, at least in part, by the reasoner's SEU of a particular set of events. Thus, in our example, we must assume that the SEU($p$) is higher for the speaker than SEU($\sim p$), because she is willing to pay a price to obtain $p$. Similarly, for the promise to be effective, the addressee must attach a higher SEU($p$ & $q$) than to SEU($\sim p$ & $\sim q$); that is, the addressee must attach a higher SEU to washing the car and receiving the money than doing neither. In order to integrate our findings with this view, we would need to make assumptions similar to those made with the other theories that we have described, and assume that variables such as necessity, sufficiency, and speaker's control can influence a reasoner's interpretation of SEU. For instance, if the speaker of a promise is perceived to have limited control over the consequent, the SEU($p$) for the addressee might be lower than SEU($\sim p$), as the fulfillment of $p$ may no longer result in the reward $q$.

*Possible extensions*

Because of the social interaction inherent in uttering inducements and advice, there may be a large number of other variables that are relevant to interpreting these statements. Indeed, the regression analyses from Experiment 1 showed that our variables account for a relatively small portion of the variance in behavioural effectiveness, suggesting that there are other, as yet unidentified, dimensions crucial to an

understanding of inducements and advice. For example, the speaker of a promise may be perceived to have a high degree of control over an outcome, but this knowledge may be of little value if the listener knows or suspects that the speaker has no intention of keeping their promise. Thus, knowledge regarding the speaker's veracity and/or intentions may mitigate the effects of other variables. For instance, a reliable speaker may lead reasoners to perceive a stronger link between *p* and *q*, and thus increase the rate at which one is willing to endorse inferences from conditional advice. Likewise, endorsement rates should decrease if a speaker of an inducement is seen as unreliable.

Similar issues arise concerning the degree to which the speaker is seen as knowledgeable or credible. A credible speaker giving advice, for instance, may persuade the listener that *q* is certain given the occurrence of *p*. Our expectation that medical doctors are highly knowledgeable about medical matters, for example, may cause their advice to sound more like an expression of what will almost certainly happen (i.e., a promise) than is the case for typical advice. In this scenario, credibility may serve to counteract the effects of speaker's control that would ordinarily make endorsement rates for advice low. Indeed, in a recent study, Stevenson and Over (2001) found that participants are more likely to accept the MP and MT inferences when premises are uttered by experts rather than novices. Thus, variables such as credibility or veracity may interact with variables such as perceived control, and serve to augment or diminish interpretations derived from other sources.

*Conclusions*

Collectively, our findings, together with much recent work, underscore the importance of understanding interpretive processes in reasoning. Many theories of

reasoning emphasise the inferential component of the reasoning process, and focus on describing how inferences are computed given a certain interpretation of the premises. Less emphasis is typically paid to the variables that determine the form of this representation, and which thereby mediate the output of the inferential mechanisms. It is clear that much of the variability in reasoning performance is left unexplained by this approach. Indeed, the processes that contextualise utterances, link them to past knowledge, assess truth, and discern intentions, appear to constitute a large part of the processes that we normally describe as reasoning behaviour.

# EXPERIMENT 3

## CONDITIONAL PROBABILITY AND PRAGMATIC CONDITIONALS: DISSOCIATING TRUTH AND EFFECTIVENESS

Much evidence has accumulated to support the view that informal everyday reasoning is uncertain or probabilistic in nature (e.g., Evans, 2002; Liu, Lo, & Wu, 1996; Oaksford & Chater, 2001, 2003; Politzer & Bourmaud, 2002). For example, despite formal instructions to assume the truth of given premises and make binary decisions about the validity of conclusions, participants view premises as uncertain and prefer to express degrees of confidence in the inferences they make (e.g., Cummins, 1995; Cummins, Lubart, Alksnis, & Rist, 1991; De Neys, Schaeken, & d'Ydewalle, 2003; George, 1995, 1997; Stevenson & Over, 1995, 2001). Most of the evidence supporting a probabilistic account of reasoning has, however, been obtained indirectly, via the use of the traditional deductive paradigm (or minor modifications thereof). Inferences about informal processes are thus derived from performance on essentially formal tasks, in which participants are presented with a set of premises and asked to indicate what follows from them (Evans & Thompson, 2004). Recently, however, researchers have begun to move away from formal paradigms, and to develop tasks that allow a more direct study of informal reasoning processes (e.g., Beller, 2002; Byrne & Egan, 2004; Bonnefon & Hilton, 2004; Feeney, Scrafton, Duckworth, & Handley, 2004; Fiddick, 2004; Fugelsang, Stein, Green, & Dunbar, 2004; Oaksford & Hahn, 2004; Ohm & Thompson, 2004; Over, Manktelow, & Hadjichristidis, 2004; Thompson & Byrne, 2002; Thompson, Evans, & Handley, 2005; van Gelder, Bissett, & Cumming, 2004; Verbrugge, Dieussaert, Schaeken, & van Belle, 2004).

In this paper, we focus on the domain of conditional reasoning. Conditionals of the form "if *p*, then *q*" have traditionally been given a truth-functional interpretation, equivalent to material implication (Edgington, 1995). The material conditional is false only when the antecedent, *p*, is true and the consequent, *q*, is false (i.e., *p*~*q*), and is otherwise true. Although much experimental work in the psychology of reasoning has been conducted within this framework, there are good reasons to doubt whether material implication provides an adequate psychological account of the ordinary conditional (e.g., Evans & Over, 2004; Oaksford & Chater, 2003; Oberauer & Wilhelm, 2003; Over & Evans, 2003). One problem is that this interpretation leads to certain counterintuitive conclusions or paradoxes; for instance, one can infer "if *p* then *q*" from ~*p* (i.e., from the fact that it is not raining, it follows that if it rains, then the streets are dry). Regardless of the logical validity of this inference, ordinary reasoners are unlikely to accept it as valid. Instead, research has shown that reasoners typically judge cases where the antecedent is false (i.e., ~*pq* and ~*p*~*q*) to be irrelevant to, rather than consistent with, the truth of the conditional rule (e.g., Evans, Newstead, & Byrne, 1993).

Evans, Handley, and Over (2003) have recently developed a suppositional theory of conditionals whose goal is to circumvent the logical paradoxes of material implication, and which naturally captures the probabilistic and uncertain nature of people's everyday reasoning. Borrowing from a tradition in philosophical logic, they propose that conditionals are evaluated with a degree of belief equal to the subjective conditional probability of *q* given *p* (see also Adams, 1965, 1975; Edgington, 1995, 2003; Evans & Over, 2004; Liu et al., 1996; Oaksford & Chater, 2001, 2003; Oaksford, Chater, & Larkin, 2000; Oberauer & Wilhelm, 2003; Over & Evans, 2003). According to this

*conditional probability hypothesis*, "*if*" triggers a process of hypothetical thinking whereby people make the antecedent *p* a hypothetical supposition and evaluate the likelihood of the consequent *q* in that context. Specifically, when given a conditional such as "If it rains, then you get wet", people apply the so-called Ramsey test by first supposing that it is raining, and then comparing the likelihood of getting wet (i.e., *pq*) to the likelihood of remaining dry (i.e., *p*~*q*). The probability that the conditional is true will be high when *pq* is judged to be more likely than *p*~*q*, that is, when the conditional probability of *q* given *p*, P(*q*/*p*), is high. One consequence of this interpretation is that people will only attend to instances where *p* holds when judging the probability of conditionals; cases where *p* is false are seen as irrelevant.

Several recent studies have provided support for the conditional probability hypothesis, using both abstract and familiar materials (Evans et al., 2003; Evans & Over, 2004; Oberauer & Wilhelm, 2003; Over & Evans, 2003). For example, Evans et al. provided explicit frequency information about abstract conditionals describing cards of different shapes and colours. Specifically, they varied the relative frequency of the four possible combinations of events (i.e., *pq*, *p*~*q*, ~*pq*, and ~*p*~*q*). Participants were then asked to judge the probability that a conditional statement was true for a card selected at random; these judgements were subsequently compared to probability estimates computed on the basis of the frequency information.

Evans et al. (2003) investigated three alternative hypotheses. The *material conditional hypothesis* states that the conditional is true in three of four possibilities (i.e., *pq*, ~*pq*, and ~*p*~*q*); accordingly, the probability that the conditional is true should be judged as 1 – *p*~*q*. The conditional probability hypothesis, in contrast, predicts that

people will equate the probability of conditionals with P($q$/$p$), computed as $pq$/($pq$ + $p\sim q$). Finally, Evans et al. speculated that some participants will cut short the process of estimating the conditional probability, and instead use the conjunction of $p$ and $q$ as the basis for their probability judgements. Accordingly, the *conjunctive probability hypothesis* predicts that people will assign P($p$&$q$) to the probability of conditionals. In a series of experiments, Evans et al. found that participants' probability judgements were better correlated with both the conditional and conjunctive probability than with the probability of the material conditional (for similar results, see Oberauer & Wilhelm, 2003).

This approach has also recently been extended to an analysis of thematic conditionals. For example, Over and Evans (2003) asked participants to judge the probability of realistic conditionals concerned with the prediction of future events (e.g., "If global warming continues, then London will be flooded"). However, instead of providing explicit frequency information as in Evans et al. (2003), they subsequently asked participants to provide relative probability estimates (summing to 100%) to each of the four truth-table combinations for these conditionals. On the basis of these truth-table estimates, Over and Evans computed probability estimates corresponding to the material conditional, P($q$/$p$), and P($p$&$q$). Replicating the results found with abstract conditionals, P($q$/$p$) was again shown to be a better predictor of truth ratings than was the probability of the material conditional. Finally, additional work suggests that the influence of conjunctive probability is also reduced for thematic conditionals (Evans & Over, 2004).

In the present paper, our aim was to extend the conditional probability approach to pragmatically rich conditionals in the form of inducements and advice. Conditional

inducements (i.e., promises and threats) are speech acts whose goal is to persuade a person to do, or refrain from doing, some action (e.g., Fillenbaum, 1986). For instance, the promise "If you wash the car, I will give you $10" encourages an action *p* by offering *q* as a reward. Conditional advice (i.e., tips and warnings), on the other hand, involves a recommendation or prediction about future behaviour, as in the tip "If you study harder, your grades will improve". Previous research has found that conditional inducements elicit more inferences than conditional advice, and are also perceived to be more effective in changing the behaviour described in the antecedent (e.g., Beller, 2002; Evans & Twyman-Musgrove, 1998; Fillenbaum, 1978, 1986; Newstead, Ellis, Evans, & Dennis, 1997; Ohm & Thompson, 2004).

Under the conditional probability hypothesis, the probability of conditional inducements and advice should be evaluated as the conditional probability of *q* given *p*. Thus, there should be a close relationship between P(*q*/*p*) and the probability that inducements and advice are perceived to be true. However, because of their pragmatic richness, we also argue that this conditional probability only provides a partial description of the meaning of these conditionals.

Specifically, because inducements and advice both entail a desire to influence or change the behaviour of the addressee, a complete representation of these conditionals must account for their role as speech acts (e.g., Beller, 2002; Fillenbaum, 1986). In these conditionals, the speaker utters *q* in order to increase or decrease the likelihood that the addressee performs *p*. Consequently, these statements are unlikely to be effective speech acts if *q* is forthcoming regardless of whether *p* is performed (e.g., if the $10 is received regardless of whether the car is washed); if it were, then there would be no reason for the

addressee to modify her behaviour (e.g., Fillenbaum, 1978). Therefore, we argue that the relevance of ~*p* should be high for conditional inducements and advice, despite the fact that it appears irrelevant to evaluating the truth of conditionals (e.g., Evans et al., 1993; Evans et al., 2003). Specifically, we predict that conditional inducements and advice should be judged as effective in changing behaviour only when the conditional probability of *q* given ~*p*, or P(*q*/~*p*), is low (i.e., when *q* is relatively unlikely to occur in the absence of *p*).

In other words, we argue that when evaluating conditional statements, the topic about which hypothetical thinking is invited may not be fixed, but instead varies as a function of the pragmatic goals of the listener and speaker as well as the type of judgment people make about the conditional. Thus, when evaluating the truth of the conditional, the relevance of the *p* world may well be highlighted, such that the relevance of the ~*p* possibility is small. For example, the conditional "If it rains, you will get wet" should be considered to be true, regardless of the number of ~*p* scenarios that will also make us wet (e.g., if you jump in the lake, you will get wet; if you go swimming, you will get wet), and also regardless of whether the probability of it not raining (~*p*) vastly exceeds the probability of it raining (*p*).

In other contexts, however, the relevance of the ~*p* possibility may play a larger role, as for example, when evaluating the effectiveness of conditional promises and threats. Intrinsic to the meaning of these statements is the inference that if the behaviour defined in *p* is not performed, the promised reward (or threatened outcome) will not be forthcoming (e.g., Evans & Over, 2004; Fillenbaum, 1978, 1986; Geis & Zwicky, 1971). Thus, we propose that these scenarios invite the listener to engage in hypothetical

thinking about the ~*p* possibility, and to evaluate the perceived effectiveness of the statement as a function of the probability with which *q* is likely to occur in that possible world. In other words, when considering whether or not the conditional will be effective in motivating the listener to comply with the statement (i.e., by carrying out, or refraining from, action *p*), people should base their decision in large part on P(*q*/~*p*).

Note that our emphasis on P(*q*/~*p*) is also consistent with recent developments of the suppositional theory. For example, when discussing people's representation of conditional inducements and advice, Evans and Over (2004) argue that pragmatic implicature may have the effect of adding the inverse conditional "if ~*p*, then ~*q*"; people may therefore apply a secondary Ramsey test involving P(*q*/~*p*) when evaluating these conditionals. In addition, Evans and Over explore the role of P(*q*/~*p*) in their recent experiments on causal conditionals, since P(*q*/*p*) must be higher than P(*q*/~*p*) for a causal relationship to exist (i.e., the delta *p* rule). In these experiments, estimates of P(*q*/~*p*) were found to predict probability judgements, although this effect was much smaller than the influence of P(*q*/*p*). However, these discussions of P(*q*/~*p*) still focus on judging the probability that conditionals are true; we argue that a consideration of the ~*p* possibility is crucial to understanding another aspect of conditional inducements and advice, namely whether they are perceived to be effective in changing behaviour.

Intuitively, one might argue that the perceived effectiveness of these conditionals should be predicted by P(*q*/*p*) as well as P(*q*/~*p*). Specifically, a statement should not be very effective in changing behaviour if the listener perceives the chances of obtaining *q* upon the fulfillment of *p* to be low, as would be the case, for example, if doubts were raised about the credibility of a speaker's promise. Nevertheless, there are reasons to

argue that P($q$/$p$) may not be a strong predictor of effectiveness, at least under normal assertibility conditions. Following Grice's (1989) maxim of quality, conversational utterances are assumed to be truthful; thus, unless there is an explicit reason to doubt the credibility of the speaker, P($q$/$p$) will usually be high. Under these circumstances, other factors may be more important when determining the effectiveness of a statement (e.g., does the cost of undertaking $p$ outweigh the benefit of the reward $q$?).

Following Over and Evans' (2003) approach, participants in the present study were asked to judge the truth of conditional inducements and advice, as well as to provide estimates of the probability that each of the four logical combinations of $p$ and $q$ (i.e., $pq$, $p\sim q$, $\sim pq$, and $\sim p\sim q$) would occur. From these estimates, we computed P($q$/$p$) and P($q$/$\sim p$). To test our hypothesis regarding the effectiveness of these conditionals as speech acts, we used a measure introduced by Ohm and Thompson (2004) in which participants were asked to judge how likely the addressee is to perform $p$ both in the absence and presence of the speaker's utterance. The difference between these scores provides an estimate of the perceived effectiveness of the statements in either raising or lowering the likelihood that the behaviour described in $p$ occurs. We predicted that whereas the truth of conditional inducements and advice would vary as a function of P($q$/$p$), their perceived effectiveness would vary mainly as a function of P($q$/$\sim p$).

In addition to computing estimates of P($q$/$p$) and P($q$/$\sim p$) using the methodology introduced by Over and Evans (2003), we also asked participants to provide more direct estimates of these conditional probabilities. A direct estimate of P($q$/$p$) can be obtained by asking for the sufficiency of $p$ in bringing about $q$ (i.e., suppose the student works harder; how likely is it that his grades will improve?). Similarly, asking for the necessity

of *p* (i.e., suppose the student does *not* work harder; how likely is it that his grades will improve?) provides a direct estimate of P(*q*/~*p*). We expected that these two ways of estimating P(*q*/*p*) and P(*q*/~*p*) should be highly correlated, and give similar patterns of results regarding ratings of truth and behavioural effectiveness.

Finally, we wanted to extend the conditional probability approach to predicting reasoning performance. To this end, participants were also asked to solve a conditional inference task. In this task, participants are presented with a conditional statement followed by a categorical premise (i.e., *p*, ~*p*, *q*, or ~*q*) and asked to evaluate the validity of four conclusions (i.e., *q*, ~*q*, *p*, or ~*p*, respectively). Our predictions for this task were derived from the probabilistic model proposed by Oaksford et al. (2000). According to this model, the willingness to endorse conditional inferences can be predicted from the conditional probability of the conclusion given the categorical premise, such that inference rates should be higher for conclusions with greater conditional probabilities. Thus, the rate of acceptance for Modus Ponens (MP: "*p*, therefore *q*") should be closely related to P(*q*/*p*), while the rate of acceptance for Denying the Antecedent (DA: "~*p*, therefore ~*q*") should vary as a function of P(~*q*/~*p*). Similarly, the computed estimates of P(*p*/*q*) and P(~*p*/~*q*) should predict participants' willingness to endorse the Affirming the Consequent (AC: "*q*, therefore *p*") and Modus Tollens (MT: "~*q*, therefore ~*p*") inferences, respectively.

Oaksford et al. (2000) used three parameters [i.e., P(*p*), P(*q*), and P(*q*/*p*)] to generate mathematical expressions for these conditional probabilities, and found that inference rates varied as predicted as a function of the probabilities derived from these parameters. Our goal was to provide converging evidence for their model using a

different means to derive the conditional probabilities. Specifically, we computed these conditional probabilities from the participants' probability estimates of the four truth-table cases, expecting a close relationship between inference rates and conditional probabilities appropriate for each inference. [5]

In summary, we aim to extend the probabilistic model of conditionals in several ways. First, we extend the approach to account for pragmatically rich statements in the form of conditional inducements and advice. Based on the findings of previous studies (e.g., Evans et al., 2003; Evans & Over, 2004; Oberauer & Wilhelm, 2003; Over & Evans, 2003), we predicted that the truth of these conditionals would vary as a function of P(*q/p*). Second, we extend the analysis to account for an important pragmatic element of these statements, namely the degree to which they are considered to be effective in changing the listener's behaviour; this variable was predicted to correlate with estimates of P(*q/~p*). Finally, we aim to use the methodology introduced by Over and Evans (2003) to predict reasoning performance, expecting that acceptance rates on the conditional inference task should be correlated with relevant conditional probabilities computed on the basis of participants' truth-table estimates.

## Method

### *Participants*

One hundred and eighty-eight University of Saskatchewan undergraduate students participated in partial fulfillment of an introductory psychology course requirement.

[5] Oaksford and colleagues have recently used a similar method of calculating conditional probabilities from estimates of the four truth-table cases to successfully account for performance on another conditional reasoning task, namely the Wason selection task (e.g., Oaksford & Moussakowski, 2004; Oaksford & Wakefield, 2003).

*Materials*

We first created three scenarios from each of four categories: tips, promises, warnings, and threats. From these 12 scenarios, a total of 36 conditionals were constructed by creating three alternative consequents for each antecedent; the purpose of this manipulation was to obtain a wide range of probability estimates [i.e., statements varying in P($q$/$p$) and P($q$/~$p$)]. An example of each category, with three different levels of $q$, is provided below:

Tip: If you show up early for work, you will get fresh coffee/ impress your boss/ be promoted.

Promise: If you help me study, I will save you a seat in class/ buy you lunch/ buy your textbooks next term.

Warning: If you close the museum, people will be upset/ tourism will decrease/ people will move to another town.

Threat: If you come home after 11, I will take $5 off your allowance this week/ take away your allowance this week/ take away your allowance for a year.

A brief paragraph preceded each statement, providing the conditional with a social context. A complete list of the materials used in this experiment is provided in Appendix C. Participants performed three separate tasks with these statements:

*Behavioural effectiveness*. This measure was derived from the difference of two scores. First, participants were provided with the social context, but not the conditional statement, and asked to estimate (on a seven-point scale) the likelihood of $p$ occurring. For the above promise, for example, this question would read: "Paul is nervous about an

upcoming exam, as he feels he doesn't understand the material. He is hoping to get some help from his friend Julian who is a good student. How likely is Julian to help Paul study?" Participants were then given the conditional statement (which discloses, in this case, what Paul is willing to offer in return for Julian's help), and asked to provide a second estimate of the likelihood of *p*. A measure of the degree to which the conditional is perceived to be effective in changing behaviour can be derived from the difference between these two estimates.

*Conditional inference task*. For this task, participants evaluated the validity of four inferences. MP asks whether they can conclude *q*, given the occurrence of *p*, while DA asks whether they can conclude ~*q*, given the absence of *p*. Conversely, AC entails concluding *p* from the occurrence of *q*, while MT involves concluding ~*p* from the absence of *q*. For each question, participants were instructed to answer "yes", "no", or "maybe" based on whether the conclusions followed necessarily from the premises.

*Probability estimates*. Probability estimates were obtained in three ways. First, participants were asked to give direct estimates of P(*q*/*p*) and P(*q*/~*p*). For the above promise, for example, they were told to either suppose that Julian helps [i.e., for P(*q*/*p*)] or does not help [i.e., for P(*q*/~*p*)] Paul study; they then indicated (on a seven-point scale) how likely Paul is to deliver the reward (e.g., save Julian a seat in class). Subsequently, participants were asked to rate the probability (on a scale from 0 to 100) that the conditional statement itself was true. Finally, they provided probability estimates for the four logical possibilities of the truth table (as per Over & Evans, 2003). That is, they were asked how likely each of the four events (i.e., *pq*, *p*~*q*, ~*pq*, and ~*p*~*q*) were to occur, under the requirement that these four estimates add up to 100%. To illustrate, for the

promise “If you help me study, I will save you a seat in class” uttered by Paul to his friend Julian, the four cases corresponded to:

Julian helps Paul study and Paul saves him a seat in class

Julian helps Paul study and Paul does not save him a seat in class

Julian does not help Paul study and Paul saves him a seat in class

Julian does not help Paul study and Paul does not save him a seat in class

*Procedure*

Each participant performed these three tasks for six conditional statements, of which three were inducements and three were advice. Half of the participants were only given statements with positive consequents (i.e., tips and promises), while the other half were only provided with statements outlining negative outcomes (i.e., warnings and threats). Participants received two conditionals from each of the three different levels of *q*, with the constraint that only one version of a particular conditional was presented to a given participant. For each participant, the order of the six statements was randomized. Participants were provided with written instructions, and were tested in one large group. On average, we collected 31 responses per conditional statement (range = 30 - 33).

The questions for each conditional statement were printed on two pages. On the first page, participants performed the behavioural effectiveness task and provided direct estimates of P(*q*/*p*) and P(*q*/~*p*). On the second page, participants performed the inference task and the probability estimates task (i.e., providing probability estimates of the conditional and the four truth-table cases). The order of these two remaining tasks was counterbalanced within participants, such that the inference task was performed second for three of the six conditionals, and third for the remaining three statements. Finally, the

questions in the inference and probability estimates tasks were presented in two quasi-random orders, with each participant receiving half of the statements in each order.

## Results

### *Inducements vs. advice*

Two analyses were first conducted to verify that we replicated the findings that inducements are perceived to be more effective in changing behaviour than advice (Ohm & Thompson, 2004), and that inducements invite more inferences than advice (e.g., Evans & Twyman-Musgrove, 1998; Newstead et al., 1997; Ohm & Thompson, 2004). Behavioural effectiveness was computed as the difference between participants' two likelihood ratings of *p* occurring (i.e., in the absence and presence of *q*). As conditionals with negative outcomes (i.e., warnings and threats) involve attempts to decrease the probability of *p* occurring, scoring for these conditionals was reversed so that higher values reflected greater effectiveness. For the conditional inference task, an inference was scored as accepted if the participant answered "yes" to indicate that the conclusion followed. To abbreviate reporting, a mean inference score was computed by combining acceptance rates for all four inferences. Unless otherwise specified, alpha was in these and all subsequent analyses set at .05.

These data are presented in Table 5, and were analysed using 2x2 mixed design ANOVAs with type of conditional (inducements vs. advice) as a within-subjects factor and valence (positive vs. negative) as a between-subjects factor. In these analyses, the data were collapsed across the different levels of *q*.[6] As anticipated, inducements were perceived to be more effective in changing behaviour than advice, $F(1, 184) = 95.87$,

[6] For the interested reader, mean inference rates for each conditional statement are presented in Appendix D. This appendix also shows, for each conditional, the four computed conditional probabilities.

Table 5

*Ratings of behavioural effectiveness and inferences rates (in percentages) for each statement type in Experiment 3*

| | | Inferences | | | | |
|---|---|---|---|---|---|---|
| Statement type | Behavioural effectiveness | MP | DA | AC | MT | Mean |
| *Advice* | | | | | | |
| Tip | -.04 | 40 | 40 | 39 | 28 | 37 |
| Warning | .76 | 49 | 31 | 34 | 40 | 39 |
| Mean | .36 | 45 | 36 | 37 | 34 | 38 |
| *Inducements* | | | | | | |
| Promise | .93 | 53 | 68 | 64 | 46 | 58 |
| Threat | 1.50 | 48 | 50 | 50 | 50 | 50 |
| Mean | 1.22 | 51 | 59 | 57 | 48 | 54 |

Note. MP = Modus Ponens, DA = Denying the Antecedent, AC = Affirming the Consequent, MT = Modus Tollens.

*MSE* = 0.71. In addition, negative statements were more effective than positive statements, $F(1, 184) = 46.08$, *MSE* = 0.95; the interaction was not significant, $F(1, 184) = 1.57$, *MSE* = 0.71. Also consistent with previous findings, participants accepted more inferences for inducements than for advice, $F(1, 181) = 99.45$, *MSE* = 3.40. The difference between negative (44%) and positive (48%) statements was not significant, $F(1, 181) = 1.21$, *MSE* = 16.77, while the interaction between type and valence was significant, $F(1, 181) = 11.76$, *MSE* = 3.40.[7]

*Probability estimates*

We computed two conditional probabilities from participants' probability estimates of the four truth-table cases:

1. $P(q/p) = pq/(pq + p\sim q)$
2. $P(q/\sim p) = \sim pq/(\sim pq + \sim p\sim q)$

The by-item correlations between these two conditional probabilities, ratings of the truth of the conditionals, and their perceived behavioural effectiveness are presented in Table 6. These correlations are first presented separately for advice and inducements, and subsequently averaged over all of the conditionals. Furthermore, the correlations are given for both the derived (i.e., by using the truth-table values) and direct methods of obtaining $P(q/p)$ and $P(q/\sim p)$. As revealed by the table, these two methods generally provide very similar results. In fact, the correlation between estimates using the two methods was .91 for $P(q/p)$ and .87 for $P(q/\sim p)$, providing validation for the indirect method used by Over and Evans (2003).

[7] This pattern was obtained when each inference type was analysed separately, except that for both DA and AC, the valence factor was significant while the interaction did not reach significance.

Table 6

*Correlations in Experiment 3 between computed conditional probabilities and 1) truth ratings and 2) behavioural effectiveness for both derived and direct probability estimates*

| | Truth ratings | | Behavioural effectiveness | |
|---|---|---|---|---|
| Conditional probability | Derived | Direct | Derived | Direct |
| *Advice* (N = 18) | | | | |
| P(*q*/*p*) | .93** | .92** | -.25 | -.21 |
| P(*q*/~*p*) | -.15 | -.02 | .23 | .15 |
| *Inducements* (N = 18) | | | | |
| P(*q*/*p*) | .93** | .94** | -.34 | -.32 |
| P(*q*/~*p*) | .10 | .26 | -.58** | -.41* |
| *Overall* (N = 36) | | | | |
| P(*q*/*p*) | .91** | .93** | -.26 | -.21 |
| P(*q*/~*p*) | -.08 | .01 | -.38* | -.40** |

Note. * $p < .05$, ** $p < .01$. One-tailed tests.

The overall correlation between participants' truth ratings and P(*q*/*p*) was highly significant, and very similar in magnitude to the values reported by Over and Evans (2003). In contrast, the correlation between truth ratings and P(*q*/~*p*) was small and non-significant. This pattern of results did not change noticeably when analysing the correlations separately for advice conditionals and inducements conditionals.

For our measure of behavioural effectiveness, however, a different picture emerged. Specifically, effectiveness ratings were overall negatively correlated with P(*q*/~*p*), while the corresponding correlation with P(*q*/*p*) was not significant. These data are consistent with our hypothesis that part of what makes these statements effective is the understanding that *q* will likely not result unless *p* is fulfilled; that is, an effective statement is one for which P(*q*/~*p*) is low. When exploring these correlations separately for advice and inducements, however, some differences are revealed. The main difference is that the correlation between behavioural effectiveness and P(*q*/~*p*) was significant only for inducements.[8] However, there is good reason to suggest that this correlation should be greater for inducements. Specifically, the manifest goal of these statements is to induce a behavioural change, while advice is given more as a recommendation or prediction about future behaviour (e.g., Fillenbaum, 1986). Thus, we might expect the relationship between P(*q*/~*p*) and behavioural effectiveness to be stronger for inducements, which, as we have shown, are perceived as more effective statements than advice.

In summary, these data suggest a distinction between the conditions under which these conditionals are judged to be true, and the conditions under which they are perceived to be effective in changing behaviour. That is, whereas judgements of truth

[8] Indeed, the correlation between P(*q*/~*p*) and behavioural effectiveness for advice (although non-significant) was in the *positive* direction. This positive correlation was, however, mostly due to one warning conditional; upon removal of this statement, this correlation was close to zero.

were strongly correlated with P($q$/$p$), the behavioural effectiveness measure was better predicted by P($q$/~$p$). Furthermore, the overall correlation between ratings of truth and behavioural effectiveness was low and non-significant ($r$ = -.17).

Although the above analyses constituted the main focus of this study, we also report the correlations for the other two hypotheses examined in Over and Evans (2003), namely the material conditional and the conjunctive probability hypotheses. The probability of the material conditional was computed as 1 – $p$~$q$, while the conjunctive probability refers to P($p$&$q$). The correlation between truth ratings and the material conditional was highly significant ($r$ = .88); the corresponding correlation for the conjunctive probability was also very high ($r$ = .78). Furthermore, the inter-correlations between the conditional probability, the material conditional probability, and the conjunctive probability were in this data set all high (smallest $r$ = .68). We therefore performed a stepwise multiple regression analysis with truth ratings as the dependent variable, and entered P($q$/$p$), P($q$/~$p$), P($p$&$q$), and P(1 – $p$~$q$) as predictors. This analysis revealed that P($q$/$p$) accounted for 83% of the variance in truth ratings, while the remaining probabilities failed to account for any of the residual variance (smallest $p$ = .30).

*Inference rates*

In Table 7, we present the correlations between the different conditional probabilities computed from the four truth-table cases and acceptance rates for each inference. Overall, our predictions were supported, in that the willingness to endorse all four inferences was highly correlated with the conditional probability predicted to be

Table 7

*Correlations between conditional probabilities and inference rates in Experiment 3*

| | Inferences | | | |
|---|---|---|---|---|
| Conditional probability | MP | DA | AC | MT |
| P(*q*/*p*) | *.78*** | .03 | .13 | .28 |
| P(*~q*/*~p*) | .10 | *.83*** | .75** | .50** |
| P(*p*/*q*) | .37* | .79** | *.81*** | .37* |
| P(*~p*/*~q*) | .65** | .17 | .20 | *.54*** |

Note. * $p < .05$, ** $p < .01$. Correlations in italics are those predicted to be most relevant for each inference type.

most relevant to each of them (all $p < .01$). Specifically, P($q$/$p$) predicted the acceptance rate for MP, while P(~$q$/~$p$), computed as ~$p$~$q$/(~$pq$ + ~$p$~$q$), predicted the acceptance rate for DA. In addition, the acceptance rate for the AC inference varied as a function of P($p$/$q$), computed as $pq$/($pq$ + ~$pq$). Finally, P(~$p$/~$q$), computed as ~$p$~$q$/($p$~$q$ + ~$p$~$q$), predicted the acceptance rate for MT.

An inspection of Table 7 also reveals a number of significant correlations between inference rates and conditional probabilities other than the one predicted to be most relevant to each inference. This situation likely reflects the fact that the four conditional probabilities, as well as the four inferences, are themselves correlated. Stepwise multiple regression analyses confirmed that for each inference, the predicted conditional probability accounted for the largest portion of variance (smallest $R^2 = .27$), with the remaining probabilities accounting for less than 10% of the residual variance.

## Discussion

Our results extend the probabilistic approach to the interpretation of conditionals in three ways. Consistent with previous findings (e.g., Evans et al., 2003; Evans & Over, 2004; Oberauer & Wilhelm, 2003; Over & Evans, 2003), we first demonstrate that estimates of P($q$/$p$) can be used to predict the truth of pragmatically rich conditionals in the form of inducements and advice. In addition, we have extended the conditional probability approach by demonstrating a distinction between the perceived truth of these conditionals and their effectiveness as speech acts. Specifically, the degree to which these conditionals are perceived to be effective in changing behaviour varies as a function of P($q$/~$p$), but not P($q$/$p$). This dissociation reinforces the view that a proper understanding of inducements and advice, as well as other pragmatic conditionals, will require an

analysis extending beyond their truth status (e.g., Beller, 2002). Finally, we have extended this approach to explaining reasoning performance, by showing that conditional probabilities derived from the truth-table estimates can be used to predict inference patterns on a conditional inference task.

*Truth and behavioural effectiveness*

Our data showed a striking dissociation between the truth of the conditional and our measure of behavioural effectiveness. That is, the degree to which a conditional is perceived to be effective in changing behaviour does not seem to depend on the degree to which it is perceived to be true, nor does it appear to depend on the conditional probability of *q* given *p* (which, to a large degree, determines its truth value). These findings are counter-intuitive, in that it seems unlikely that someone will change their behaviour in response to a conditional regardless of whether they perceive the speaker of that conditional to be lying or telling the truth, and whether or not the promised (or threatened) outcome is, in fact, likely to be forthcoming.

There are a number of artifactual explanations for this null relationship. The first is that we have observed a Type II error. This explanation is rendered less likely by the fact that we have subsequently replicated the main findings of this experiment very closely with different participants and different items (see Experiments 4 and 5). The second explanation is that the problem is due to a truncated range in one or both of these variables. Specifically, our measure of behavioural effectiveness involves a difference score between two likelihood estimates, each rated on a seven-point scale; not surprisingly, the range in effectiveness scores is rather restricted. In addition, while we attempted to increase the range of probability estimates by using different consequents

for our conditionals (see method section), most statements were given moderately high truth ratings (mean = 65%), with very few perceived to be more likely false than true (i.e., below 50%).

A third possibility is that the relationship between truth and behavioural effectiveness is not linear. It is quite possible, for example, that behavioural effectiveness varies with the perceived truth of a conditional only up to a criterion, such that a conditional will be ineffective when it is perceived to have a low truth-value, but beyond a modest value, the perceived truth of a conditional may not predict its effectiveness. Thus, a promise will be perceived as ineffective when the listener does not believe the speaker to be telling the truth (i.e., truth is necessary for effectiveness); however, the promise may still be ineffective when its truth is not in doubt (e.g., when the listener is indifferent about the reward). This explanation is given credence by the fact that our conditional statements received relatively high truth ratings; it is possible that a relationship between truth and behavioural effectiveness would emerge with the inclusion of more conditionals with lower truth values.[9] This explanation is similar to the arguments made by Hilton and Erb (1996), who distinguish between the truth of causal conditionals and their explanatory relevance. Specifically, they argue that relevance may only be a factor when a causal explanation has some minimal truth-value; thus, truth may be a necessary condition for relevance to come into play.

Despite these potential problems, our results clearly show that it is possible to uncover relationships with these variables. Thus, despite a restricted range of effectiveness scores, the degree to which these pragmatic conditionals are perceived to be

[9] To deal with the possibility of non-linearity, we applied a square root transformation to the truth variable. However, the correlation between behavioural effectiveness and truth ratings remained unchanged after this transformation.

effective in changing behaviour was found to correlate with estimates of the conditional probability of *q* given ~*p*. This correlation was furthermore only significant for inducement conditionals, where the relationship between effectiveness and P(*q*/~*p*) might be expected to be stronger. In other words, while the problems of restricted range and non-linearity may have acted to suppress the intuitive relationship between behavioural effectiveness and truth [as well as between effectiveness and P(*q*/*p*)], we can nevertheless be confident in the conclusion that P(*q*/~*p*) plays a crucial role in accounting for the perceived effectiveness of these speech acts.

*Conditional probabilities and representation*

Thus far, we have shown that two conditional probabilities, P(*q*/*p*) and P(*q*/~*p*), accurately predict a number of judgments that participants make about conditional inducements and advice. These probabilities, however, are not explanatory constructs. They are, rather, mathematical summaries that represent the culmination of one or more underlying representational processes. Although the factors underlying probability judgments have not been examined directly, there have been a number of explanations whose goal is to explain the probabilistic nature of conditional reasoning that can readily be extended as plausible candidates to mediate psychological estimates of P(*q*/*p*) and P(*q*/~*p*).

For example, the number and availability of counter-examples are known to mediate reasoning performance (e.g., Byrne, Espino, & Santamaria, 1999; Cummins 1995; Cummins et al., 1991; De Neys et al., 2003; Thompson, 1995). Indeed, the retrieval of a single counter-example (Markovits & Quinn, 2002), or a judgment about the probability with which such an example would take effect (Thompson, 2000), is

sufficient to produce probabilistic patterns in people's inferences. Thus, it seems likely that probability estimates may be mediated by the availability of instances of the form "*p* and ~*q*", or "~*p* and *q*", that come to mind.

Over and Evans (2003) offered a similar proposal, suggesting that psychological estimates of P(*q*/*p*) may be based on memorial representations of the relative frequency of two event combinations, namely "*p* and *q*" versus "*p* and ~*q*". Alternatively, they suggested that the availability or vividness of specific scenarios may underlie such judgments. In addition, reasoners may base their judgments on their causal models of the mechanisms mediating the relationship between *p* and *q* (e.g., Fugelsang & Thompson, 2003). For example, if asked to estimate the probability that John will pay me $10 if I mow his lawn, I might rely on the fact that, in my experience, most such transactions have been successfully completed in the past. Alternatively, I might anchor my judgment in a highly memorable case in which I was ripped off after mowing someone's lawn, or base my judgments on my knowledge of John's character.

*Conditional probabilities and necessity/ sufficiency relations*

Finally, it is worth noting that the conditional probabilities defined by P(*q*/*p*) and P(*q*/~*p*) overlap conceptually to a large extent with the constructs of perceived sufficiency and necessity. These constructs predict up to 70% of the variance in the rate at which conditional inferences are endorsed for a large variety of conditional statements (Thompson, 1994, 1995, 2000; Thompson & Mann, 1995). Perceived sufficiency reflects the extent to which the antecedent *p* guarantees the outcome *q*, whereas perceived necessity reflects the extent to which the antecedent is a required condition for the

consequent to occur. Thus, when a relationship is completely sufficient, P(*q*/*p*) will equal one; and when a relationship is necessary, P(*q*/~*p*) will equal zero.

One contribution of this paper is to document the close relationship between these constructs, as the derived estimates of P(*q*/*p*) and P(*q*/~*p*) computed on the basis of the truth-table cases correlate very highly with our direct estimates of these probabilities, which in effect ask for judgements of sufficiency and necessity. Furthermore, like conditional probabilities, the constructs of perceived necessity and sufficiency are continuous variables, taking a range of values from zero to 1; also like conditional probabilities, they are meta-theoretical constructs that subsume a number of underlying processes (e.g., the availability of counter-examples). Under this model, therefore, the truth of a conditional statement will vary as a function of the sufficiency of *p* for *q*, and the perceived behavioural effectiveness as a function of the necessity of *p* for *q*.

These constructs do not overlap entirely, and the extent to which they are likely to predict performance in a given circumstance will depend on the degree to which the context takes advantage of the subtle differences between them. Specifically, perceived necessity and sufficiency carry a causal connotation, in which one condition is denoted as the temporally prior event. For some conditional reasoning tasks, this temporal sequence is a key element of the task, for instance when reasoning about statements phrased as "*p* only if *q*" (e.g., Evans, 1977; Thompson & Mann, 1995). In these cases, the temporal ordering inherent in the necessity/ sufficiency constructs gives them an explanatory edge.

In other cases, the a-temporal relations implied by conditional probabilities will be more appropriate. Take for example a conditional in which the relationship is epistemic in nature (i.e., where knowing that *p* is enough to know that *q*):

If the flag is at half-mast, then someone important has died.

Here, *p* cannot be read to be a sufficient condition for *q*; indeed, the causality works in the opposite direction (i.e., someone important dying leads to the flag being flown at half-mast). Although it is possible to give an adequate account of these epistemic relations by analysing the necessity and sufficiency of the underlying causal sequence (Thompson, 1994), the conditional probability approach offers a more straightforward account. That is, these, like other conditional relations, will be perceived to be true to the extent that P(*q*/*p*) is high. In this case, the conditional is true to the extent that the probability that someone important having died, given the flag at half-mast, exceeds the probability that no one important has died, given the flag at half-mast.

*Conclusions*

Overall, the findings of this study provide further support for a probabilistic interpretation of conditional statements. Thus, consistent with related probabilistic accounts (e.g., Evans et al. 2003, Evans & Over, 2004; Liu et al., 1996; Oaksford & Chater, 2001, 2003; Oaksford et al., 2000; Oaksford & Moussakowski, 2004; Oaksford & Wakefield, 2003; Oberauer & Wilhelm, 2003; Over & Evans, 2003), we have shown that conditional probabilities can be used to predict a number of judgments that people make about conditional inducements and advice, including judgments about their truth status, their effectiveness as speech acts, and the inferences they invite. However, little is at present known about the representational processes that underlie these probability judgments; while we have suggested some factors that may determine such judgments (e.g., availability of counter-examples, the vividness of specific situations), it is clear that much research is needed to elucidate this issue further.

# EXPERIMENTS 4 & 5

# EXAMINING THE INDEPENDENCE BETWEEN TRUTH AND EFFECTIVENESS

## Experiment 4

The results of Experiment 3 seem to suggest a distinction between the conditions that influence the truth of conditional inducements and advice, and the conditions that contribute to their effectiveness as speech acts. That is, consistent with previous research (e.g., Evans, Handley, & Over, 2003; Evans & Over, 2004; Oberauer & Wilhelm, 2003; Over & Evans, 2003), estimates of P($q$/$p$) strongly predicted the probability that these conditionals were perceived to be true. However, this conditional probability did not predict the degree to which inducements and advice were perceived to be effective in changing behaviour; these ratings were instead (negatively) correlated with P($q$/~$p$). In addition, the correlation between ratings of truth and behavioural effectiveness was non-significant, further supporting the dissociation between these two measures.

### *Exploring the independence hypothesis*

As previously discussed, however, this apparent dissociation between truth and effectiveness seems to be somewhat counter-intuitive. In particular, these results suggest that the degree to which a conditional is effective in changing the listener's behaviour does not depend on whether the speaker is perceived to be lying or telling the truth, nor on whether the likelihood of obtaining the consequent $q$ (upon the fulfillment of the antecedent $p$) is high or low. To appreciate the counter-intuitive nature of these findings, consider again the promise "If you wash the car, I will give you $10". In this example, a moderately high truth value would seem to be a necessary condition for this statement to be effective, as it would be imprudent for the listener to incur the cost of washing the car

if the speaker is perceived to be lying and, consequently, if $10 is unlikely to be forthcoming. Given that there are good reasons to expect a relationship between behavioural effectiveness and truth, a replication of Experiment 3 is clearly needed; one goal of the current study was thus to test the reliability of this dissociation.

Another goal was to provide a stronger test of this apparent dissociation. Thus, if truth and effectiveness are independent, these measures should also be affected by different variables. The results of Experiment 3 offered support for this hypothesis, as truth ratings were highly correlated with estimates of P(*q*/*p*) and effectiveness ratings instead correlated with P(*q*/~*p*). In the current experiment, a different design was employed to further test the hypothesis that truth and effectiveness are independent and affected by different variables. Specifically, two pragmatic factors were manipulated, which were expected to affect either P(*q*/*p*) or P(*q*/~*p*); the impact of these manipulations on our measures of truth and effectiveness was then explored.

In an attempt to vary estimates of P(*q*/*p*), the perceived *credibility* of conditional inducements and advice was manipulated. To illustrate, the credibility of the tip "If you show up early for work, you will impress the boss" was manipulated by changing the speaker giving this advice (for a similar manipulation, see Stevenson & Over, 2001). We reasoned that this tip would be perceived as less credible when uttered by a friend, as compared to a senior colleague in the company (who presumably would have a better idea of what is likely to impress the boss). Furthermore, we expected that this manipulation would affect estimates of P(*q*/*p*), such that the likelihood of impressing the boss upon showing up early would be higher when this tip was given by the high-credibility speaker (i.e., the colleague).

A different pragmatic factor was manipulated in an attempt to vary estimates of P(*q*/~*p*). Specifically, we provided an *alternative antecedent* (i.e., besides *p*) for why the consequent *q* could come about (for similar manipulations, see Bonnefon & Hilton, 2004; Byrne, 1989; Byrne, Espino, & Santamaria, 1999; Chan & Chua, 1994; De Neys, Schaeken, & d'Ydewalle, 2003; Rumain, Connell, & Braine, 1983; Stevenson & Over, 1995). In our tip, for example, the speaker would indicate that another way to impress the boss (besides showing up early) was by working late. We expected that providing an explicit alternative would suggest that *q* could occur even in the absence of *p*; providing such an alternative should thus increase estimates of P(*q*/~*p*), relative to only presenting the original statement.

To recap, we tested the independence hypothesis by manipulating two pragmatic factors, which in turn were expected to affect estimates of P(*q*/*p*) and P(*q*/~*p*). Based on the findings of Experiment 3, we would predict that varying estimates of P(*q*/*p*), achieved by manipulating the credibility of the conditional, should only affect truth ratings. Similarly, varying P(*q*/~*p*), achieved by manipulating the presence of an alternative antecedent, should only have an impact on effectiveness ratings.

However, given that there are good reasons to expect a relationship between the perceived truth and the perceived effectiveness of conditional inducements and advice, a different pattern of results might emerge from these manipulations. Hence, another possibility is that manipulating credibility will also influence the degree to which inducements and advice are effective in changing behaviour. In other words, by reducing the credibility of the statement, not only should its truth be questioned, but the effectiveness of this conditional may also be reduced. In this scenario, the perceived

effectiveness of conditional inducements and advice will no longer be independent of their truth status. That is, by varying estimates of P(*q*/*p*), which are known to be strong predictors of truth ratings, we will also be changing effectiveness ratings.

On the other hand, there may be less reason to expect that the alternative antecedent manipulation will affect truth ratings. While providing an alternative antecedent should reduce effectiveness ratings [by increasing P(*q*/~*p*)], doing so should not question the truth of the original conditional. For example, the truth of the tip "If you show up early for work, you will impress the boss" should not be reduced when the speaker suggests that the same outcome may also result from working late.

Finally, note that a proper examination of the independence hypotheses requires that our manipulations *selectively* affect P(*q*/*p*) and P(*q*/~*p*). For instance, when exploring the impact of the credibility manipulation on measures of truth and effectiveness, this manipulation should ideally change estimates of P(*q*/*p*) while keeping estimates of P(*q*/~*p*) constant. That is, if manipulating credibility also affects P(*q*/~*p*), we would in fact expect a corresponding change in effectiveness ratings, a change not due to the impact of P(*q*/*p*). We will return to this issue when reporting the results.

To explore the independence hypothesis, we again used the probability estimates task introduced by Over and Evans (2003). As in Experiment 3, participants were asked to rate the probability that the conditional statement was true. In addition, they provided probability estimates (summing to 100%) for each of the four possible outcomes of *p* and *q*; from these estimates, we computed P(*q*/*p*) and P(*q*/~*p*). Finally, a measure of behavioural effectiveness was obtained by asking participants to judge the probability of *p* occurring both before and after the conditional was given; the difference between these

two scores provides an indication of the effectiveness of the conditional in bringing about the behaviour described in *p* (e.g., how effective the promise of $10 is in inducing the listener to wash the car).

Unlike our previous experiments, however, we only included conditional inducements and advice with positive outcomes in this study. That is, we focused on promises and tips, and omitted threats and warnings. The main reason for excluding these conditionals was that we were more interested in exploring the relationship between the conditional probabilities of P(*q*/*p*) and P(*q*/~*p*) and ratings of truth and effectiveness than in investigating any differences between positive and negative statements. Furthermore, omitting threats and warnings kept the complexity of our design to a manageable level (i.e., by including three, rather than four, factors).

*Exploring truth-table performance*

In Experiment 3, we also demonstrated that the conditional probability approach (e.g., Evans & Over, 2004) could be extended to predict reasoning performance. Specifically, acceptance rates on the conditional inference task were found to be highly correlated with the conditional probability relevant to each inference type. For instance, inference rates for AC (i.e., "*q*; therefore *p*") correlated .81 with estimates of the conditional probability of *p* given *q*, P(*p*/*q*). These findings are consistent with the view that people's reasoning performance should be interpreted within a probabilistic framework (e.g., Oaksford & Chater, 2001, 2003; Oaksford, Chater, & Larkin, 2000). A final goal of the current study was to provide further support for this view by extending the conditional probability approach to a different reasoning task. Consequently, participants were given another formal deductive task: the truth table evaluation task.

In this task, participants are provided with the four cases of the truth table and are asked to indicate whether these cases are true or false (i.e., consistent or inconsistent) with respect to the conditional statement (e.g., Evans, Newstead, & Byrne, 1993; Manktelow, 1999). As compared to the conditional inference task, a slightly different notation is used for the truth table task. Specifically, the four different combinations of *p* and *q* (e.g., *pq*, *p*~*q*, ~*pq*, and ~*p*~*q*) are instead described with reference to the truth or falsity of the antecedent and consequent terms (i.e., TT, TF, FT, and FF). For example, FT describes the outcome where *p* is false and *q* is true.

Formal logic recognizes two interpretations of the truth table: *material implication* and *material equivalence* (e.g., Evans et al., 1993). Under the material implication interpretation, the conditional is only falsified when *p* is true and *q* is false (i.e., TF), and is true in the other three cases (see Table 8). Material equivalence differs from material implication in that FT also falsifies the conditional, such that *p* implies *q* and *q* also implies *p*; note that this difference makes material equivalence the truth table representation of the biconditional "if and only if *p*, then *q*". Logic only permits propositions to take two values (i.e., true or false); however, participants typically classify false antecedent cases (i.e., FT and FF) as irrelevant (e.g., Johnson-Laird & Tagart, 1969). Allowing for a third "irrelevant" response for these cases thus creates two further truth table interpretations: *defective implication* and *defective equivalence*.

In their study exploring reasoning with realistic conditionals, Newstead, Ellis, Evans, and Dennis (1997: Experiments 1-3) found systematic differences in performance on the truth table task between conditional inducements and advice. In particular, the FT case was more often classified as false for inducements, while irrelevant was the

Table 8

*Four truth table interpretations of the conditional*

| Case p | Case q | Material implication | Material equivalence | Defective implication | Defective equivalence |
|---|---|---|---|---|---|
| T | T | T | T | T | T |
| T | F | F | F | F | F |
| F | T | T | F | I | F |
| F | F | T | T | I | I |

T = True: F = False; I = Irrelevant

dominant response for advice. In other words, obtaining *q* in the absence of *p* was more likely to be interpreted as contradicting inducements than advice. As Table 8 shows, a false response to the FT outcome leads to an equivalence reading; consequently, equivalence interpretations were more prevalent for inducements than for advice.

Based on the findings of Newstead et al. (1997), therefore, conditional promises were in the current study predicted to generate more false responses to the FT case, relative to conditional tips. In addition, however, we were interested in exploring the relationship between performance on this reasoning task and the conditional probability approach (e.g., Evans & Over, 2004). That is, can the conditional probability approach be used to explain truth table responses in the same way it was found to predict acceptance rates on the conditional inference task in Experiment 3?

In fact, some relevant research suggests that it might. Specifically, Thompson (1994, 2000) reported that the constructs of perceived sufficiency and necessity were excellent predictors of performance on the truth table task. Recall that sufficiency reflects the extent to which the occurrence of *p* guarantees the occurrence of *q*, while necessity refers to the degree to which the absence of *q* is guaranteed by the absence of *p*. These constructs are conceptually very close to our measures of P(*q*/*p*) and P(*q*/~*p*); indeed, the derived estimates of P(*q*/*p*) and P(*q*/~*p*) were in Experiment 3 found to correlate about .90 with the corresponding direct estimates of these conditional probabilities, which in effect ask for ratings of sufficiency and necessity.

Thompson (1994, 2000) observed that the tendency to respond false to TF and FT correlated with ratings of perceived sufficiency and necessity, respectively. Given the substantial degree of overlap between these constructs and our measures of P(*q*/*p*) and

P($q$/~$p$), we therefore expected that estimates of these conditional probabilities could be used to predict truth table responses. Specifically, responses to TF were expected to correlate with P($q$/$p$), and FT responses to correlate with P($q$/~$p$).

In summary, this experiment had three main goals. First, we aimed to replicate the results of Experiment 3, and in particular, the counter-intuitive finding suggesting that the effectiveness of conditional inducements and advice is unrelated to the probability that they are judged to be true. Second, to further test the hypothesis that the perceived truth and the perceived effectiveness of these statements are independent, and affected by different variables, we attempted to manipulate two pragmatic factors; manipulating the credibility of the conditional was expected to affect P($q$/$p$), while providing an alternative antecedent was expected to affect P($q$/~$p$). Finally, we examined the relationship between the conditional probability approach and reasoning performance on the truth table evaluation task.

## Method

### *Participants*

A total of 79 University of Saskatchewan undergraduate students participated in partial fulfillment of an introductory psychology course requirement.

### *Materials*

The conditionals used in this experiment were initially selected from a larger pool of statements on the basis of a pilot study. For each of the conditionals in this initial pool, we created four conditions (see below for examples of conditionals included in the study proper). These conditions corresponded to the four combinations of our two pragmatic

factors: credibility (high vs. low) and alternative antecedent (absent vs. present). Note that the conditional statement itself remained the same across these four conditions.

This larger pool of conditionals was assessed according to a number of selection criteria. First, we asked participants to indicate which of several definitions best described each conditional, and selected the statements with the highest level of category agreement. In addition to the categories of promise (i.e., "an inducement, where a reward is offered to encourage a specific action") and tip (i.e., "a recommendation highlighting positive outcomes if a specific action is carried out"), we also provided definitions specifying other types of conditionals, including causal statements (i.e., "a statement describing a cause and effect relationship") and threats (i.e., "a deterrent, where punishment is used to discourage a specific action"). Second, statements were evaluated on the degree of speaker's control, a crucial variable differentiating conditional inducements and advice (see Experiment 1). Finally, we obtained ratings of sufficiency and necessity, which served as rough estimates of P($q$/$p$) and P($q$/~$p$), and examined each conditional with regard to the effects of our pragmatic manipulations. The selected statements were those that conformed best to the intended effects of these manipulations (i.e., statements where the credibility manipulation selectively affected sufficiency ratings and where the alternative antecedent manipulation selectively influenced necessity ratings).

Based on the ratings of the pilot study, eight conditionals were selected for this experiment, of which four were conditional tips and four were conditional promises. Appendix E includes a complete list of these materials.

To illustrate the manipulations of our two pragmatic factors, consider the tip "If you show up early for work, you will impress the boss". Credibility was manipulated by varying the speaker of this advice, uttered either by a senior colleague (high credibility) or a friend (low credibility). For the other manipulation, an alternative antecedent was provided, which described another means by which the consequent could come about. In this particular tip, the speaker would continue by saying that another way of impressing the boss was by working late. In the conditions where the alternative antecedent was provided, it would always be presented after the conditional statement. Furthermore, both the original and alternative antecedents were designed to be singly, as opposed to jointly, sufficient to bring about the consequent (i.e., impressing the boss would likely result from either showing up early *or* working late).

For conditional promises, credibility was manipulated slightly differently. Specifically, rather than changing the speaker, participants were given additional information intended to reduce the credibility of the statement. As an example, one promise read "If you take your medication this week, you can go home for the weekend", spoken by a nurse to a psychiatric patient resisting treatment. In the high credibility condition, no further information was provided. However, in the low credibility condition, participants were informed that it is in fact the attending psychiatrist who determines whether patients are allowed to go home for the weekend. This additional information should have the effect of questioning the credibility of the nurse's promise. The alternative antecedent manipulation was, however, similar for promises as for tips; in this particular example, the nurse would tell the patient that attending group therapy during the week would also allow him to go home for the weekend.

*Design and Procedure*

Each participant was presented with (one condition of) all eight conditional statements. This experiment used a 2x2x2 mixed design, with statement type (tip vs. promise) and credibility as within-subjects factors and the alternative antecedent manipulation as a between-subjects factor. The order of the eight statements was randomized for each participant. For each statement, participants were asked to complete three tasks, which are described below.

*Behavioural effectiveness task*. As in Experiment 3, this task involved a difference score between two estimates of the likelihood of *p*. First, only the social context was provided (without presenting the conditional statement), and participants were asked to estimate the probability of *p* occurring. For example, in the promise described above, participants were told that a patient was resisting treatment, and were asked to indicate how likely the patient was to take his medication that week. Subsequently, they were provided with the conditional statement (which informs the patient that taking medication will grant him a trip home for the weekend); a second estimate of the probability of *p* was then solicited. For both of these questions, an 11-point probability scale with values between 0 and 100 was used. By subtracting the first estimate from the second, we obtain a measure of the degree to which the conditional statement was perceived to be effective in changing the behaviour described in *p*.

*Probability estimates task*. The methodology introduced by Over and Evans (2003) was again used to obtain truth ratings of the conditionals and to compute conditional probability estimates. Specifically, participants were first asked to rate the probability that the conditional statement itself was true. Next, they rated the likelihood

of each of the four events corresponding to the different combinations of *p* and *q*. As an example, the four events for the promise "If you take your medication this week, you can go home for the weekend" are as follows:

The patient takes his medication and goes home for the weekend (*pq*)
The patient takes his medication and does not go home for the weekend (*p*~*q*)
The patient does not take his medication and goes home for the weekend (~*pq*)
The patient does not take his medication and does not go home for the weekend (~*p*~*q*)

For this task, participants were asked to indicate the chances that these events would occur, ensuring that the four estimates added up to 100%. Furthermore, they were asked to read through all four events before providing their estimates.

*Truth table evaluation task*. For this reasoning task, participants were again provided with the same four cases (corresponding to TT, TF, FT, and FF). However, they were here asked to indicate whether each of these outcomes 1) support, 2) contradict, or 3) tell us nothing about the conditional statement. This format is somewhat different from the traditional truth table task, wherein participants are asked whether these cases are true, false, or irrelevant with respect to a conditional rule (e.g., Evans et al., 1993; Manktelow, 1999). The traditional format was originally used to explore truth table responses on relatively abstract statements; the realistic conditionals used in this experiment require a slightly different wording (see also Newstead et al., 1997). Participants were instructed to give only one response to each of the four cases.

Participants always performed the behavioural effectiveness task first. The order of the two remaining tasks was counterbalanced within participants, such that the probability estimates task was performed second for four of the statements, and third for the remaining four statements. Finally, we presented the four cases of the probability estimates and truth table tasks in two quasi-random orders; for each participant, the standard order (as given above) was presented for half of the statements, while the four cases were given in a reversed order for the remaining statements.

Participants were tested in one large group and were given written instructions. Participants were asked to complete the questions in the order they appeared; furthermore, they were told to work at their own pace.

## Results and Discussion

One participant failed to comply with instructions and was therefore removed from data analysis. For all subsequent analyses, an alpha of .05 was used.

### *Replication of previous findings*

*Inducements vs. advice*. To verify that conditional inducements are perceived to be more effective than conditional advice in changing behaviour, a difference score between participants' two likelihood ratings of $p$ (i.e., before and after knowledge of $q$) was computed. Consistent with the findings from Experiments 1 and 3, effectiveness ratings were higher for conditional promises (mean score = 23.97) than conditional tips (mean score = 18.90), $F(1, 76) = 10.32$, $MSE = 193.84$.

*Correlations*. Next, we repeated the main analyses from Experiment 3. Specifically, from the ratings of the probability estimates task, $P(q/p)$ was computed as $pq/(pq + p\sim q)$ and $P(q/\sim p)$ as $\sim pq/(\sim pq + \sim p\sim q)$. The correlations between these

conditional probabilities, ratings of the truth of the conditionals, and their effectiveness scores are presented in Table 9. Note that, as was the case in Experiment 3, items were used as the units of analysis.

As can clearly be seen, we replicated the results of Experiment 3 very closely. That is, the probability that these conditionals were judged to be true was highly correlated with computed estimates of P(*q*/*p*), while the corresponding correlation with P(*q*/~*p*) was non-significant. In contrast, effectiveness ratings were negatively correlated with P(*q*/~*p*), and unrelated to P(*q*/*p*). Furthermore, the correlation between truth ratings and effectiveness ratings was non-significant, and in the negative direction ($r$ = -.30).

These findings thus provide further support for the dissociation observed in Experiment 3 between the truth status of conditional inducements and advice and their effectiveness as speech acts. In addition, the degree to which these conditionals are perceived to be effective in changing behaviour again is unrelated to the perceived likelihood of obtaining *q* upon the fulfillment of *p*, that is, on P(*q*/*p*). Overall, this replication renders less likely the possibility that the counter-intuitive findings of Experiment 3 were due to Type II errors.

*Credibility and alternative antecedent manipulations*

For the next set of analyses, we performed 2x2x2 mixed design ANOVAs to explore the effects of our credibility and alternative antecedent manipulations. Two of these factors, credibility (low vs. high) and statement type (tip vs. promise), were within-subjects, while the manipulation of alternative antecedent was a between-subjects factor. We performed these analyses on four dependent variables: ratings of P(*q*/*p*), P(*q*/~*p*), truth, and behavioural effectiveness. Ratings on these variables are reported in Table 10,

Table 9

*By-item correlations (N = 32) between computed conditional probabilities, truth ratings, and behavioural effectiveness ratings in Experiment 4*

| Conditional probability | Truth ratings | Effectiveness |
|---|---|---|
| P(*q*/*p*) | .86** | .001 |
| P(*q*/~*p*) | .15 | -.38* |

Note. * $p < .05$, ** $p < .01$.

Table 10

*Ratings on P(q/p), P(q/~p), truth, and behavioural effectiveness as a function of credibility and alternative antecedent in Experiment 4*

| Manipulations | P(*q*/*p*) | P(*q*/~*p*) | Truth | Effectiveness |
|---|---|---|---|---|
| Credibility | | | | |
| High | 71.19 | 34.42 | 67.90 | 22.73 |
| Low | 66.02 | 31.90 | 61.74 | 20.15 |
| Alternative antecedent | | | | |
| No alternative | 70.18 | 33.40 | 65.78 | 24.16 |
| Alternative | 67.03 | 32.92 | 63.86 | 18.72 |

separately for the manipulations of credibility and alternative antecedent. [10] For ease of presentation, these analyses are grouped by the independent variables (i.e., the manipulations of credibility and alternative antecedent), rather than by the dependent variables. In addition, none of the following main effects were qualified by higher-order interactions. That is, the two-way or three-way interactions were not significant for any of these analyses. [11]

*Manipulating credibility*. As expected, we found that reducing perceived credibility decreased ratings of P(*q*/*p*), $F(1, 76) = 12.93$, $MSE = 161.15$. Although significant, an inspection of Table 10 reveals that the magnitude of the difference between high-credibility and low-credibility conditionals was not very large. However, ratings of P(*q*/~*p*) did not differ as a function of credibility, $F(1, 76) = 2.22$, $MSE = 222.15$. In other words, the manipulation of credibility was fairly successful, in that it was found to selectively affect estimates of P(*q*/*p*).

Consequently, we are justified in exploring the impact of this manipulation on measures of truth and behavioural effectiveness. As expected, probability ratings of truth were higher in the high-credibility condition, $F(1, 72) = 12.30$, $MSE = 228.19$. This finding confirms the close relationship that exists between P(*q*/*p*) and the perceived truth of conditional inducements and advice. That is, higher estimates of P(*q*/*p*) resulted in increased truth ratings.

---

[10] The third independent variable (i.e., statement type) was of less importance in this study, but here are the relevant results. As previously reported, promises were perceived to be more effective than tips; given that effectiveness correlated negatively with P(*q*/~*p*), promises (28.70%) were also given lower scores than tips (37.62%) on this conditional probability, $F(1, 76) = 21.80$, $MSE = 283.99$. Furthermore, the main effect of statement type was significant for both P(*q*/*p*) and truth ratings [$F(1, 76) = 14.90$, $MSE = 221.88$, and $F(1, 72) = 13.12$, $MSE = 200.58$, respectively]. Specifically, both ratings were higher for promises than for tips [P(*q*/*p*): 71.86% vs. 65.35%, truth: 67.80% vs. 61.84%, respectively).

[11] However, the power to detect these interactions was generally quite low, especially for the three-way interaction (range = .06 - .10). Even for the two-way interactions, power was low (range = .06 - .49).

Reducing credibility did not, however, result in decreased effectiveness ratings, $F(1, 76) = 2.81$, $MSE = 184.59$ ($p = .10$). In other words, consistent with the correlational data indicating that effectiveness is not predicted by estimates of P($q$/$p$), manipulating the credibility of these conditional promises and tips [and consequently P($q$/$p$)] did not change the degree to which they were perceived to be effective in changing behaviour. However, our power to reject this particular null hypothesis was only .38, and Table 10 does reveal a trend for lower effectiveness ratings in the low-credibility condition. Combined with the relatively modest magnitude of the credibility manipulation on estimates of P($q$/$p$), these findings suggest that our experiment may not have been powerful enough to detect a relationship between P($q$/$p$) and behavioural effectiveness. Nevertheless, it seems safe to say that the main role of P($q$/$p$) is to determine the truth of conditional inducements and advice.

*Manipulating the alternative antecedent*. As is clear from Table 10, manipulating the presence of an alternative antecedent did not produce the expected effects on ratings of P($q$/~$p$). Specifically, providing an alternative antecedent did not result in higher estimates of P($q$/~$p$), $F(1, 76) = .03$, $MSE = 652.84$. In other words, the probability of obtaining the consequent $q$ in the absence of the original antecedent $p$ did not increase when an alternative means of obtaining $q$ was explicitly provided. This failure to affect P($q$/~$p$) in the predicted manner obviously qualifies the conclusions one can draw about the remaining findings concerning this manipulation. Nevertheless, providing an alternative antecedent did not affect P($q$/$p$) ratings, $F(1, 76) = 1.10$, $MSE = 702.82$, nor ratings of truth, $F(1,72) = .48$, $MSE = 568.85$. Hence, as predicted, specifying an alternative way of obtaining $q$ did not question the sufficiency of the original antecedent

in bringing about the consequent, and neither did it reduce the truth of the conditional statement itself.

This manipulation was, however, found to influence behavioural effectiveness. Specifically, while providing an alternative antecedent did not increase estimates of P($q$/~$p$), it still resulted in lower effectiveness ratings, $F(1, 76) = 4.25$, $MSE = 542.91$. Thus, when informing participants that $p$ was only one of two possible ways to bring about $q$, they judged the conditional statement to be less effective in changing the behaviour described in $p$. This finding makes intuitive sense, as the speaker in effect is giving the listener a choice between two options by which $q$ will occur; the likelihood that any of these two options is chosen should decrease relative to the scenario in which only a single course of action is specified.

While intuitive, this finding is also rather puzzling. The correlational data from Experiment 3 and its replication in the current study showed that behavioural effectiveness was predicted by P($q$/~$p$). The results of the alternative antecedent manipulation, however, suggest that effectiveness may also be influenced *without* varying estimates of P($q$/~$p$). In fact, the magnitude of the correlation coefficients between P($q$/~$p$) and effectiveness observed in these experiments (around .40) support the view that other, as yet unidentified, factors may play a role in determining the effectiveness of conditional inducements and advice.

For example, perhaps one consequence of the alternative antecedent manipulation was to reduce the perceived *importance* of $p$. Usually, the manifest goal of a speaker uttering a promise, for instance, is to persuade the listener to perform $p$; indeed, the inducement in $q$ is offered precisely to increase the likelihood of this goal. When the

speaker instead offers an alternative means by which the reward will be given, however, the importance of this original goal may be reduced. Consider the promise "If you take your medication this week, you can go home for the weekend". In this example, the importance that the nurse attaches to *p* will be reduced when attending group therapy is offered as an alternative. Furthermore, sensing this reduction in importance, listeners may become less willing to perform *p*. This possibility may thus help in explaining how the provision of an additional antecedent may have reduced effectiveness ratings without influencing ratings of P(*q*/~*p*).

However, why were estimates of P(*q*/~*p*) unaffected by the alternative antecedent manipulation? One possibility is, of course, that this failure reflects a Type II error. Another possibility is that giving participants an explicit alternative may not be needed; they may be thinking of such alternatives spontaneously. Consider the tip "If you show up early for work, you will impress the boss". When interpreting this tip, world knowledge may suggest that showing up early is only one way of impressing one's boss. For example, a boss is also likely to be impressed by employees who work hard or stay late. Reasoners may therefore spontaneously generate potential counterexamples in the form *~p and q* (e.g., Byrne et al., 1999; Thompson, 1994, 1995, 2000). Note that this interpretation seems less plausible for promises, since the biconditional nature of these statements normally suggests that *p* is a necessary condition for *q* to occur (i.e., it would seem more difficult to imagine counterexamples for promises).

Nevertheless, if reasoners in fact engage in spontaneous generation of such counterexamples, then the presentation of an alternative antecedent would be unlikely to change estimates of P(*q*/~*p*). That is, providing an alternative antecedent may have failed

to increase estimates of P($q$/~$p$) because doing so was largely redundant, as participants who were not provided with one were generating such alternatives by themselves. To counteract this strategy, one option would be to explicitly inform participants in this condition that $p$ is the *only* behaviour that can result in $q$.

*Truth table responses*

Table 11 shows the percentages of true, false, and irrelevant responses on the truth table task, separately for conditional tips and conditional promises. Overall, the main results of Newstead et al. (1997) were confirmed. Specifically, a majority of the responses to the TT and TF cases were classified as true and false, respectively. Furthermore, while Newstead et al. found both false and irrelevant responses to be frequent for FT, this case was primarily judged as false in the current study. Finally, the dominant response for the FF case was true, although the irrelevant response was also quite frequent. A look back at Table 8 reveals that the dominant truth table pattern in this experiment thus conforms to the material equivalence interpretation (i.e., TFFT).

In addition, we replicated the main difference between conditional inducements and advice identified by Newstead et al. (1997). Specifically, the percentage of false responses to FT was higher for promises than for tips, $t(30) = 3.31$, SE = 4.32. In other words, the scenario in which the consequent $q$ comes about in the absence of the antecedent $p$ was more often interpreted as contradicting a conditional promise than a conditional tip. As previously discussed, a false response to FT results in an equivalence interpretation; consequently, promises (62%) were more likely to be given equivalence interpretations (both material and defective) than tips (50%), $t(30) = 2.54$, SE = 4.75.

Table 11

*Percentages of true, false, and irrelevant responses on the truth table task as a function of statement type in Experiment 4*

| | Truth table case | | | |
|---|---|---|---|---|
| Response | TT | TF | FT | FF |
| *Tip* | | | | |
| True | 93 | 6 | 6 | 53 |
| False | 3 | 83 | 66 | 13 |
| Irrelevant | 5 | 11 | 28 | 34 |
| *Promise* | | | | |
| True | 96 | 7 | 4 | 59 |
| False | 2 | 79 | 80 | 10 |
| Irrelevant | 2 | 14 | 16 | 31 |

Next, based on the findings of Thompson (1994, 2000), we explored whether truth table performance could be predicted by conditional probability ratings computed on the basis of the probability estimates task. Specifically, Thompson reported that responses to TF were related to ratings of perceived sufficiency, while perceived necessity predicted responses to FT. Recall that the constructs of sufficiency and necessity overlap substantially with ratings of P(*q*/*p*) and P(*q*/~*p*), respectively. Consistent with these previous findings, the correlation between P(*q*/*p*) and the likelihood of responding true to TF was -.35 (significant under one-tailed test). Thus, when participants judged that the occurrence of *p* was likely to result in *q*, they were less likely to indicate that the TF outcome supported the conditional statement. In addition, the correlation between P(*q*/~*p*) and the likelihood of responding true to the FT case was also significant ($r$ = .39). In this case, when *p* was seen as necessary for *q* [i.e., when P(*q*/~*p*) was low], the scenario in which the consequent occurs in the absence of *p* (i.e., FT) was less frequently classified as supporting the statement.

These findings show that the conditional probability approach can be extended to predict reasoning performance on the truth table evaluation task, and thus provide further support for a probabilistic interpretation of conditional statements (e.g., Evans & Over, 2004; Oaksford & Chater, 2001, 2003). Specifically, this experiment offers some evidence that conditional probability estimates [e.g., P(*q*/~*p*)] can be used to predict responses on the truth table task (e.g., FT). It is also clear, however, that the conditional probability approach corresponds better to the conditional inference task than to the truth table task. For instance, the correlations between inference rates and relevant conditional

probabilities in Experiment 3 (mean $r = .74$) were substantially higher than the correlations just reported for truth table responses in the current study.

Why is the relationship between the conditional probability approach and the truth table task not as strong as it is for the inference task? One possibility is that it may be more difficult to map conditional probabilities onto truth table cases than onto the different inference types. For example, the conditional probability of *q* given *p* (i.e., suppose *p*; how likely is *q*?) seems to map fairly directly onto the MP inference (i.e., suppose *p*; can we conclude *q*?), but does not seem to correspond to any of the outcomes of the truth table (e.g., *p* occurs and *q* occurs).

Another factor may be that there is virtually no variability to explain for two of the truth table cases. Specifically, there is usually very little disagreement about the TT and TF outcomes, which are overwhelmingly classified as true and false, respectively. Such ceiling effects are not as problematic for the inference task, where variability can be much greater. Even in the case of MP, where acceptance rates can be close to 100% in abstract tasks (Evans et al., 1993), inference rates can vary much more when realistic material is introduced (e.g., Newstead et al., 1997).

*Conclusions*

In summary, this experiment provides additional support for the dissociation between perceived truth and perceived effectiveness identified in Experiment 3. We first replicated the finding that probability judgments of the truth of conditional inducements and advice are unrelated to their effectiveness in changing behaviour; furthermore, truth was strongly predicted by P(*q*/*p*), while effectiveness was negatively correlated with P(*q*/~*p*). In other words, we replicated the counter-intuitive finding that the degree to

which these conditionals are perceived to be effective in changing *p* does not seem to depend on whether performing *p* is likely to result in *q*, nor on how likely the speaker is perceived to be telling the truth.

This pattern was also found when we attempted to directly manipulate these estimates. Specifically, reducing the credibility of conditional tips and promises decreased estimates of P(*q*/*p*); while this manipulation reduced truth ratings, it nevertheless failed to reduce ratings of effectiveness. It was also suggested, however, that a stronger manipulation may be needed to reveal these intuitive relationships. Finally, we showed that the conditional probability approach (e.g., Evans & Over, 2004) can be used to explain performance on another formal reasoning task, the truth table evaluation task. At the same time, it is also clear that this approach was more successful in predicting the conditional inference data observed in Experiment 3.

## Experiment 5

Thus far, there is little evidence to suggest that the effectiveness of conditional inducements and advice in changing behaviour is tied to the probability that they are perceived to be true. In two experiments, we have found that effectiveness ratings do not correlate with truth ratings; furthermore, these two measures are affected by different variables, such that truth is highly correlated with estimates of P(*q*/*p*), while effectiveness is predicted by P(*q*/~*p*). Finally, attempts to manipulate P(*q*/*p*) in Experiment 4 produced a corresponding change in truth ratings but failed to influence judgments of effectiveness. There are, however, reasons to suggest that this manipulation may not have been sufficiently strong. Specifically, reducing the credibility of the conditionals in

Experiment 4 reduced estimates of P(*q*/*p*) by an average of only 5%; moreover, the power to detect a change in effectiveness ratings with this manipulation was rather weak.

Consequently, the main goal of this experiment was to apply a stronger credibility manipulation. In doing so, we also address a potential explanation for the dissociation identified in our previous experiments. Specifically, one possible reason for why truth and effectiveness were found to be unrelated is that most of the conditionals used in Experiments 3 and 4 received relatively high truth ratings [as well as high estimates of P(*q*/*p*)]; it could be that these measures are in fact independent under these conditions. In other words, while a moderately high truth value would seem to be necessary for effectiveness (e.g., a promise will be ineffective when the speaker is perceived to be lying), other factors may be more important to the effectiveness of a conditional when its truth is *not* in doubt. For example, if the listener judges that the cost of performing *p* outweighs the benefit of *q*, or is indifferent about the reward offered, even a promise perceived to be truthful will likely by ineffective.

Although questioning the credibility of the conditionals in Experiment 4 reduced truth ratings, the statements in the low-credibility condition were still perceived to be more true than false (mean truth rating = 62%). Consider again the tip "If you show up early for work, you will impress your boss". While this statement may be judged as less true when uttered by a friend than by a colleague, it would seem to be a good piece of advice regardless of who the speaker is (i.e., showing up early for work will probably have the effect of impressing most employers, and it definitely would not hurt). Instead, what is needed is a manipulation where the credibility, and hence the truth, of conditional inducements and advice is questioned to a much greater degree.

To illustrate the manipulation adopted in the current study, the credibility of the above tip was reduced by changing it to "If you show up early for work, you will be promoted". There would seem to be good reason to doubt the truth of this advice. That is, while impressing one's boss may be a likely consequence of showing up for work early, being promoted would seem to be excessive. In this case, showing up early will, on its own, likely be insufficient reason for a promotion, as other factors relevant to work performance would seem to be more important (doing a good job, being respectful towards customers, not being tardy, etc.). Consequently, we predicted that manipulating credibility in this way would result in substantially lower estimates of P($q$/$p$). [12]

We further reasoned that this manipulation would enable a relationship between truth and effectiveness to emerge. That is, as in Experiment 4, decreasing P($q$/$p$) via this manipulation was expected to reduce truth ratings; unlike the results of Experiment 4, however, this manipulation was also predicted to decrease effectiveness ratings. In our example, since a promotion would seem to be an unlikely consequence of showing up for work early, we expected that this tip would not only receive a low truth rating, but the listener should also become less likely to follow the advice.

For another illustration of this manipulation, P($q$/$p$) should decrease substantially when replacing the consequent of the promise "If you wash the car, I will give you $10" with a $200 reward. That is, $200 would seem to be an excessive reward for washing someone's car, and doubts should arise as to whether the speaker is likely to keep such a promise. While this manipulation should reduce perceived truth, it should also result in a

[12] In fact, there is evidence from Experiment 3 to support this prediction. Recall that we attempted to obtain a wide range of probability estimates by creating three consequents for each conditional. One of the statements used in Experiment 3 was exactly the tip described here; changing the consequent from "you will impress the boss" to "you will be promoted" was found to reduce P($q$/$p$) from 73% to 53%.

decreased effectiveness rating (i.e., the listener should be less willing to wash the car with a $200 reward, since this promise is unlikely to come true). In contrast, if truth and effectiveness in fact are independent (as Experiments 3 and 4 seem to indicate), then the credibility manipulation would be expected to, once again, only affect truth ratings.

In addition to re-examining the relationship between ratings of truth and effectiveness, another goal of this experiment was to explore a different aspect of conditional inducements and advice. One could argue that the essence of conditional advice, for instance, is not accurately or fully captured by our measure of behavioural effectiveness; indeed, the results from Experiments 1-4 indicate that neither tips nor warnings are perceived to be very effective in changing the listener's behaviour. Instead, the meaning of advice may better be captured by assessing their quality (e.g., is this tip a good piece of advice?). Consequently, participants in this experiment were also asked to rate how "good" they perceived conditional tips to be (for a similar type of judgment, see Over, Manktelow, & Hadjichristidis, 2004).

What about conditional promises? While our measure of effectiveness seems to reflect the purpose of promises fairly well, asking for judgments of their quality may still shed light on a different aspect of these speech acts. Note, however, that the quality of a promise does not seem to be captured by asking how good it is; instead, participants were asked to judge how "reasonable" they perceived conditional promises to be. Intuitively, a reasonable promise would seem to be one in which the proposed exchange is perceived to be fair (i.e., a fair balance between the cost of performing *p* and the benefit of receiving *q*).

What might we predict for these ratings of quality? First, we might expect that our credibility manipulation will have an impact on quality judgments. For example, the "car wash" promise should be judged less reasonable with a $200 reward, as this benefit would seem to vastly exceed the cost of washing the car. Likewise, a low-credibility tip should be judged as less good than a high-credibility tip. Furthermore, since the credibility manipulation was intended to affect P($q$/$p$), quality ratings may also be tied to estimates of this conditional probability. Specifically, when performing $p$ is relatively unlikely to result in $q$ [i.e., when P($q$/$p$) is low], tips should be judged as less good and promises as less reasonable.

Finally, there are also reasons to suggest that quality ratings may be correlated with our measures of truth and effectiveness. When doubts are raised about the truthfulness of a tip, for example, the degree to which it is judged to be good should also be reduced. Similarly, since a listener may comply with a promise only when it sounds reasonable, we might expect this judgment to be tied to the degree to which promises are effective.

To summarize, this experiment used a stronger credibility manipulation in the hope of reducing estimates of P($q$/$p$) to a greater degree than in Experiment 4. Using the same methodology, we in turn examined whether this manipulation only affected truth ratings (as in Experiment 4), or also resulted in decreased effectiveness ratings. Finally, in order to capture a different aspect of conditional tips and promises, we obtained ratings of quality by asking participants to judge how good (in the case of tips) or reasonable (in the case of promises) they perceived these conditionals to be.

## Method

### *Participants*

Forty-nine undergraduate students at the University of Saskatchewan participated in this experiment in partial fulfillment of an introductory psychology course requirement.

### *Materials*

Eight scenarios (four tips and four promises) were used in this experiment. From these eight scenarios, 16 conditionals were constructed by creating two consequents for each antecedent. These two consequents formed the two credibility conditions. To illustrate, the credibility of the promise "If you wash the car, I will give you $10" was manipulated by changing the consequent to "I will give you $200"; this latter statement referred to the low-credibility condition. For a full list of the conditional statements, refer to Appendix F.

### *Design and Procedure*

This experiment used a 2x2 within-subjects design, with factors of statement type (promise vs. tip) and credibility (low vs. high). Each participant was presented with all eight scenarios; for each statement type, participants received two high-credibility and two low-credibility conditionals. The order of the eight statements was randomized for each participant.

Participants were asked to complete three tasks for each statement: the behavioural effectiveness task, the probability estimates task, and the quality ratings task. The methodology of the first two tasks was identical to that of Experiment 4. Thus, in the behavioural effectiveness task, two likelihood ratings of *p* were obtained, one before and

one after presentation of the conditional statement. For these two ratings, an 11-point scale with values between 0 and 100 was used. The difference between these two scores gives an indication of how effective the conditional statement is perceived to be in changing the behaviour described in *p*.

In the probability estimates task (Over & Evans, 2003), participants were first asked to rate the probability that the conditional statement was true, and then to rate the likelihood of the four possible truth-table cases (i.e., *pq*, *p*~*q*, ~*pq*, and ~*p*~*q*). As before, they were instructed to read through all four events before providing final estimates, and to ensure that these estimates added up to 100% for each statement.

For the quality ratings task, participants completed one question. In the case of conditional tips, they were asked to judge how good the advice given by the speaker was. In the case of conditional promises, participants instead were asked how reasonable they judged the speaker's utterance to be. In both cases, an 11-point Likert scale with values between 0 (very poor/unreasonable) and 100 (very good/reasonable) was used.

For all statements, participants completed the behavioural effectiveness task first, the probability estimates task second, and the quality ratings task last. The four cases of the probability estimates task were presented in two quasi-random orders; for each participant, the standard order (i.e., *pq*, *p*~*q*, ~*pq*, ~*p*~*q*) was presented for half of the statements, while the cases were given in the reversed order for the remaining four statements. Participants were instructed to complete the questions in the order they appeared and to work at their own pace. Participants were recruited from one large class; roughly half the students participated in this experiment, while the remaining took part in Experiment 6.

## Results and Discussion

This section is divided into three parts: replicating previous findings, exploring the credibility manipulation, and investigating ratings of quality. Alpha was in all these analyses set at .05.

*Replication of previous findings*

As before, a difference score between the two likelihood ratings of *p* (i.e., before and after presentation of the conditional) was computed to obtain a measure of behavioural effectiveness. Consistent with previous findings showing how inducements are more effective in changing behaviour than advice, conditional promises (mean score = 37.14) were given higher effectiveness ratings than conditional tips (mean score = 9.61), $F(1, 47) = 185.26$, $MSE = 196.31$. Notice, also, that the difference between tips and promises was substantially larger in this experiment than in Experiment 4; one reason for this difference is that the credibility manipulation actually *increased* effectiveness ratings for promises (e.g., contrary to predictions, a $200 reward was more effective than a $10 reward in the "car wash" promise; see below).

Next, we computed estimates of P(*q*/*p*) and P(*q*/~*p*) from the ratings on the probability estimates task. Table 12 shows the correlations between these two conditional probabilities and ratings of truth and effectiveness. An inspection of this table reveals that we replicated the basic pattern from Experiments 3 and 4. Specifically, while P(*q*/*p*) strongly predicted truth ratings, effectiveness ratings were instead (negatively) correlated with P(*q*/~*p*). Moreover, ratings of truth and effectiveness were again not correlated (r = -.09). Collectively, these findings provide strong support for the apparent dissociation between the probability that conditional inducements and advice are judged to be true and

Table 12

*By-item correlations (N = 16) between computed conditional probabilities, truth ratings, and behavioural effectiveness ratings in Experiment 5*

| Conditional probability | Truth ratings | Effectiveness |
|---|---|---|
| P(*q*/*p*) | .86** | .15 |
| P(*q*/~*p*) | .34 | -.48* |

Note. * $p < .05$, ** $p < .01$.

their effectiveness in changing behaviour.

*Manipulating credibility*

To explore the effects of our credibility manipulation, we performed a series of 2x2 repeated measures ANOVAs with credibility (high vs. low) and statement type (tip vs. promise) as factors. These analyses were performed on four dependent variables: P(*q*/*p*), P(*q*/~*p*), truth, and behavioural effectiveness. Table 13 presents the ratings on these variables across the two credibility conditions. [13]

There was a significant main effect of credibility on P(*q*/*p*), $F(1, 43) = 27.43$, $MSE = 603.32$. Specifically, estimates of P(*q*/*p*) were lower for low-credibility conditionals than for high-credibility conditionals; furthermore, an inspection of Table 13 reveals that the magnitude of this difference was much greater than in Experiment 4 (19% vs. 5%). Consistent with this reduction in ratings of P(*q*/*p*), low-credibility conditionals also received lower truth ratings, $F(1, 41) = 26.34$, $MSE = 565.46$. The mean truth rating for these conditionals was below 50%, suggesting that the credibility manipulation in this experiment was indeed successful in making participants doubt the truth of these statements.

The main objective of this experiment was to examine whether a stronger credibility manipulation would also influence ratings of behavioural effectiveness. However, despite the success of this manipulation in substantially reducing estimates of P(*q*/*p*) and truth, effectiveness ratings were not found to be lower in the low-credibility

[13] Statement type was again deemed to be of less importance in this study. As previously reported, effectiveness ratings were higher for promises; given than effectiveness correlated negatively with P(*q*/~*p*), computed estimates of this conditional probability were also lower for promises (16.03) than for tips (22.70), $F(1, 41) = 6.48$, $MSE = 288.82$. However, the main effect of statement type was not significant on ratings of truth, $F(1, 41) = .51$, $MSE = 407.16$, or P(*q*/*p*), $F(1, 43) = 2.75$, $MSE = 197.40$. Finally, the interaction between statement type and credibility was significant for only one of these DVs, namely effectiveness, $F(1, 47) = 4.36$, $MSE = 213.36$.

Table 13

*Ratings on P(q/p), P(q/~p), truth, and behavioural effectiveness as a function of credibility in Experiment 5*

| Credibility | P(*q*/*p*) | P(*q*/~*p*) | Truth | Effectiveness |
|---|---|---|---|---|
| High | 68.41 | 21.55 | 65.79 | 21.35 |
| Low | 49.02 | 17.18 | 46.96 | 25.39 |

condition than in the high-credibility condition, $F(1, 47) = 2.88$, $MSE = 271.82$ ($p = .10$). In fact, there was a trend in the opposite direction, that is, for low-credibility conditionals to be perceived as *more* effective. In other words, the degree to which conditional tips and promises were effective in changing the listener's behaviour did not decrease for conditionals where the probability of obtaining the positive outcome in $q$ (upon the fulfillment of $p$) was at chance level, or, alternatively, for conditionals that were perceived to be just as likely false as true!

This picture is, however, complicated somewhat by the finding that the credibility manipulation also had an effect on estimates of P($q$/~$p$), $F(1, 41) = 4.98$, $MSE = 160.61$. That is, unlike Experiment 4 where credibility was found to selectively (albeit weakly) affect P($q$/$p$), reducing credibility in this experiment also reduced P($q$/~$p$) ratings. To illustrate, suppose that the listener in the "car wash" promise fails to wash the car (i.e., ~$p$). In this scenario, participants judged the chances that the reward $q$ would still come about as better in the high-credibility condition (i.e., $10) than in the low-credibility condition (i.e., $200), suggesting that it may have been easier to think of counterexamples of the form ~$p$ *and* $q$ for high-credibility conditionals (e.g., a "soft" speaker may decide to give $10 even though the listener fails to wash the car, an outcome less likely to occur with a $200 reward).

The finding that credibility also affected P($q$/~$p$) does, however, raise a potential problem. Specifically, since P($q$/~$p$) correlates negatively with effectiveness, we would in fact expect a reduction in P($q$/~$p$) to result in a corresponding increase in effectiveness ratings, a trend we do observe in this experiment. This pattern of results raises the

possibility that reducing credibility may have failed to reduce effectiveness ratings because this manipulation decreased estimates of P($q$/~$p$) as well as P($q$/$p$).

However, this possibility is given less credence when examining tips and promises separately. Follow-up t-tests revealed that the credibility manipulation decreased P($q$/~$p$) ratings only for tips, $t(43) = 2.21$, $SE = 3.28$, and not for promises, $t(44) = .86$, $SE = 2.38$. Yet, this manipulation was found to increase effectiveness ratings only for promises, $t(48) = 2.18$, $SE = 3.69$, and had no impact on the effectiveness of tips, $t(47) = .15$, $SE = 2.48$. Thus, it seems unlikely that the overall reduction in P($q$/~$p$) ratings for low-credibility conditionals was the reason that the credibility manipulation failed to reduce effectiveness ratings.

In addition, we performed an ANCOVA (through the SPSS MANOVA program) to explore the effects of the credibility manipulation on effectiveness ratings after adjusting for differences on P($q$/~$p$). For this analysis, we collapsed ratings of effectiveness and P($q$/~$p$) across statement type. A one-way repeated-measures ANCOVA with credibility as IV, effectiveness as DV, and the two levels of P($q$/~$p$), corresponding to the high and low credibility conditions, as covariates was performed (see Tabachnick & Fidell, 2001a, 2001b, on how to deal with repeatedly measured covariates). Holding estimates of P($q$/~$p$) constant, the manipulation of credibility had a significant impact on effectiveness ratings, $F(1, 40) = 4.20$, $MSE = 145.42$, $p = .047$; contrary to predictions, however, low-credibility conditionals were in fact given *higher* effectiveness ratings than high-credibility conditionals. Thus, we conclude that the failure of the credibility manipulation in reducing effectiveness ratings was not due to the concurrent effects of this manipulation on P($q$/~$p$).

On the basis of Experiments 3-5, then, participants' effectiveness ratings of inducements and advice do not seem to be tied to their probability judgments of the truth of these conditionals. That is, we have consistently observed that ratings of truth and effectiveness are not correlated; moreover, truth and effectiveness are affected by different variables [i.e., P($q$/$p$) and P($q$/~$p$)]. Even when participants in the current experiment judged conditionals to be just as likely false as true (i.e., 50%), they did not evaluate these conditionals to be any less effective than those whose truth was in much less doubt. We have previously emphasized the counter-intuitive nature of the apparent dissociation between truth and effectiveness; yet, the consistency of these findings suggests that this dissociation is reliable and in need of an explanation.

One possible explanation is that participants may be interpreting truth and effectiveness on two distinct levels. In fact, an examination of the wording of our questions may provide some support for this suggestion. Specifically, when judging truth, participants are asked to give an estimate reflecting their own opinion; when evaluating effectiveness, in contrast, they are asked (twice) about the likelihood that another person (i.e., the listener) will perform some action (i.e., the behaviour in $p$). Although this explanation would seem to attribute rather shallow processing on the part of our participants, a relationship between truth and effectiveness may be more likely to emerge if they were instead asked to reason on these questions from the same perspective (i.e., judge truth from the listener's perspective or evaluate effectiveness from their own).

Evidence that such perspective shifts can matter was recently provided by Thompson, Evans, and Handley (2005) in a study investigating conditionals they referred to as "persuasions" and "dissuasions". In this study, a number of differences were found

in the way participants responded depending on the specific perspective they were asked to take. For instance, participants who were asked to reason according to what they thought the writer intended to communicate often responded differently to invited inferences (e.g., if you want to achieve *q*, you should do *p*) than those reasoning from their own perspective.

Another possibility suggested by Experiment 5 is that even though the outcome (i.e., *q*) of low-credibility conditionals is unlikely to come true, there may still be some value or utility in performing *p*. Consider again the tip "If you show up early for work, you will be promoted". While there are good reasons to doubt that performing *p* will bring about *q* in this tip, showing up early may still be a good idea, in that other (more realistic) benefits may result (e.g., impressing the boss, getting a good parking spot). Likewise, while there are doubts that a promise such as "If you wash the car, I will give you $200" will come true, the speaker may still somehow reimburse the listener who performs *p* (e.g., by giving a lesser reward), especially if another encounter between the two is anticipated. Under this interpretation, then, any residual utility of performing *p* may partly explain how perceptions of effectiveness can dissociate from ratings of truth. This explanation is, however, contingent on the assumption that people spontaneously apply a substantial amount of background beliefs and extraneous information when interpreting inducements and advice, as these conditionals would need to cue or remind participants about any unstated benefit (i.e., other than *q*) associated with performing *p*.

*Quality ratings*

Finally, we present the results of participants' quality ratings. Recall that the quality of a conditional tip was assessed by asking how good it was; the quality of a

conditional promise was assessed by asking for its reasonableness. High estimates on these questions corresponded to high quality ratings.

As predicted, the credibility manipulation had an impact on these quality judgments. Specifically, low-credibility tips (mean rating = 47.11) were perceived to be less good than high-credibility tips (mean rating = 69.00), $t(44) = 5.19$, $SE = 4.22$. Similarly, the degree to which promises were judged to be reasonable decreased from the high-credibility condition (mean rating = 62.78) to the low-credibility condition (mean rating = 47.56), $t(44) = 4.48$, $SE = 3.40$. In other words, when the credibility of these conditionals was reduced, judgments of their quality also decreased.

Consistent with this effect of credibility, quality judgments were also found to correlate highly with computed estimates of $P(q/p)$. Thus, a tip ($r = .84$) was perceived to be good, and a promise ($r = .83$) to be reasonable, to the extent that performing $p$ was likely to result in $q$. Furthermore, quality ratings correlated highly with truth ratings (overall $r = .94$). In other words, when doubts are raised about the truthfulness of these statements, tips are judged to be less good and promises as less reasonable. The high coefficient of this correlation in fact suggests a substantial degree of overlap between the constructs of truth and quality.

However, quality ratings did not correlate with effectiveness ratings (overall $r = -.15$). Thus, similar to the dissociation identified between truth and effectiveness, the degree to which conditional tips and promises are perceived to be effective does not seem to depend on how good or reasonable they are judged to be. Again, this finding is somewhat counter-intuitive; for example, it seems strange to claim that a listener in a

promise will incur the cost of *p* regardless of how reasonable or fair the proposed exchange is perceived to be.

Yet, this finding may not be that surprising after all. An inspection of Appendix F reveals that the low-credibility promises used in this experiment seem to be unreasonable from the speaker's, rather than the listener's, perspective. In the "car wash" promise, for example, the cost of giving *q* (i.e., $200) outweighs the benefit of receiving *p* (i.e., car wash) for the speaker; this exchange (if it were to occur) actually describes a good deal for the listener. Given that effectiveness is measured from the listener's perspective, maybe the dissociation between quality and effectiveness is not that odd. Furthermore, given the high degree of overlap between truth and quality, we might have expected a similar pattern of results for these two constructs [i.e., high correlations with P(*q*/*p*) and low correlations with effectiveness].

*Conclusions*

The results of Experiment 5 provide additional support for the dissociation between the perceived truth of inducements and advice and the effectiveness of these conditionals in changing behaviour. As in Experiments 3 and 4, ratings of truth and effectiveness did not correlate, and were affected by different variables [i.e., P(*q*/*p*) and P(*q*/~*p*), respectively]. Furthermore, despite the success of the credibility manipulation in substantially reducing estimates of P(*q*/*p*) and truth, participants did *not* judge low-credibility conditionals to be any less effective than high-credibility conditionals. Finally, by asking participants to rate the quality of these conditionals, we found that the degree to which tips are perceived to be good, and promises to be reasonable, is highly correlated with ratings of truth, but unrelated to ratings of effectiveness.

# EXPERIMENT 6

## EXPLORING INDUCEMENTS AND ADVICE USING A DECISION-THEORETIC ANALYSIS

The principal methodological tool used in Experiments 3-5 consisted of asking participants for likelihood ratings of the four possible outcomes defined by the truth table (i.e., *pq*, *p*~*q*, ~*pq*, and ~*p*~*q*), from which conditional probability estimates could be computed (Over & Evans, 2003). The results of these experiments demonstrated that this task can be used to predict a number of judgments people make about conditional inducements and advice, including their effectiveness in changing behaviour, the probability that they are true, how good or reasonable they are perceived to be, and the reasoning patterns they invite on formal deductive tasks. Despite the undeniable utility of this methodology, people's representation of such pragmatic conditionals obviously involves more than these fairly simple likelihood estimations. The aim of this sixth, and final, experiment was therefore to investigate people's understanding of inducements and advice from a different perspective, and by using a different task.

Specifically, in this study we adopt a *decision theoretic analysis* (e.g., Manktelow & Over, 1991, 1995; Over, Manktelow, & Hadjichristidis, 2004). This analysis is grounded in the principles of *subjective expected utility* (SEU) theory, the accepted normative theory of decision making (see Baron, 2000, 2004; Manktelow, 1999). Formally, this theory states that when deciding among several options, each option should be assessed with reference to both the probability (i.e., the likelihood) and the utility of its outcome. Judgments of utility, furthermore, are said to depend on people's perceptions of expected costs and benefits, which in turn influence the subjective

*preferences* people have among possible outcomes. In this experiment, we therefore explored participants' perceptions of costs and benefits by asking them to provide preference judgments to different possibilities.

The main motivation for this experiment was a recent study by Over et al. (2004). Participants in this study were given two types of deontic conditionals: obligations (e.g., "If the fire alarm goes off, then you must leave the building") and permissions (e.g., "If the fire alarm goes off, then you may leave the building"). Over et al. were interested in exploring the conditions under which these conditionals are interpreted as "acceptable" rules for guiding behaviour. To this end, these researchers proposed a move away from the conditional probability hypothesis (e.g., Evans, Handley, & Over, 2003; Evans & Over, 2004; Oberauer & Wilhelm, 2003; Over & Evans, 2003) and towards what they termed the "conditional expected utility hypothesis".

Participants in Over et al.'s (2004) study were first asked to rate the "goodness" of conditional obligations and permissions. Subsequently, they were presented with the four possible outcomes of *p* and *q*; however, instead of providing likelihood estimates to these possibilities (as per Over & Evans, 2003), participants were asked to arrange them in order of preference (henceforth called *the preference ratings task*). In this task, they were instructed to give a score of 1 to the outcome that described the best situation, a score of 2 to the next best situation, and so forth. Over et al. reasoned that there would be a relationship between these preference ratings and the degree to which obligations and permissions were judged to be good.

Specifically, Over et al. (2004) predicted that these rules would be judged good to the extent that the outcome where both the antecedent and consequent terms come about

(i.e., *pq*) was preferred to the outcome where only the antecedent is fulfilled (i.e., *p*~*q*). The results of two experiments provided strong support for this prediction. In the "fire alarm" obligation, for example, goodness ratings were higher when the situation where the fire alarm goes off and people leave the building (i.e., *pq*) was preferred to the situation where the alarm goes off but people do not leave the building (i.e., *p*~*q*). In other words, this conditional is an acceptable rule for guiding behaviour because the expected costs of people staying inside during a fire alarm (e.g., poor evacuation procedures, injury/death in the case of a real fire) are perceived to outweigh any potential benefits (e.g., convenience, continued work).

Can we apply this "conditional expected utility hypothesis" (Over et al., 2004) to conditional inducements and advice? Intuitively, an analysis based on notions such as subjective preferences, costs, and benefits would seem to offer great promise in capturing some of the dimensions explored in our previous experiments. Thus, we reasoned that preference ratings could feasibly be tied to such judgments as the perceived effectiveness and quality of inducements and advice. However, we also argue that the truth table possibilities of most relevance to these conditionals are different from those emphasized by Over et al. Specifically, whereas Over et al. focused on preference ratings of *pq* and *p*~*q* when evaluating conditional obligations and permissions, we propose that the outcomes most important to inducements and advice instead are *pq* and ~*p*~*q*. [14]

To illustrate, consider the promise "If you wash the car, I will give you $10". For this promise to be effective, the listener must perceive that the benefit of receiving $10

[14] In fact, one would expect very little variation in preference ratings for the other two truth table outcomes (i.e., *p*~*q* and ~*pq*). Consider a promise such as "If you wash the car, I will give you $10". For the listener, ~*pq* (i.e., receiving the reward without washing the car) will always represent the best case scenario, while *p*~*q* (i.e., washing the car without receiving the money) will be the least preferred outcome. This pattern is reversed for the speaker, such that *p*~*q* will be the most, and ~*pq* the least, preferred outcome.

exceeds, or at least equals, the cost incurred by washing the car. Indeed, if the perceived cost of *p* outweighs the perceived benefit of *q*, the listener may instead decide to forgo the reward by refraining from performing *p*. In other words, for this promise to be an effective inducement, the listener should attach a higher subjective expected utility (SEU) to washing the car and receiving $10 (i.e., *pq*) than to doing neither (i.e., ~*p*~*q*). Translating this scenario to Over et al.'s (2004) preference ratings task, we would predict that conditional inducements and advice will be judged effective in changing the listener's behaviour to the extent that *pq* is preferred to ~*p*~*q*.

To see how these preference ratings can also capture the perceived quality of conditional inducements and advice, consider the tip "If you arrive 10 minutes early for work, you will get a good parking spot", uttered by a colleague to a new employee. This utterance will only be judged a good tip if the cost of arriving early for work does not outweigh the benefit of a good parking spot, that is, when the listener attaches a higher SEU to arriving early in order to get a good parking spot (i.e., *pq*) than to relinquishing the parking spot by arriving later (i.e., ~*p*~*q*). This analysis can again be translated to performance on the preference ratings task by predicting that a tip will be judged a good piece of advice (and a promise a reasonable inducement) to the extent that *pq* is preferred to ~*p*~*q*.

As previously discussed, we can normally assume that the listener of a promise or a tip will attach a higher subjective expected utility to *pq* than to ~*p*~*q*, and consequently will prefer the former to the latter possibility. In an attempt to manipulate this particular preference pattern, we therefore created a second condition by changing the content of each conditional statement. To illustrate, the promise "If you wash the car, I will give you

$10" was changed by replacing the consequent with a $1 reward. Whereas $10 may seem to be an appropriate reward for washing your brother's car, a $1 reward would seem to be insufficient (and perhaps somewhat insulting). That is, the perceived benefit of $1 is surely outweighed by the cost of washing the car. In this "reversed" condition, we reasoned that the listener will attach a *lower* subjective expected utility to *pq* (i.e., washing the car and receiving the reward) than to ~*p*~*q* (i.e., forgo the reward by not washing the car), and consequently, will prefer the ~*p*~*q* possibility to *pq*. As a result of this preference reversal, we further predicted that ratings of both effectiveness and quality would decrease in this condition.

The goal of this study was thus to apply a decision theoretic analysis to the interpretation of conditional tips and promises, expecting that judgments of effectiveness and quality would be related to performance on the preference ratings task (Over et al., 2004). As before, behavioural effectiveness was measured by obtaining two ratings of the likelihood of *p* (i.e., with and without knowledge of *q*); quality was again assessed by asking for the "goodness" of tips and the "reasonableness" of promises. Preference ratings were obtained by instructing participants to arrange the four possible outcomes (i.e., *pq*, *p*~*q*, ~*pq*, ~*p*~*q*) in order of preference. Note that participants in Over et al.'s experiments either gave preference ratings from their own, the speaker's, or the listener's perspective; all participants in the current study were asked to complete the preference ratings task from the listener's perspective.

## Method

### *Participants*

A total of 32 University of Saskatchewan undergraduate students participated in partial fulfillment of an introductory psychology course requirement.

### *Materials*

As in Experiments 4 and 5, only conditional inducements and advice specifying positive outcomes (i.e., promises and tips) were used in this experiment. A total of 16 conditional statements (eight promises and eight tips) were constructed on the basis of eight basic scenarios. For each scenario, two conditions were created: "normal" and "reversed". The latter condition described a situation in which we expected a preference reversal for the *pq*/~*p*~*q* pair.

The reversed condition was created differently for conditional promises and conditional tips. For promises, the conditional was changed by substituting the consequent term. To illustrate, the reversed condition of the normal promise "If you wash the car, I will give you $10" was constructed by replacing the consequent with a $1 reward. In contrast, creating reversed conditional tips was achieved by replacing the antecedent term. For example, the antecedent of the tip "If you arrive 10 minutes early, you will get a good parking spot" was changed to "If you arrive 1 hour early". In this example, we reasoned that the cost of arriving a full hour early would be perceived as exceeding the benefit of a good parking spot, such that ~*p*~*q* would more likely be preferred to *pq*. Appendix G gives a complete list of these materials.

*Design and Procedure*

Every participant received eight conditional statements, four of each statement type (i.e., tips and promises). Furthermore, for each statement type, both conditions of our attempted preference manipulation (i.e., normal and reversed) were presented twice. Hence, both of these factors were manipulated within-subjects. The order of the eight statements was randomized for each participant.

Participants were asked to complete the following tasks for each conditional. First, they completed the behavioural effectiveness task by providing two likelihood ratings of *p* occurring (one before and one after knowledge of *q*). For these ratings, an 11-point scale with values between 0 and 100 was again used. After this task, participants were also asked to estimate the probability that the conditional statement was true (Evans et al., 2003).

Next, participants completed the preference ratings task (Over et al., 2004). In this task, participants were given the four possible outcomes of the truth table and were asked to arrange them in order of preference from the listener's perspective. For the tip "If you arrive 10 minutes early, you will get a good parking spot", these outcomes read:

Rachel arrives 10 minutes early and gets a good parking spot (*pq*)
Rachel arrives 10 minutes early and does not get a good parking spot (*p*~*q*)
Rachel does not arrive 10 minutes early and gets a good parking spot (~*pq*)
Rachel does not arrive 10 minutes early and does not get a good parking spot (~*p*~*q*)

The instructions to this task were, for this particular conditional, as follows:

"Please rate the following four situations in order of preference from **Rachel's** perspective, using the numbers 1 to 4. Give a score of 1 to the situation that you think Rachel will **most** prefer; a score of 2 to the situation Rachel will next prefer; and so on."

The next task, the quality ratings task, was identical to that of Experiment 5. Specifically, using an 11-point scale with values between 0 and 100, participants were asked to rate how good conditional tips were, and how reasonable conditional promises were.

Finally, participants were requested to give direct estimates of P(*q*/*p*) and P(*q*/~*p*). Recall from Experiment 3 that these questions refer to perceived sufficiency and necessity, respectively. For the above tip, for example, these questions read as follows:

> P(*q*/*p*): Suppose Rachel arrives 10 minutes early. How likely do you think it is that she will get a good parking spot?
>
> P(*q*/~*p*): Suppose Rachel does *not* arrive 10 minutes early. How likely do you think it is that she will get a good parking spot?

For these questions, a likelihood scale between 0 (very unlikely) and 100 (very likely) was used.

These tasks were always presented in the order just described. However, the four outcomes in the preference ratings task were presented in one of two quasi-random orders. Each participant received the standard order (as given above) for half of the

statements, while these outcomes were given in the reversed order for the remaining four statements.

Testing took place in one large class simultaneously with Experiment 5. Participants were instructed to complete the questions in the given order and to work at their own pace; most were finished within 20 minutes.

## Results and Discussion

### *Correlational analyses*

In order to examine whether the main findings of previous experiments were replicated, we first present the by-item correlations between direct conditional probability estimates and ratings of truth, effectiveness and quality. In this experiment, estimates of P(*q*/*p*) and P(*q*/~*p*) were obtained by asking for sufficiency and necessity ratings, respectively (see also Experiment 3).

Table 14 reveals that we replicated previous findings regarding ratings of P(*q*/*p*). Specifically, consistent with Experiments 3-5, P(*q*/*p*) correlated very highly with truth ratings, but was unrelated to effectiveness ratings. Furthermore, as in Experiment 5, P(*q*/*p*) correlated highly with quality ratings. This pattern also held when analyzing quality ratings separately for tips and promises: the degree to which conditional tips were judged to be good correlated .63 with estimates of P(*q*/*p*), while the corresponding correlation for reasonableness ratings of promises was .52 (although this latter value was non-significant due to a correlational analysis based on only 8 items).

Also consistent with previous findings, the correlation between ratings of truth and effectiveness did not reach significance ($r$ = .29), providing further support for the dissociation between these two measures. Moreover, while truth ratings and quality

Table 14

*By-item correlations (N = 16) between computed conditional probabilities, truth ratings, behavioural effectiveness ratings, and quality ratings in Experiment 6*

| Conditional probability | Truth | Effectiveness | Quality |
|---|---|---|---|
| P(*q*/*p*) | .78** | .08 | .66** |
| P(*q*/~*p*) | .08 | .15 | -.23 |

Note. * $p < .05$, ** $p < .01$, one-tailed tests. Quality refers to goodness for tips, and reasonableness for promises.

ratings were positively correlated (overall $r$ = .75), judgments of quality did not correlate with effectiveness ratings (overall $r$ = .07). These findings suggest that whereas the probable truth and the quality of these conditionals overlap substantially, these constructs are both independent of the degree to which inducements and advice are perceived to be effective.

The only difference in this study from previous experiments concerned ratings of P($q$/~$p$). Specifically, P($q$/~$p$) did *not* correlate negatively with effectiveness ratings, in contrast to Experiments 3-5. There are no straightforward explanations for this inconsistency. For example, this pattern held for both conditional tips and conditional promises. In addition, the range in P($q$/~$p$) ratings was similar to that of previous experiments.

*Preference data*

Preference ratings were overall quite stable. Collapsed across the normal and reversed conditions, the dominant choice for the most preferred outcome was ~$pq$ (83%), while $p$~$q$ was most frequently judged to be the least preferred outcome (84%). Thus, from the listener's perspective, the best scenario is almost always the one in which the cost in $p$ is not performed yet the benefit in $q$ still comes about; the worst case is where the benefit $q$ is not fulfilled even though $p$ is. The dominant choices for the second and third best outcomes were $pq$ (75%) and ~$p$~$q$ (74%), respectively.

Of more importance, however, was whether our preference manipulation was successful. Recall that we expected a preference reversal for the $pq$/~$p$~$q$ pair to occur in the reversed condition. In general, preference ratings to these outcomes were found to be quite resistant to change. Nevertheless, the percentage of responses where $pq$ was

preferred to ~*p*~*q* differed significantly between the normal condition (93.65%) and the reversed condition (75.59%), $\chi^2(1) = 15.82$. This pattern was found for both conditional tips (96.83% vs. 78.13%), $\chi^2(1) = 10.04$, and conditional promises (90.48% vs. 73.02%), $\chi^2(1) = 6.43$. Thus, although the preference manipulation did not result in an actual *reversal* in preference ratings, there was still a significant decrease in the tendency to prefer *pq* to ~*p*~*q* in the reversed condition.

Further support for this conclusion is provided by an analysis that takes into account the magnitude, and not just the direction, of the preference scores for *pq* and ~*p*~*q*. Specifically, we computed a difference score by subtracting *pq* from ~*p*~*q*; given that a high number on the preference ratings task signified a low preference to that outcome, a high difference score (~*p*~*q* – *pq*) thus reflects a greater preference for *pq*. This difference score was higher in the normal condition (mean = 1.03) than in the reversed condition (mean = 0.51), $t(28) = 3.73$, $SE = .14$. In other words, the preference manipulation was successful in reducing the degree to which *pq* was preferred to ~*p*~*q*; we are therefore justified in exploring the influence of this manipulation on ratings of effectiveness and quality.

*Effectiveness ratings*

Table 15 shows effectiveness and quality ratings across the two conditions of the preference manipulation, separately for conditional tips and promises. To analyse effectiveness ratings, we performed a 2x2 repeated measures ANOVA with statement type (tip vs. promise) and preference ordering (normal vs. reversed) as factors. Neither the main effect of statement type, $F(1, 29) = 2.96$, $MSE = 310.96$, nor of preference ordering, $F(1, 29) = 3.18$, $MSE = 154.28$, reached significance, although an examination

Table 15

*Ratings of effectiveness and quality in Experiment 6 as a function of statement type and preference condition*

| | Effectiveness | | Quality | |
|---|---|---|---|---|
| Condition | Tip | Promise | Tip | Promise |
| Normal | 9.83 | 26.42 | 71.09 | 76.33 |
| Reversed | 16.83 | 11.33 | 49.53 | 65.50 |

Note. The quality of tips refers to their goodness, while the quality of promises refers to their reasonableness.

of the means in both cases reveals a trend in the expected direction (i.e., tip = 13.33 vs. promise = 18.88; normal = 18.13 vs. reversed = 14.08). These results are qualified, however, by the presence of a significant interaction, $F(1, 29) = 33.08$, $MSE = 110.57$.

In an attempt to decompose this interaction, paired-samples t-tests were performed separately for tips and promises. In the case of promises, effectiveness ratings were lower in the reversed condition than in the normal condition, $t(29) = 4.96$, $SE = 3.04$. Thus, as expected, the degree to which a promise is perceived to be effective in changing the listener's behaviour is reduced for conditionals where the listener is less likely to prefer $pq$ to $\sim p \sim q$. Using the "car wash" promise as an example, a $1 reward is less effective than a $10 reward because the cost of washing the car is more likely to be perceived as exceeding the benefit of the former than the latter reward.

An inspection of Table 15 reveals a very different picture for conditional tips, however. Contrary to predictions, effectiveness ratings were *higher* in the reversed condition than in the normal condition, $t(31) = 2.24$, $SE = 2.86$. Thus, although participants in this study were less likely to judge that the listener would prefer $pq$ to $\sim p \sim q$ in the reversed condition, these conditionals were perceived to be *more* effective in changing the listener's behaviour. To illustrate the paradoxical nature of these findings, consider the reversed tip "If you arrive 1 hour early, you will get a good parking spot". This tip was rated more effective than its normal counterpart (which instead recommends arriving 10 minutes early), even though the cost of arriving an hour early was more likely to be interpreted as exceeding the benefit of a good parking spot.

How can these findings be reconciled? One possibility is that the reversed tips are more effective because they are perceived to be more certain than their normal

counterparts, in the sense that fulfillment of *p* is more likely to result in *q*. For example, your chances of finding a good parking spot are probably better if you arrive a full hour early for work (when the parking lot presumably is empty) than if you arrive only 10 minutes early. If this hypothesis is correct, then we would expect higher estimates of P(*q*/*p*) for the reversed condition. However, an examination of participants' direct conditional probability estimates for conditional tips revealed that P(*q*/*p*) ratings were in fact *lower* in the reversed condition (mean = 65.22%) than in the normal condition (mean = 73.75%), $t(31) = 2.64$, $SE = 3.23$. Thus, consistent with the finding that effectiveness ratings do not correlate positively with P(*q*/*p*), the greater effectiveness of reversed tips does not seem to be due to higher ratings of P(*q*/*p*).

Another possibility is that the preference manipulation might have affected estimates of the conditional probability of *q* given ~*p*, P(*q*/~*p*). Recall that effectiveness has previously (Experiments 3-5) been found to correlate negatively with P(*q*/~*p*); thus, perhaps the reversed tips were more effective because they received lower ratings of P(*q*/~*p*). In our tip, for example, failing to arrive a full hour early (i.e., ~*p*) would still leave plenty of time to get a good parking spot, whereas this window of opportunity is much smaller in the normal "10 minutes early" version. However, there was little support for this hypothesis: direct estimates of P(*q*/~*p*) for tips did not differ between the normal (mean = 37.34%) and the reversed (mean = 40.31%) conditions, $t(31) = 0.92$, $SE = 3.22$.[15] In addition, effectiveness did not correlate with P(*q*/~*p*) in this particular experiment (in contrast to Experiments 3-5), further questioning this hypothesis.

[15] However, the preference manipulation did affect estimates of P(*q*/~*p*) for promises. Specifically, P(*q*/~*p*) ratings were lower in the reversed condition (mean = 21.00%) than in the normal condition (mean = 30.08%), $t(29) = 2.37$, $SE = 3.83$.

Finally, this counter-intuitive finding regarding effectiveness ratings for conditional tips could be a consequence of our difficulty in reversing preference patterns. As reported above, the preference manipulation did not actually *reverse* ratings for the *pq*/~*p*~*q* pair; rather, it reduced the tendency to prefer *pq* over ~*p*~*q*. Yet, even in the reversed condition, participants were found to prefer *pq* roughly three times out of four; consequently, we performed a more fine-grained analysis by identifying those responses where participants actually preferred ~*p*~*q* to *pq* for conditional tips.

Of a total of 64 responses in the reversed condition, participants preferred *pq* on 50 (78%) and ~*p*~*q* on the remaining 14 (22%). We computed mean effectiveness ratings for these two types of responses and found that effectiveness ratings were higher for responses where *pq* was preferred (19.00) than where ~*p*~*q* was preferred (7.14). Thus, when separating responses into those where either *pq* or ~*p*~*q* was preferred, a trend in the expected direction was found [the same analysis for promises yielded a similar pattern: effectiveness was higher for responses where *pq* was preferred (12.83) than where ~*p*~*q* was preferred (5.29)].

*Quality ratings*

To explore whether the preference manipulation was successful in changing ratings of quality, two paired-samples t-tests were performed. Recall that the quality of a tip was assessed by asking for its goodness, and the quality of a promise by asking for its reasonableness (refer back to Table 15 for relevant means). Consistent with our predictions, tips were rated as less good in the reversed condition than in the normal condition, $t(31) = 5.72$, $SE = 3.77$; similarly, reasonableness ratings for promises were also found to be lower in the reversed condition, $t(29) = 3.10$, $SE = 3.49$.

These findings show that the decision theoretic analysis adopted by Over et al. (2004) can usefully be extended to conditional inducements and advice. Over et al. found that the degree to which conditional obligations (i.e., if *p* happens, then you must do *q*) and conditional permissions (i.e., if *p* happens, then you may do *q*) are judged to be good rules in guiding behaviour can be predicted by people's preference ordering of two possible outcomes, namely *pq* and *p*~*q*. Similarly, we have shown that ratings of the goodness of conditional tips, as well as the reasonableness of conditional promises, can be tied to the extent that the *pq* possibility is preferred to the ~*p*~*q* possibility. Collectively, these findings suggest that an analysis based on utilities, perceptions of costs and benefits, and subjective preferences can fruitfully be applied to the study of pragmatic conditionals.

How could this "conditional expected utility hypothesis" (Over et al., 2004) be further extended for conditional inducements and advice? One possibility is suggested by Over et al.'s paper. Specifically, while we only asked for preference judgments from the listener's perspective, we could alternatively instruct participants to make these judgments from the speaker's perspective. To illustrate the potential utility of this approach, a promise would not seem to be reasonable from the speaker's perspective if the cost associated with *q* (i.e., giving the reward) exceeded any benefit resulting from the listener performing *p*. While it is unlikely that a speaker would utter such a promise in the first place, we should still expect quality ratings to be tied to preference ratings from the speaker's perspective.

Another potential extension would be to explore negatively-phrased inducements and advice. For example, consider the threat "If you show up late for work, I will fire

you". For this threat to be an effective and reasonable deterrent, the listener would have to prefer $\sim p \sim q$ (i.e., showing up on time and avoiding dismissal) to $pq$ (i.e., getting fired after showing up late); hence, we would expect opposite preference patterns for conditionals specifying positive and negative consequences. A third option would be to combine preference judgments (i.e., perceptions of utility) with subjective probability judgments (e.g., Manktelow & Over, 1995). Consistent with a traditional subjective expected utility analysis, we might expect that the dimensions of inducements and advice we have explored (i.e., effectiveness, quality, truth) should depend not just on preference ratings of alternative outcomes, but also on the likelihood that these outcomes will in fact occur.

*Conclusions*

In this final experiment, we have attempted to show how a decision theoretic analysis based on subjective utilities and preferences can be applied to an understanding of conditional inducements and advice. Specifically, we found that ratings of quality (i.e., goodness and reasonableness) and effectiveness could be tied to participants' preference ordering of two possible outcomes, namely $pq$ and $\sim p \sim q$. These findings illustrate that people's representations of such pragmatic conditionals are likely to involve much more than the fairly simple probability judgments investigated in the "conditional probability approach" (e.g., Evans & Over, 2004; Oaksford & Chater, 2001) and in Experiments 3-5.

# GENERAL DISCUSSION

The overall aim of this dissertation research was to show the potential benefits of applying an informal approach to the study of conditional statements. For this purpose, two types of conditionals were investigated: inducements (i.e., promises and threats) and advice (i.e., tips and warnings). Because of their pragmatic richness, I have argued that a traditional formal approach grounded in the deductive paradigm is, by itself, inadequate for a complete understanding of these conditionals. In particular, an approach focusing only on the inferential use of conditional inducements and advice is unlikely to provide an account of their role as speech acts uttered to achieve some communicative or social purpose.

Consequently, the experiments in this dissertation adopted an informal approach, exploring the interpretation of inducements and advice without the restrictions imposed by the deductive paradigm. While some of these experiments included standard deductive tasks (e.g., the conditional inference task), I argue that the main contribution of this work instead involves the knowledge gained by the use of a methodology different from that traditionally favoured by reasoning researchers. Before concluding this dissertation with a general theoretical discussion, these contributions are first briefly summarized.

## *Summary of experiments*

Previous psychological research with inducements and advice (e.g., Evans & Twyman-Musgrove, 1998; Fillenbaum, 1978, 1986; Newstead, Ellis, Evans, & Dennis, 1997) suggested that people interpret and reason with these conditionals differently. For instance, Newstead et al. reported that inducements elicit higher acceptance rates on the conditional inference task, relative to advice. In an attempt to explain this difference,

Experiment 1 explored a number of pragmatic variables. Despite some redundancy among these variables, inducements were differentiated from advice on such variables as the degree of speaker's control over *q*, obligation to speaker, and necessity/sufficiency relations. In addition, participants were found to be sensitive to a purposive difference between these statements, whereby inducements were perceived to be more effective in changing the behaviour described in the antecedent of the conditional (a result replicated in subsequent experiments). The results of this study thus offer direct support for the claim that people recruit a substantial amount of background knowledge when interpreting these pragmatic conditionals.

The goal of Experiment 2 was to explore whether the variables identified in Experiment 1 could be tied to the differences in reasoning performance reported earlier. Consequently, participants performed the conditional inference task, and the degree to which these variables could predict inference rates was investigated. In addition to replicating Newstead et al.'s (1997) finding that inferences (of all types) were endorsed more frequently for inducements than for advice, this study found that the variables of speaker's control, sufficiency, and necessity were the strongest predictors of reasoning performance, thus confirming the importance attributed to these variables in past studies (e.g., Evans & Twyman-Musgrove, 1998; Thompson, 1994, 1995, 2000).

In the next set of experiments (Experiments 3-5), the main purpose was to apply a probabilistic approach to the study of conditional inducements and advice. Over the last decade, a number of authors have argued that reasoning is uncertain in nature and should be interpreted within a probabilistic, rather than a logical, framework (e.g., Evans & Over, 2004; George, 1997; Liu, Lo, & Wu, 1997; Oaksford & Chater, 2001; Politzer &

Bourmaud, 2002; Stevenson & Over, 1995). Applied to the area of conditional reasoning, a probabilistic interpretation suggests that people evaluate conditional statements as the conditional probability of *q* given *p*, or P(*q*/*p*).

In Experiment 3, this "conditional probability hypothesis" was extended to inducements and advice. With the use of Over and Evans' (2003) probability estimates task, a distinction between the truth status of these conditionals and their effectiveness as speech acts was demonstrated. Specifically, probability judgments of truth were highly correlated with computed estimates of P(*q*/*p*), while effectiveness ratings were instead predicted by the conditional probability of *q* given not-*p*, or P(*q*/~*p*); furthermore, ratings of truth and effectiveness were not themselves correlated. Finally, by showing how conditional probability estimates could predict inference patterns, this experiment also provides strong support for a probabilistic interpretation of conditional reasoning.

The main goal of Experiment 4 was to test the hypothesis, suggested by the results of Experiment 3, that the perceived truth and the perceived effectiveness of inducements and advice are independent, and also affected by different variables. In addition to replicating the main findings of Experiment 3, reducing estimates of P(*q*/*p*), achieved by manipulating the perceived credibility of the statement, resulted in a reduction in truth ratings but *not* in a corresponding decrease in effectiveness ratings. In this experiment, the conditional probability approach was also found to predict performance on another deductive reasoning task, namely the truth-table evaluation task.

A final inquiry into the seemingly counter-intuitive dissociation between truth and effectiveness was conducted in Experiment 5. In this study, a stronger credibility manipulation proved successful in substantially reducing estimates of P(*q*/*p*); yet, this

manipulation was again found to selectively affect truth ratings, suggesting that this dissociation is indeed robust. This experiment also aimed to capture a different aspect of inducements and advice, by asking participants to judge the "quality" (goodness and reasonableness for tips and promises, respectively) of these conditionals. The results showed that ratings of quality were highly correlated with truth ratings, but unrelated to ratings of effectiveness.

Finally, Experiment 6 explored inducements and advice from a decision theoretic perspective (e.g., Manktelow & Over, 1995; Over, Manktelow, & Hadjichristidis, 2004). Participants were instructed to list the four truth-table possibilities in order of preference; subsequently, the extent to which the listener was perceived to prefer *pq* to ~*p*~*q* was found to correlate with judgments of quality and effectiveness. For example, conditional promises were perceived to be both less effective and less reasonable when it was suggested that the cost of *p* outweighs the benefit of *q*, in which case the listener is likely to forgo the reward *q* by failing to perform *p* (i.e., ~*p*~*q*). These results suggest that a complete understanding of inducements and advice, as well as other pragmatic conditionals, will require a consideration of subjective utilities and preferences associated with alternative actions.

To what extent are the contributions of this dissertation restricted to the particular content investigated? In other words, do the aforementioned conclusions apply only to conditional inducements and advice, or can they be generalized to a broader class of conditionals (or even to other areas of reasoning)? At least to a certain degree, some of the conclusions do seem to be specific to inducements and advice. For instance, some of the pragmatic factors explored in Experiment 1 (e.g., speaker's control and obligation)

may not be crucial to other types of (even pragmatically-rich) conditionals. Similarly, the importance of the ~*p* possibility emphasized in Experiment 3 may not generalize from inducements and advice to other scenarios, where people may instead routinely focus on the *p* possibility to the exclusion of ~*p* (Over & Evans, 2003).

While some of the conclusions of this dissertation may indeed be specific to inducements and advice, I argue that the main contributions of this work are more general. Thus, this dissertation, along with other similar work emerging recently (e.g., Beller, 2002; Feeney, Scrafton, Duckworth, & Handley, 2004; Fiddick, 2004; Oaksford & Hahn, 2004; Over, et al., 2004; Thompson & Byrne, 2002; Thompson, Evans, & Handley, 2005), illustrates the potential benefits of an informal approach to the study of everyday reasoning. When extending such an approach to different types of statements and forms of reasoning, the particular questions *are* likely to vary; indeed, the onus is on the individual researcher to choose and develop the most appropriate questions for whatever content is under investigation.

*Implications for traditional theories of reasoning*

It should be clear that this dissertation did not explicitly aim to adjudicate between the various theories that have been proposed over the years in the psychology of reasoning. In fact, this omission was a deliberate choice, since these theories are, in my view, unable to provide an account of informal reasoning. Like the deductive paradigm from which they originate, most of these theories are geared towards explaining reasoning performance on formal tasks. In such tasks, people are in effect asked to engage in logical reasoning (e.g., determine whether conclusions necessarily follow from true premises). While few theorists nowadays remain faithful to the doctrine of logicism,

the original emphasis on logical reasoning has constrained the types of inferential processes investigated; consequently, these theories were not designed to explain the types of informal reasoning processes investigated in the current series of experiments.

For example, while both the mental logic (e.g., Braine & O'Brien, 1991) and mental model (e.g., Johnson-Laird & Byrne, 1991, 2002) approaches have added pragmatic mechanisms to explain the errors people make in deductive tasks, they still share a common agenda: to account for the ability of humans to reason deductively (Evans, 2002). As discussed in the introduction of this dissertation, however, deductive reasoning is unlikely to provide an appropriate framework for most reasoning in the real world. Other theories (e.g., Cosmides, 1989; Cheng & Holyoak, 1985; Manktelow & Over, 1991; Sperber, Cara, & Girotto, 1995; Thompson, 2000), which do not explicitly propose to account for deductive competence, nevertheless attempt to explain how reasoning performance on deductive tasks is modified by different content and various pragmatic factors; as previously argued, the continued reliance on these tasks is likely to impede our progress in exploring the complexities of informal reasoning.

Of course, these theories could always be modified or extended to provide broader accounts of human reasoning able to incorporate informal reasoning phenomena. For example, the mental model theory has recently been developed to explain probabilistic reasoning (Johnson-Laird, Legrenzi, Girotto, Legrenzi, & Caverni, 1999). According to this account, the probability of an event is based on the proportion of mental models constructed in which the event occurs. With this extension, the mental model theory could be viewed as consistent with, or at least reflecting, the uncertain nature of informal reasoning.

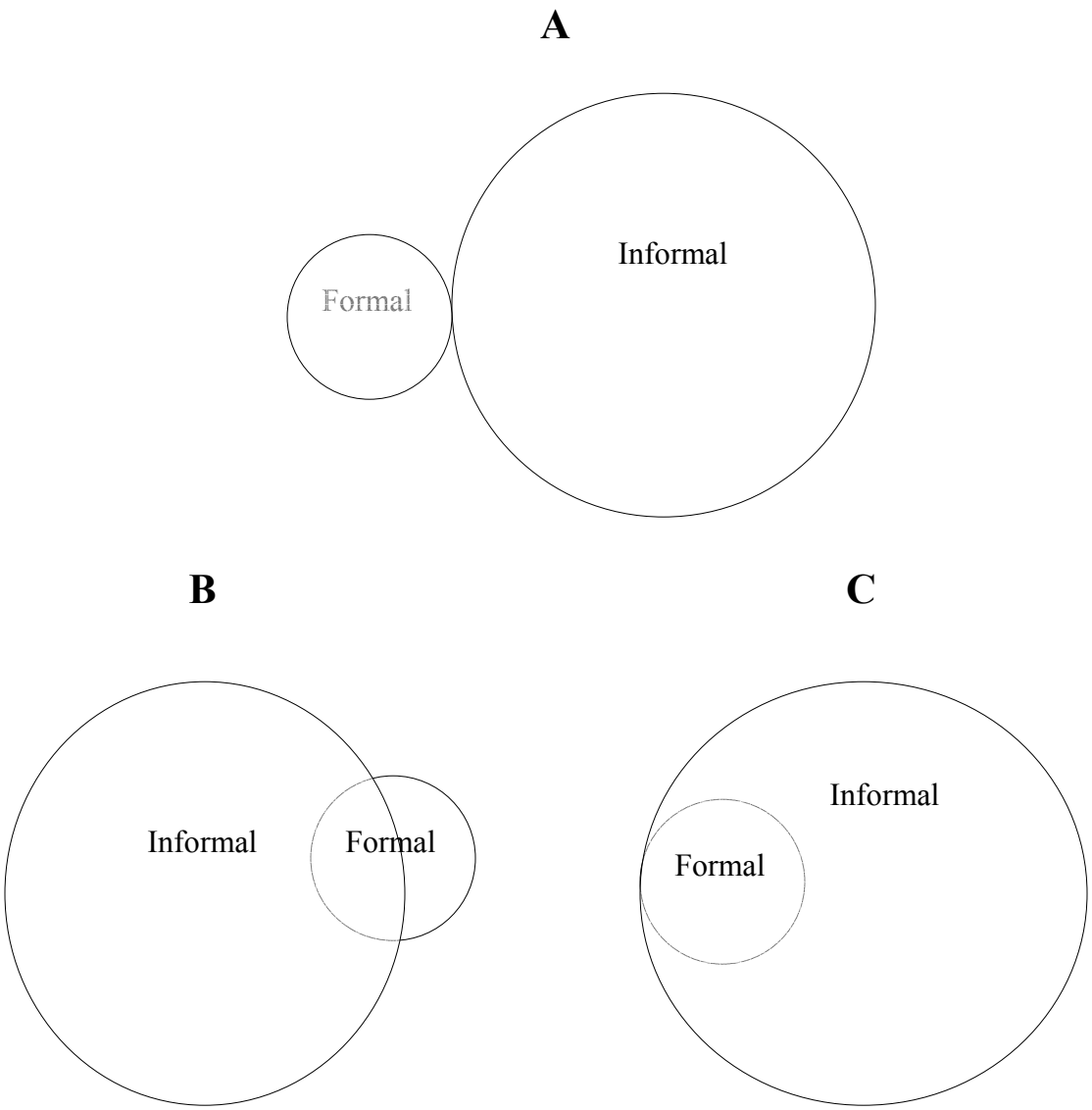
However, given that these traditional theories were proposed with a different purpose in mind, it might be overly optimistic to think that they can naturally be extended to the domain of informal reasoning. As our exploration into informal reasoning continues to make empirical progress, we should witness a corresponding theoretical development (Evans & Thompson, 2004). Thus, in order to explain skill in informal reasoning, a new set of theories is likely to emerge in the field. A better role for the traditional theories, then, might be to focus on the questions they were designed to tackle (i.e., explaining deductive competence or performance on deductive tasks); as will soon be discussed, this role will likely remain an important one.

*Formal vs. informal reasoning revisited*

In the remainder of this discussion, I will again focus on the relationship between formal and informal reasoning. From the introduction, recall that despite the primacy of informal reasoning in daily life, traditional psychological investigation has mainly explored formal reasoning. Recall also that due to the fundamental differences between these two domains, I have argued that a formal approach grounded in the deductive paradigm is unlikely to reveal the complexities and subtleties of everyday informal reasoning. In the following sections, I will discuss a few remaining theoretical issues regarding the formal vs. informal distinction; in particular, I will attempt to map this distinction onto other theoretical dichotomies proposed in the literature.

To facilitate this discussion, a graphical illustration of the relationship between formal and informal reasoning is presented in Figure 1. This figure depicts three possible representations of how formal and informal reasoning could be related to each other.

*Figure 1*. Three possible representations of the relationship between formal and informal reasoning



Common to all three is the greater size of the Euler circles representing informal reasoning, thus illustrating the primacy of this domain in daily life. However, the diagrams differ in how they depict the nature of the relationship between formal and informal reasoning.

One possibility (diagram A) is that the two domains are completely separate. This depiction is perhaps the impression readers come away with from the introduction of this dissertation; specifically, formal and informal reasoning are fundamentally different from one another, and knowledge gained in the former domain does not extend to the latter. However, this representation is unlikely to appeal to most reasoning researchers, who would surely dispute the claim that the contributions of their formal studies have no relevance at all to our understanding of everyday reasoning (e.g., Galotti, 1989; Oaksford & Chater, 2002). Another reason to be skeptical about this depiction is the evidence suggesting a convergence between the two domains. For instance, the difficulty in separating prior belief from the process of argument evaluation has been documented in both formal (e.g., Evans, Barston, & Pollard, 1983) and informal (e.g., Klaczynski & Gordon, 1996; Neuman, 2003) settings.

These arguments, therefore, suggest further possibilities. In diagram B, the two domains partially overlap, such that some of the processes investigated in studies of formal reasoning will extend to the informal domain. Nevertheless, certain aspects of formal reasoning (e.g., the principle of logical necessity) are still incompatible with the nature of informal reasoning. Finally, diagram C shows a third possibility. According to this representation, formal reasoning is actually a (very small) part of informal reasoning. In other words, while there is much more to informal reasoning than formal reasoning, a

formal approach nonetheless explores some of the reasoning processes we engage in during everyday life. Note that according to the last two diagrams, humans sometimes will make use of their formal reasoning skill when reasoning informally. The implications of this view will be discussed in the following sections.

*Formal vs. informal reasoning and dual-process theories.* The relationship between formal and informal reasoning bears an intuitive resemblance to a distinction recently made by several dual-process theories of reasoning (e.g., see Evans, 2002, 2003; Evans & Over, 1996, 1997; Sloman, 1996; Stanovich, 1999, 2004; Stanovich & West, 2000). While these theories differ somewhat in terminology (e.g., implicit vs. explicit, automatic vs. controlled, associative vs. rule-based) and descriptive details, they unite in postulating the existence of two separate reasoning systems. A summary of the main differences between these systems, generically labeled System 1 and System 2 by Stanovich and West (2000), is presented in Table 16.

In the dual-process framework, System 1 processes are described as automatic, largely unconscious, fast, and relatively undemanding of computational capacity (i.e., they can operate in parallel). Furthermore, processing is highly contextualised and pragmatic in nature, properties that contribute to what Stanovich (1999) termed the *fundamental computational bias* in human reasoning. Applied to the area of deductive reasoning, for instance, the operation of this bias is thought to be responsible for a number of pragmatic (i.e., non-deductive) strategies, such as the tendency to recruit any prior beliefs and background knowledge deemed relevant to the task at hand (Stanovich & West, 2000).

In contrast, System 2 processes are controlled, conscious, slow, and relatively

Table 16

*Properties of two reasoning systems in dual-process theories*

| System 1 | System 2 |
|---|---|
| Automatic | Controlled |
| Implicit | Explicit |
| Unconscious | Conscious |
| Parallel | Sequential |
| Rapid | Slow |
| Associative | Rule-based, symbolic |
| Undemanding of cognitive capacity | Demanding of cognitive capacity |
| Contextualised | Decontextualised |
| Evolved early | Evolved late |
| Shared with non-human animals | Unique to humans |
| Independent of intelligence | Related to intelligence |
| Acquired by biology and experience | Acquired by cultural/formal tuition |

Main sources: Evans (2002, 2003), Sloman (1996), Stanovich & West (2000)

demanding of computational capacity; thus, they are considered to be sequential and effortful in nature. Compared to System 1, System 2 is evolutionarily recent, distinctly human, and correlated with measures of general intelligence (e.g., working memory capacity or processing speed). Finally, the origins of these two systems are said to be different: while System 1 processes are acquired by biology, exposure, or personal experience, the acquisition of System 2 processes requires cultural or formal tuition (Stanovich & West, 2000).

To what extent does this dual-process framework correspond to the distinction between formal and informal reasoning? In the author's view, there seems to be a substantial degree of overlap between formal reasoning and the properties descriptive of System 2 processing. In fact, I propose that one consequence of applying System 2 is precisely to enhance skill in formal reasoning. As Stanovich and West (2000) argue, an especially important function of System 2 is to override the fundamental computational bias by suppressing the automatic contextualization of problems. This ability to decontextualise is, furthermore, a crucial skill in the domain of formal reasoning, where correct (i.e., normative) performance requires a suppression of pragmatic influences and a de-coupling of prior beliefs.

In addition, an inspection of Table 16 reveals that many of the properties of System 2 processing seem to be fairly accurate descriptions of formal reasoning. Thus, if asked to describe the processes involved when people evaluate logical arguments, most theorists would probably agree that these processes are controlled, conscious, sequential, and relatively slow. Moreover, such processes are likely to be demanding of computational capacity (i.e., effortful) and therefore correlated with measures of general

intelligence, a finding confirmed by the individual differences approach of Stanovich and West (1998, 2000). This degree of convergence between the properties of System 2 and formal reasoning may, in fact, suggests that these two terms are nearly synonymous and could be used interchangeably.

Not everyone would endorse this assertion, however. For example, advocates of the mental logic/rules approach (e.g., Braine, 1978; Braine & O'Brien, 1991; O'Brien, 2004) argue that humans, even those untutored in formal logic, possess an in-built mental logic (or "natural deduction system") consisting of a set of inferences rules; they would therefore likely dispute the claim that formal reasoning, like System 2 processing, needs to be acquired by means of cultural or formal tuition. Another likely point of contention for these theorists is the claim that formal reasoning processes are controlled, slow, and demanding of computational capacity. For instance, in Braine and O'Brien's proposal of conditional reasoning, some inference rules or reasoning schemas (e.g., for Modus Ponens), are said to be made routinely and without much effort (i.e., automatically).

Despite these potential disagreements, I think it is fair to say that the relationship between formal reasoning and System 2 is closer than the corresponding relationship between informal reasoning and System 1. One reason for this discrepancy may be the greater difficulty in providing a precise definition of informal reasoning (defined in this dissertation very broadly as the reasoning people engage in during everyday life); nevertheless, only a few of the properties of System 1 in Table 16 seem to be accurate descriptions of informal reasoning. For instance, one would certainly agree with the claim that informal reasoning processes are, like those of System 1, contextualized and

pragmatic in nature. Furthermore, informal processes may be more likely than formal ones to be acquired through personal experience rather than by tuition.

These similarities between informal reasoning and System 1 are, however, overshadowed by major differences. For example, I would argue that the informal processes engaged when people are trying to, say, justify their position on a topic or provide an effective counter-argument, are not accurately described as automatic, outside consciousness, rapid, or effortless. Furthermore, the claim that informal reasoning skill is independent of cognitive ability seems somewhat counter-intuitive; an obvious source of individual differences in informal reasoning, such as the skills just described, would seem to be differences in cognitive ability. However, the literature on this issue is inconclusive (e.g., Ceci & Liker, 1986; Gottfredson, 1997; Klaczynski & Gordon, 1996; Klaczynski & Robinson, 2000; Perkins, 1985; Stanovich & West, 1997, 1998; Toplak & Stanovich, 2003).

On the basis of these differences, it is unlikely that informal reasoning will map directly onto System 1 in the dual-process framework. Consequently, I propose that informal reasoning involves more than System 1 processing. Specifically, I suggest that informal reasoning can employ *both* System 1 and System 2, and the relative degree to which an act of reasoning engages these two systems is likely to vary. In some scenarios, for instance when the ability to evaluate the plausibility of utterances or to attribute intentions is important, informal reasoning will require the type of pragmatic and contextualized processing provided by System 1.

At other times, informal reasoning will instead rely on controlled System 2 processing. For example, consider the Argument Evaluation Test (AET, Stanovich &

West, 1997), which presents a number of debates about real social and political issues (e.g., taxes, crime, gun control). The participants' task is to evaluate the strength (i.e., the quality) of a rebuttal made by one of the contestants. This test is presented as an informal task: participants are not asked to determine whether conclusions necessarily follow from premises (as in formal tasks), but are instead assessed on "argument evaluation skills of the type studied in the informal reasoning literature" (Stanovich & West, 1998, p. 165). However, in the instructions to this task, participants are told to assume that the statements made by the contestants are factually correct and to disregard their own opinion about the issue debated; these requirements clearly rely on the ability to decontextualise, which is a fundamental attribute of System 2 processing.

In fact, the discussion of the AET highlights a fundamental problem, that of providing a clear-cut distinction between formal and informal reasoning. On the one hand, the inferential process participants are asked to engage in (i.e., judge the quality of an argument) does not rely on the formal criterion of logical necessity; in this regard, the AET could be considered an informal task. On the other hand, the instructions have a striking resemblance to the requirements of traditional deductive tasks (i.e., assume the truth of premises, disregard prior belief). The difficulty in deciding whether problems are formal or informal in nature obviously complicates the task of assessing the relationship between formal and informal reasoning.

However, can the preceding discussion of the dual-process framework help in this endeavour? To recap, I have suggested that there is a high degree of overlap between System 2 and formal reasoning. Furthermore, informal reasoning makes use of both System 1 and System 2. A critical task, therefore, involves figuring out whether the part

of System 2 that is engaged in informal reasoning is also engaged in (i.e., shared with) formal reasoning. Unfortunately, this issue can presently not be resolved, as no research explicitly addresses it. Consequently, the state of affairs I have suggested is consistent with all three possibilities illustrated in Figure 1. Thus, the part of System 2 engaged in informal reasoning could be completely separate from (diagrams A, B, C), overlap partially with (diagrams B and C), or overlap completely with (diagrams B and C) formal reasoning. Later I will make some general recommendations that could go some way in resolving the interplay between formal and informal reasoning in System 2 processing.

*Informal reasoning and induction.* In the last section, I argued that an application of the dual-process framework to informal reasoning is not straightforward. A more appropriate comparison might instead be to liken informal reasoning to inductive reasoning. A general distinction is often made between deductive and inductive reasoning (e.g., see Manktelow, 1999). As discussed throughout this dissertation, the former involves drawing necessary conclusions on the basis of premises whose truth can be assumed. These conclusions are truth-preserving (i.e., correct use of logic ensures the endorsement of only true conclusions); however, they do not allow you to learn anything new, but merely make explicit what is already implicit in the premises. In contrast, inductive arguments result in an increase in semantic information, but at the expense of truth-preservation (i.e., the premises of an inductive argument only make the conclusion more probable). When people reason inductively, they make generalizations on the basis of some evidence or observation (i.e., form general rules from particular instances). For example, from the observation that Mary frowned at you, you might infer that Mary in fact dislikes you.

With this distinction in mind, there seems to be a considerable degree of overlap between informal reasoning and inductive reasoning. An especially important similarity is that both informal and inductive inferences are likely to yield logically *invalid* conclusions. That is, while these conclusions may be plausible (even highly so), there is no guarantee that they are true. For example, it is always possible that an induction (such as concluding, from the fact that the sun has risen every day, that it will rise again tomorrow) will be falsified by new evidence. One reason why validity is usually precluded in both informal and inductive reasoning is that people tend to go beyond given information. In the case of informal reasoning, people might need to make invited or implied inferences; similarly, inductive reasoning, by definition, involves an expansion of knowledge. Another reason why informal and inductive inferences are logically invalid is that conclusions in both domains are defeasible or non-monotonic (e.g., Evans, 2002; Manktelow, 1999; Oaksford & Chater, 2001, 2002). In other words, inductive conclusions are, like informal ones, provisional and subject to change if new (especially contradictory) evidence comes to light.

A final similarity between informal and inductive reasoning is that both involve probabilistic thinking. As discussed in various sections of this dissertation, substantial evidence, even from within the formal paradigm, has accumulated in support of a probabilistic interpretation of human reasoning (e.g., Evans & Over, 2004; George, 1997; Liu et al., 1996; Oaksford & Chater, 2001; Politzer & Bourmaud, 2002; Stevenson & Over, 1995). For example, people are less willing to endorse conclusions from premises viewed as uncertain, and prefer to express degrees of confidence in their conclusions (see introduction). Similarly, inductive inferences are probabilistic in nature; indeed, the aim

of an important and voluminous area of inductive reasoning is precisely to explore how people judge the probability of uncertain events (e.g., Baron, 2000; Gilovich, Griffin, & Kahneman, 2002; Manktelow, 1999).

On the basis of these similarities, I therefore propose that most informal reasoning is inductive, rather than deductive, in nature. Thus, rather than drawing conclusions from what necessarily follows, most informal inferences are probabilistic, in that people draw conclusions with some degree of belief. Furthermore, rather than restricting themselves to the information explicitly provided, people reason from all relevant belief; in other words, informal reasoning is likely to result in an increase in semantic information. However, while induction may provide a better description of the inferential *mechanism* involved in informal reasoning, there are still reasons to suggest that informal reasoning cannot be completely captured by, or reduced to, inductive reasoning.

Specifically, most inductive reasoning tasks present participants with a well-defined set of premises and ask them to engage in a particular type of inferential process (e.g., forming a general rule from particular instances). However, most arguments in everyday life (e.g., a dialogue between two people, a newspaper editorial) do not contain clearly defined premises on which to base conclusions, and some premises may not be stated at all, but must somehow be inferred by the reasoner (e.g., Evans & Thompson, 2004; Shaw, 1996). In addition, as argued in this dissertation and by others (e.g., Beller, 2002; Evans, 2005b), an analysis based on the inferences people are willing to draw is, by itself, unlikely to fully capture the complexities of informal arguments (such as their social and communicative functions). Therefore, while induction may offer a more appropriate description of the inferential mechanism of informal reasoning than

deduction does, an inductive paradigm will likely also fall short in providing a complete account of informal reasoning.

Referring back to Figure 1, can the deductive vs. inductive dichotomy shed some light on the relationship between formal and informal reasoning? Assume first that formal reasoning is equivalent to deductive reasoning. [16] If informal reasoning is inductive, then we might argue that formal and informal reasoning are separate domains, a depiction consistent with diagram A. However, this argument rests on the assumption that informal reasoning is never deductive in nature, an assumption that most theorists would probably disagree with (but see Oaksford & Chater, 2001, 2002). If this assumption is relaxed, then diagrams B and C are also possible.

*In defense of formal reasoning*. Throughout this dissertation, I have argued that the study of formal reasoning is too restrictive to capture the complexities of real world reasoning, and, consequently, that researchers will need to make use of a more informal approach. However, this claim should not necessarily be interpreted as an argument for the abandonment of formal reasoning research. I have previously outlined some of the important empirical and theoretical contributions of the formal approach (see introduction); here, I aim to provide additional justification for the continued existence of formal reasoning research.

Recall that the study of formal reasoning has in this dissertation been defined as the processes involved in the evaluation of logical arguments. The original research agenda was geared towards assessing the ability of naïve adults (i.e., untutored in logic) to reason in accordance with the rules of logic (i.e., to draw deductively valid

[16] While formal reasoning, for the purpose of this dissertation, was restricted to deductive reasoning, it can be expanded to include the literature on judgment and decision making; in this case, the equivalence between formal and deductive reasoning obviously no longer would apply.

conclusions); furthermore, it was assumed that this ability was a cornerstone of rational thought. While belief in logic as an appropriate benchmark for human rationality has diminished over the years (Evans, 2002), most theorists, and indeed most people, would probably still agree that the ability to reason logically is an important and highly valued skill. Thus, one argument in support of formal reasoning research is that it allows us to study the degree to which people are logical and, perhaps more importantly, to identify the types of logical blunders we are prone to make.

Of course, this argument presupposes that logic, on some level, plays an important role in daily life. In other words, if the ability to follow the laws of logic confers no real benefit in the real world, then we should not be overly concerned when people violate them. So, in what ways could logic be important in everyday life? A fundamental contribution of logic is that it provides the *certainty* we need in our pursuit of knowledge and truth. As Bennett (2004) explains, logic was developed in the hope it could become a universal language by which controversies and disagreements could be resolved, in the same way that the laws of algebra can solve mathematical problems. For instance, suppose that you disagree with a statement your opponent has made. If, by using a logical chain of reasoning, you can show that this premise leads to a contradictory or impossible conclusion, then you have proved that this statement must be false.

Furthermore, the application of logical rules ensures that our reasoning is *consistent*. Imagine, for example, the chaos that would ensue if people did not comply with the law of non-contradiction, which states that a proposition cannot be both true and false at the same time (Bennett, 2004). Without this basic logical principle, it should be obvious that we could never hope to resolve differences or arrive at universal truths.

Finally, the ability to reason logically has also been said to have important real-life consequences. For instance, Johnson-Laird (1999) argues that a contributing factor in the Chernobyl accident may have been a failure to draw the valid Modus Tollens inference from the rule "If the test is to continue, then the turbine must be rotating fast enough".

On the other hand, there are also reasons to question whether an understanding of, and a willingness to conform to, logical principles is always useful or even appropriate. One problem concerns the difficulty in translating logical terms into their everyday language counterparts (e.g., Manktelow, 1999). For example, while the logical meaning of *some* is "at least one and possibly all", most people would instead interpret this quantifier to mean "at least one but not all". Thus, asserting "Some A are B" when you know that all A are in fact B would be, at best, very misleading.

Another problem is that some of the inferences sanctioned by logic are not terribly informative, and some are downright absurd. For example, from the premise "The keys are on the table", a conclusion such as "The keys are on the table, or they are in my pocket, or both" follows logically, even though it is clearly less informative than the initial premise (Johnson-Laird, 1999; see also Over & Evans, 2003, on the paradoxes following from the traditional logical interpretation of the conditional "if *p*, then *q*"). In fact, some authors have argued that since so many conclusions arrived at through the use of deductive logic are uninteresting, very little of the reasoning humans do should be characterized as deductive or logical in nature (e.g., Oaksford & Chater, 2002).

Perhaps a stronger argument in favour of the continued existence of formal reasoning research is the need for System 2 in an increasingly technological and complex world. Recall from the discussion of dual-process theories that an important function of

System 2 processing is to suppress the tendency towards the automatic contextualization of problems (Stanovich & West, 2000). While the tendency to recruit prior knowledge and reason from all relevant belief can often be adaptive (e.g., Evans, Over, & Manktelow, 1993), sometimes successful reasoning instead requires the ability, provided by System 2, to decontextualise and to engage in cognitive abstraction.

For example, Evans (2003; see also Evans & Over, 1996, 1997) points out that the ability of System 2 to engage in abstract hypothetical thinking may be crucial to our very existence. Thus, the engagement of System 1 processing, which relies on our ability to learn from past experience, will likely be insufficient if humans hope to avoid the catastrophic effects of potential disasters such as nuclear war or global warming. Another arena where the requirement for cognitive abstraction is becoming increasingly important is the job sector. Specifically, skills of abstraction are crucial to our success in dealing with cognitive complexity, which is thought to be the major distinction among jobs today (Gottfredson, 1997). In fact, the importance of System 2 may generalize to numerous aspects of daily life: many everyday tasks (e.g., filing tax returns, dealing with bureaucracies) now routinely require the ability to think abstractly (Gottfredson, 1997; Stanovich & West, 2000).

*Recommendations for the future*. It should be clear from the preceding sections that we have not been able to resolve the issue of how formal and informal reasoning are related. Keeping in mind the relative dearth of informal research in the reasoning literature, perhaps this impasse should not be that surprising. Here, I would like to conclude by making a few recommendations for future research.

First, there is a particular need to investigate the relationship between formal and informal reasoning directly. Most reasoning researchers would probably claim that formal reasoning competence can predict skill in everyday or informal reasoning; the focus on formal reasoning could thus be defended by arguing that it allows researchers to investigate informal reasoning under more controlled circumstances (Galotti, 1989). In other words, the "messy" nature of informal reasoning can be studied with the use of a well-established methodology presenting few practical difficulties to the researcher (e.g., in differentiating poor from good reasoning). However, as I have argued, the fundamental differences between formal and informal reasoning provide strong reasons to doubt that research in the former domain can generalize to the latter.

To my knowledge, there are only a handful of studies that include both formal and informal tasks (Neuman, 2003; Radvansky & Copeland, 2004; Ricco, 2003; Stanovich & West, 1998); furthermore, these studies fail to provide a clear answer to the question of whether formal reasoning competence is correlated with informal reasoning skill. Hence, a greater empirical effort is needed to address this issue; only then can we determine whether the knowledge generated by formal reasoning research in fact can be generalized to informal settings, or is restricted to explaining performance on a small number of deductive tasks. In addition, these studies should help in answering the question of whether reasoning processes are similar across the two domains.

A direct comparison of formal and informal reasoning should also assist in the quest to accurately portray the relationship between these two domains. The research described in this dissertation cannot distinguish between the different representations depicted in Figure 1; worse still, it is sometimes difficult to determine whether a reasoner

is in fact engaging in formal or informal reasoning. This state of affairs obviously complicates matters. As an example, I previously mentioned that an argument in favour of the continuation of the formal paradigm involves the importance of System 2 processing (and especially the ability to decontextualise). However, I also argued that informal reasoning engages both System 1 and System 2; if System 2 processing can be studied in the informal domain, then what use do we have for the formal paradigm? This problem suggests that we need to provide better definitions of formal and especially informal reasoning.

I would furthermore suggest that researchers make efforts to improve the external validity of reasoning research. While a greater emphasis on informal reasoning no doubt will assist in such an endeavour, there is also a specific need to extend laboratory phenomena to more naturalistic settings (for a discussion on the relation between laboratory work and the real world, see Fischhoff, 1996). A good illustration of such work is provided by Dunbar (e.g., 2001), who advocates a two-pronged approach to the study of scientific reasoning. Specifically, cognitive processes are identified by observing and analyzing scientists "live" in their laboratories; controlled experiments are then conducted to further examine these processes. Fugelsang, Stein, Green, and Dunbar (2004) recently demonstrated the utility of this *in vivo*/*in vitro* approach when exploring the oft-reported confirmation bias (e.g., Nickerson, 1998), showing that an initial reluctance to accept inconsistent data is usually overcome with repeated observations of such data.

Finally, I would like to highlight some of the additional benefits that likely will ensue from adopting an informal approach. One benefit is that an emphasis on informal

reasoning should give the field a better opportunity to have some practical application. Specifically, such research may have the potential to inform people on how to become better everyday reasoners (e.g., detect weak arguments, make strong counter-arguments), a potential that would seem more limited for research in the deductive paradigm. Another attribute of an informal approach is that researchers will likely make greater cross-reference with related fields in psychology (e.g., decision making, social cognition) as they attempt to explore an increasing number of informal reasoning phenomena.

# Conclusions

Collectively, the experiments in this dissertation have illustrated the utility of applying an informal approach to the study of everyday reasoning. By offering a detailed analysis of how people interpret and reason with pragmatic conditionals in the form of inducements and advice, this work suggests that a formal approach grounded in the traditional deductive paradigm is, by itself, inadequate to explore the complexities of everyday reasoning. Thus, the most intriguing findings from this dissertation did not involve the use of deductive tasks, but were instead gained by using a methodology tailored to the conditionals under investigation. In the future, more contributions, both empirical and theoretical in nature, will surely follow from this type of informal approach; these contributions should also be directly relevant to the issue of how we should view the relationship between formal and informal reasoning.

References

Adams, E. W. (1965). The logic of conditionals. *Inquiry, 8*, 166-197.

Adams, E. W. (1975). *The logic of conditionals: An application of probability to deductive logic*. Dordrecht, Holland: Reidel.

Ajzen, I. (1991). The theory of planned behavior. *Organizational behavior and human decision processes, 50*, 179-211.

Baron, J. (1995). Myside bias in thinking about abortion. *Thinking and Reasoning, 1*, 221-235.

Baron, J. (2000). *Thinking and deciding* (3rd edition). New York: Cambridge University Press.

Baron, J. (2004). Normative models of judgment and decision making. In D. J. Kohler & N. Harvey, *Blackwell handbook of judgment and decision making* (pp. 19-36). Malden, MA: Blackwell.

Beller, S. (2002). Conditional promises and threats: Cognition and emotion. In W. D. Gray, & C. D. Schunn (Eds.), *Proceedings of the Twenty-fourth Annual Conference of the Cognitive Science Society* (pp. 113-118). Mahwah, NJ: Lawrence Erlbaum.

Bennett, D. J. (2004). *Logic made easy: How to know when language deceives you*. New York: W. W. Norton & Co.

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

Braine, M. D. S. (1978). On the relation between the natural logic of reasoning and standard reasoning. *Psychological Review, 85*, 1-21.

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.

Brem, S. K., & Rips, L. J. (2000). Explanation and evidence in informal argument. *Cognitive Science, 24*, 573-604.

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

Byrne, R. M. J., & Egan, S. M. (2004). Counterfactual and prefactual conditionals. *Canadian Journal of Experimental Psychology, 58*, 82-89.

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

Ceci, S. J., & Liker, J. K. (1986). A day at the races: A study of IQ, expertise, and cognitive complexity. *Journal of Experimental Psychology: General, 115*, 255-266.

Chan, D., & Chua, F. (1994). Suppression of valid inferences: Syntactic views, mental models, and relative salience. *Cognition, 53*, 217-238.

Chase, V. M., Hertwig, R., & Gigerenzer, G. (1998). Visions of rationality. *Trends in Cognitive Sciences, 2*, 206-214.

Chater, N., & Oaksford, M. (1999). The probability heuristics model of syllogistic reasoning. *Cognitive Psychology, 38*, 191-258.

Chater, N., & Oaksford, M. (2000). The rational analysis of mind and behavior. *Synthese, 122*, 93-131.

Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology, 17,* 391-416.

Cheng, P. W., Holyoak, K. J., Nisbett, R. E., & Oliver, L. M. (1986). Pragmatic versus syntactic approaches to training deductive reasoning. *Cognitive Psychology, 18*, 293-328.

Cohen, L. J. (1981). Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences, 4*, 317-370.

Conison, J. (1997). The pragmatics of promise. *Canadian Journal of Law and Jurisprudence, 10*, 273-322.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition, 31*, 187-276.

Cummins, D. D. (1995). Naïve theories and causal deduction. *Memory and Cognition, 23*, 646-658.

Cummins, D. D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory and Cognition, 19*, 274-282.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2003). Inference suppression and semantic memory retrieval: Every counterexample counts. *Memory and Cognition, 31*, 581-595.

Dieussart, K., Schaeken, W., De Neys, W., & d'Ydewalle, G. (2000). Initial belief state as a predictor of belief revision. *Current Psychology of Cognition, 19*, 277-288.

Dunbar, K. (2001). What scientific thinking reveals about the nature of cognition. In K. Crowley, C. D. Schunn, & T. Okada (Eds.), *Designing for science: Implications*

*from everyday, classroom, and professional settings* (pp. 115-140). Mahwah, NJ: Lawrence Erlbaum Associates.

Edgington, D. (1995). On conditionals. *Mind, 104*, 235-329.

Edgington, D. (2003). What if? Questions about conditionals. *Mind and Language, 18*, 380-401.

Edwards, K., & Smith, E. E. (1996). A disconfirmation bias in the evaluation of arguments. *Journal of Personality and Social Psychology, 71*, 5-24.

Elio, R., & Pelletier, F. J. (1997). Belief change as propositional update. *Cognitive Science, 21*, 419-460.

Evans, J. St. B. T. (1977). Linguistic factors in reasoning. *Quarterly Journal of Experimental Psychology, 29*, 297-306.

Evans, J. St. B. T. (1989). *Bias in human reasoning: Causes and consequences*. London; Lawrence Erlbaum.

Evans, J. St. B. T. (2002). Logic and human reasoning: An assessment of the deduction paradigm. *Psychological Bulletin, 128*, 978-996.

Evans, J. St. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences, 7*, 454-459.

Evans, J. St. B. T. (2005a). Deductive reasoning. In K. J. Holyoak, & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning*. New York: Cambridge University Press.

Evans, J. St. B. T. (2005b). The social and communicate function of conditional statements. *Mind and Society, 4*, 97-113.

Evans, J. St. B. T., Barston, J. L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory and Cognition, 11*, 295-306.

Evans, J. St. B. T., Handley, S. J., & Over, D. E. (2003). Conditionals and conditional probability. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 321-335.

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

Evans, J. St. B. T., & Over, D. E. (1996). *Rationality and Reasoning*. Hove, UK; Psychology Press.

Evans, J. St. B. T., & Over, D. E. (1997). Rationality in reasoning: The problem of deductive competence. *Current Psychology of Cognition, 16*, 3-38.

Evans, J. St. B. T., & Over, D. E. (2004). *If*. Oxford, UK: Oxford University Press.

Evans, J. St. B. T., Over, D. E., & Manktelow, K. I. (1993). Reasoning, decision making, and rationality. *Cognition, 49*, 165-187.

Evans, J. St. B. T., & Thompson, V. A. (2004). Informal reasoning: Theory and method. *Canadian Journal of Experimental Psychology, 58,* 69-74.

Evans, J. St. B. T., & Twyman-Musgrove, J. (1998). Conditional reasoning with inducements and advice. *Cognition, 69*, B11-B16.

Feeney, A., Scrafton, S., Duckworth, A., & Handley, S. J. (2004). The story of some: Everyday pragmatic inference by children and adults. *Canadian Journal of Experimental Psychology, 58*, 90-101.

Fiddick, L. (2004). Domains of deontic reasoning: Resolving the discrepancy between the cognitive and moral reasoning literatures. *Quarterly Journal of Experimental Psychology, 57A*, 447-474.

Fiddick, L., Cosmides, L., & Tooby, J. (2000). No interpretation without representation: The role of domain-specific representations and inferences in the Wason selection task. *Cognition, 77*, 1-79.

Fillenbaum, S. (1974). Information amplified: Memory for counterfactual conditionals. *Journal of Experimental Psychology, 102*, 44-49.

Fillenbaum, S. (1975). If: Some uses. *Psychological Research, 37*, 245-260.

Fillenbaum, S. (1976). Inducements: On the phrasing and logic of conditional promises, threats, and warnings. *Psychological Research, 38*, 231-250.

Fillenbaum, S. (1977). A condition on plausible inducements. *Language and Speech, 20*, 136-141.

Fillenbaum, S. (1978). How to do some things with IF. In J. W. Cotton, & R. L. Klatzky (Eds.), *Semantic factors in cognition* (pp. 169-214). Hillsdale, NJ: Lawrence Erlbaum.

Fillenbaum, S. (1986). The use of conditionals in inducements and deterrents. In E. C. Traugott, A. T. Meulen, J. S. Reilly, & C. A. Ferguson (Eds.), *On conditionals* (pp. 179-195). Cambridge: Cambridge University Press.

Fischhoff, B. (1996). The real world: What good is it? *Organizational Behavior and Human Decision Processes*, 65, 232-248.

Fugelsang, J. A., Stein, C. B., Green, A. E., & Dunbar, K. N. (2004). Theory and data interactions of the scientific mind: Evidence from the molecular and the cognitive laboratory. *Canadian Journal of Experimental Psychology, 58*, 132-141.

Fugelsang, J. A., & Thompson, V. A. (2003). A dual-process model of belief and evidence interactions in causal reasoning. *Memory and Cognition, 31*, 800-815.

Galotti, K. M. (1989). Approaches to studying formal and everyday reasoning. *Psychological Bulletin, 105*, 331-351.

Geis, M. L., & Zwicky, A. M. (1971). On invited inferences. *Linguistic Inquiry, 2*, 561-566.

George, C. (1995). The endorsement of the premises: Assumption-based or belief-based reasoning. *British Journal of Psychology, 86*, 93-111.

George, C. (1997). Reasoning from uncertain premises. *Thinking and Reasoning, 3*, 161-189.

George, C. (1999). Evaluation of the plausibility of a conclusion derivable from several arguments with uncertain premises. *Thinking and Reasoning, 5*, 245-281.

Gigerenzer, G., & Hug, K. (1992). Domain-specific reasoning: Social contracts, cheating, and perspective change. *Cognition, 43*, 127-171.

Gigerenzer, G., & Selten, R. (2001). Rethinking rationality. In G. Gigerenzer, & R. Selten (Eds.), *Bounded Rationality: The Adaptive Toolbox* (pp. 1-12). Cambridge, MA: MIT Press.

Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgment*. New York: Cambridge University Press.

Gottfredson, L. S. (1997). Why *g* matters: The complexity of everyday life. *Intelligence, 24*, 79-132.

Grice, H. P. (1975). Logic and conversation. In P. Cole, & J. L. Morgan (Eds.), *Syntax and semantics, Vol. 3: Speech acts* (pp. 41-58). London: Academic Press.

Grice, P. (1989). *Studies in the way of words*. Cambridge, MA; Harvard University Press.

Henle, M. (1962). On the relation between logic and thinking. *Psychological Review, 69*, 366-378.

Hilton, D. J. (1995). The social context of reasoning: Conversational inference and rational judgment. *Psychological Bulletin, 118*, 248-271.

Hilton, D. J., & Erb, H-P. (1996). Mental models and causal explanation: Judgements of probable cause and explanatory relevance. *Thinking and Reasoning, 2*, 273-308.

Hilton, D. J., Jaspars, J. M. F., & Clarke, D. D. (1990). Pragmatic conditional reasoning: Context and content effects on the interpretation of causal assertions. *Journal of Pragmatics, 14*, 791-812.

Johnson-Laird, P. N. (1999). Deductive reasoning. *Annual Review of Psychology, 50*, 109-135.

Johnson-Laird, P. N. (2001). Mental models and deduction. *Trends in Cognitive Sciences, 5*, 434-442.

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

Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review, 109*, 646-678.

Johnson-Laird, P. N., Legrenzi, P., Girotto, V., Legrenzi, M. S., & Caverni, J-P. (1999). Naïve probability: A mental model theory of extensional reasoning. *Psychological Review, 106*, 62-88.

Johnson-Laird, P. N., & Tagart, J. (1969). How implication is understood. *American Journal of Psychology, 2*, 367-373.

Kirby, K. N. (1994). Probabilities and utilities of fictional outcomes in Wason's four card selection task. *Cognition, 51*, 1-28.

Klaczynski, P. A., & Gordon, D. H. (1996). Self-serving influences on adolescents' evaluations of belief-relevant evidence. *Journal of Experimental Child Psychology, 62*, 317-339.

Klaczynski, P. A., & Robinson, B. (2000). Personal theories, intellectual ability, and epistemological beliefs: Adult age differences in everyday reasoning biases. *Psychology and Aging, 15*, 400-416.

Koehler, D. J., & Harvey, N. (2004). *Blackwell handbook of judgment and decision making*. Malden, MA: Blackwell.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review, 96*, 674-689.

Kuhn, D. (1992). Thinking as argument. *Harvard Educational Review, 62*, 155-178.

Kuhn, D. (2001). How do people know? *Psychological Science, 12*, 1-8.

Light, P., Girotto, V., & Legrenzi, P. (1990). Children's reasoning on conditional promises and permissions. *Cognitive Development, 5*, 369-383.

Liu, I.-M., Lo, K.-C., & Wu, J.-T. (1996). A probabilistic interpretation of "If-Then". *Quarterly Journal of Experimental Psychology, 49A*, 828-844.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology, 37*, 2098-2109.

Manktelow, K. I. (1999). *Reasoning and thinking*. Hove, UK: Psychology Press.

Manktelow, K. I., & Fairley, N. (2000). Superordinate principles in reasoning with causal and deontic conditionals. *Thinking and Reasoning, 6*, 41-65.

Manktelow, K. I., & Over, D. E. (1991). Social roles and utilities in reasoning with deontic conditionals. *Cognition, 39*, 85-105.

Manktelow, K. I., & Over, D. E. (1995). Deontic reasoning. In S. E. Newstead, & J. St. B. T. Evans (Eds.), *Perspectives on thinking and reasoning: Essays in honour of Peter Wason* (pp. 91-114). Hove, UK: Lawrence Erlbaum.

Manktelow, K. I., Sutherland, E. J., & Over, D. E. (1995). Probabilistic factors in deontic reasoning. *Thinking and Reasoning, 1*, 201-220.

Markovits, H. (1984). Awareness of the "possible" as a mediator of formal thinking in conditional reasoning problems. *British Journal of Psychology, 75*, 367-376.

Markovits, H., & Lesage, C. (1990). Pragmatic reasoning schemas for conditional promises: Context and representation. In J.-P. Caverni, J.-M. Fabre, & M. Gonzalez (Eds.), *Cognitive biases* (pp. 183-192). New York: Elsevier Science.

Markovits, H., & Quinn, S. (2002). Efficiency of retrieval correlates with "logical" reasoning from causal conditional premises. *Memory and Cognition, 30*, 696-706.

Markovits, H., & Savary, F. (1992). Pragmatic schemas and the selection task: To reason or not to reason. *Quarterly Journal of Experimental Psychology, 45A*, 133-148.

McCawley, J. D. (1993). *Everything that linguists have always wanted to know about logic but were ashamed to ask (2nd Edition)*. Chicago: University of Chicago Press.

Neuman, Y. (2003). Go ahead, prove that God does not exist! On high school students' ability to deal with fallacious arguments. *Learning and Instruction, 13*, 367-380.

Neuman, Y., Glassner, A., & Weinstock, M. (2004). The effect of a reason's truth-value on the judgment of a fallacious argument. *Acta Psychologia, 116*, 173-184.

Neuman, Y., & Weizman, E. (2003). The role of text representation in students' ability to identify fallacious arguments. *Quarterly Journal of Experimental Psychology, 56A*, 849-864.

Newstead, S. E., Ellis, M. C., Evans, J. St. B. T., & Dennis, I. (1997). Conditional reasoning with realistic material. *Thinking and Reasoning, 3*, 49-76.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*, 175-220.

Oaksford, M., & Chater, N. (2001). The probabilistic approach to human reasoning. *Trends in Cognitive Sciences, 5*, 349-357.

Oaksford, M., & Chater, N. (2002). Commonsense reasoning, logic, and human rationality. In R. Elio (Ed.), *Common sense, reasoning, and rationality*. New York: Oxford University Press.

Oaksford, M., & Chater, N. (2003). Conditional probability and the cognitive science of conditional reasoning. *Mind and Language, 18*, 359-379.

Oaksford, M., Chater, N., & Grainger, B. (1999). Probabilistic effects in data selection. *Thinking and Reasoning, 5*, 193-243.

Oaksford, M., Chater, N., & Larkin, J. (2000). Probabilities and polarity biases in conditional inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 883-899.

Oaksford, M., & Hahn, U. (2004). A Bayesian approach to the argument from ignorance. *Canadian Journal of Experimental Psychology, 58*, 121-131.

Oaksford, M., & Moussakowski, M. (2004). Negations and natural sampling in data selection: Ecological versus heuristic explanations of matching bias. *Memory and Cognition, 32*, 570-581.

Oaksford, M., & Wakefield, M. (2003). Data selection and natural sampling: Probabilities do matter. *Memory and Cognition, 31*, 143-154.

Oberauer, K., & Wilhelm, O. (2003). The meaning(s) of conditionals: Conditional probabilities, mental models, and personal utilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 680-693.

O'Brien, D. P. (2004). Mental-logic theory: What it proposes, and reasons to take this proposal seriously. In J. P. Leighton, & R. J. Sternberg (Eds.), *The nature of reasoning*. New York: Cambridge University Press.

Ohm, E., & Thompson, V. A. (2004). Everyday reasoning with inducements and advice. *Thinking and Reasoning, 10*, 241-272.

Ohm, E., & Thompson, V. A. (2005). Conditional probability and pragmatic conditionals: Dissociating truth and effectiveness. *Thinking and Reasoning, 11*, xx-xx.

Over, D. E., & Evans, J. St. B. T. (2003). The probability of conditionals: The psychological evidence. *Mind and Language, 18*, 340-358.

Over, D. E., Manktelow, K. I., & Hadjichristidis, C. (2004). Conditions for the acceptance of deontic conditionals. *Canadian Journal of Experimental Psychology, 58*, 111-120.

Peetz, V. (1977). Promises and threats. *Mind, 86*, 578-581.

Perkins, D. N. (1985). Postprimary education has little impact on informal reasoning. *Journal of Educational Psychology, 77*, 562-571.

Politzer, G. (1986). Laws of language use and formal logic. *Journal of Psycholinguistic Research, 15*, 47-92.

Politzer, G., & Bourmaud, G. (2002). Deductive reasoning from uncertain conditionals. *British Journal of Psychology, 93*, 345-381.

Politzer, G., & Carles, L. (2001). Belief revision and uncertain reasoning. *Thinking and Reasoning, 7*, 217-234.

Politzer, G., & Nguyen-Xuan, A. (1992). Reasoning about conditional promises and warnings: Darwinian algorithms, mental models, relevance judgements or pragmatic schemas? *Quarterly Journal of Experimental Psychology, 44A*, 401-421.

Polk, T. A., & Newell, A. (1995). Deduction as verbal reasoning. *Psychological Review, 102*, 533-566.

Quinn, S., & Markovits, H. (1998). Conditional reasoning, causality, and the structure of semantic memory: Strength of association as a predictive factor for content effects. *Cognition, 68*, B93-B101.

Radvansky, G. A., & Copeland, D. E. (2004). Reasoning, integration, inference alteration, and text comprehension. *Canadian Journal of Experimental Psychology, 58*, 102-110.

Ricco, R. B. (2003). The macrostructure of informal arguments: A proposed model and analysis. *Quarterly Journal of Experimental Psychology, 56A*, 1021-1051.

Rips, L. J. (1998). Reasoning and conversation. *Psychological Review, 105*, 411-441.

Rips, L. J. (2002). Circular reasoning. *Cognitive Science, 26*, 767-795.

Rumain, B., Connell, J., & Braine, M. D. S. (1983). Conversational comprehension processes are responsible for reasoning fallacies in children as well as adults: *If* is not the biconditional. *Developmental Psychology, 19*, 471-481.

Shaw, V. (1996). The cognitive processes in informal reasoning. *Thinking and Reasoning, 2*, 51-80.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*, 3-22.

Sloman, S. A., & Lagnado, D. A. (2005). The problem of induction. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 95-116). New York: Cambridge University Press.

Sperber, D., Cara, F., & Girotto, V. (1995). Relevance theory explains the selection task. *Cognition, 57*, 31-95.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and Cognition*. Oxford, UK; Blackwell.

Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Mahway, NJ: Lawrence Erlbaum.

Stanovich, K. E. (2004). *The robot's rebellion: Finding meaning in the age of Darwin*. Chicago: University of Chicago Press.

Stanovich, K. E., & West, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology, 89*, 342-357.

Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General, 127*, 161-188.

Stanovich, K. E., & West, R. F. (1999). Discrepancies between normative and descriptive models of decision making and the understanding/acceptance principle. *Cognitive Psychology, 38*, 349-385.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences, 23*, 645-726.

Stevenson, R. J., & Over, D. E. (1995). Deduction from uncertain premises. *Quarterly Journal of Experimental Psychology, 48A*, 613-643.

Stevenson, R. J., & Over, D. E. (2001). Reasoning from uncertain premises: Effects of expertise and conversational context. *Thinking and Reasoning, 7*, 367-390.

Tabachnick, B. G., & Fidell, L. S. (2001a). *Using multivariate statistics* ($4^{th}$ edition). Boston: Allyn & Bacon.

Tabachnick, B. G., & Fidell, L. S. (2001b). *Computer-assisted research design and analysis*. Boston: Allyn & Bacon.

Thompson, V. A. (1994). Interpretational factors in conditional reasoning. *Memory and Cognition, 22*, 742-758.

Thompson, V. A. (1995). Conditional reasoning: The necessary and sufficient conditions. *Canadian Journal of Experimental Psychology, 49*, 1-60.

Thompson, V. A. (1996). Reasoning from false premises: The role of soundness in making logical deductions. *Canadian Journal of Experimental Psychology, 50*, 315-319.

Thompson, V. A. (1997). On interpreting inferences: Comment on Evans and Over. *Current Psychology of Cognition, 16*, 230-237.

Thompson, V. A. (2000). The task-specific nature of domain-general reasoning. *Cognition, 76*, 209-268.

Thompson, V. A., & Byrne, R. M. J. (2002). Reasoning counterfactually: Making inferences about things that didn't happen. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 1154-1170.

Thompson, V. A., Evans, J. St. B. T., & Handley, S. J. (2005). Persuading and dissuading by conditional argument. *Journal of Memory and Language*, *53*, 238-257.

Thompson, V. A., & Mann, J. (1995). Perceived necessity explains the dissociation between logic and meaning: The case of "only if". *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 1554-1567.

Toplak, M. E., & Stanovich, K. E. (2003). Associations between myside bias on an informal reasoning task and amount of post-secondary education. *Applied Cognitive Psychology, 17*, 851-860.

van Gelder, T., Bissett, M., & Cumming, G. (2004). Cultivating expertise in informal reasoning. *Canadian Journal of Experimental Psychology, 58*, 142-152.

Verbrugge, S., Dieussaert, K., Schaeken, W., & van Belle, W. (2004). Promise is debt, threat another matter: The effect of credibility on the interpretation of conditional promises and threats. *Canadian Journal of Experimental Psychology, 58*, 75-81.

Weinstock, M., Neuman, Y., & Tabak, I. (2004). Missing the point or missing the norms? Epistemological norms as predictors of students' ability to identify fallacious arguments. *Contemporary Educational Psychology, 29*, 77-94.

# Appendix A

## Ethics Approval

**UNIVERSITY ADVISORY COMMITTEE**
**ON ETHICS IN BEHAVIOURAL SCIENCE RESEARCH**

**NAME:** Valerie Thompson, Psychology **BSC:** 2001-186
Eyvind Ohm

**DATE:** 08-Nov-2001

The University Advisory Committee on Ethics in Behavioural Science Research has reviewed the Application for Ethics Approval for your study "Reasoning with Realistic Material" (2001-186).

1. Your study has been APPROVED
2. Any significant changes to your proposed study should be reported to the Chair for Committee consideration in advance of its implementation.
3. The term of this approval is for 5 years.
4. In order to maintain ethics approval, a status report must be submitted to the Chair for Committee consideration within one month of the current expiry date each year the study remains open, and upon study completition. Please refer to the website for further instructions: http://www.usask.ca/research/ethics.shtml

I wish you a successful and informative study.

Linda Kaeser - E Ohm for the Committee, Vice Chair

Valerie Thompson, Chair
University Advisory Committee
on Ethics in Behavioural Science Research

**Office of Research Services, University of Saskatchewan**
Kirk Hall Room 210, 117 Science Place, Saskatoon SK S7N 5C8 CANADA
Telephone: (306) 966-8576 or (306) 966-2084 Facsimile: (306) 966-8597 http://www.usask.ca/research/

Appendix B

*Conditional statements used in Experiment 1*

**Tips**:

1. Paula has recently found a new job. While talking to her friend, she is told "If you show up early for work, you will impress your boss".
2. Reggie has recently encountered problems at school. He is falling behind his classmates in a few courses. While talking to his mother, Reggie is told "If you listen to your tutor, your grades will improve".
3. Karen feels that she has few friends in her life. She is discussing ways to find new friends with her mother. Her mother tells Karen "If you show people more respect, you will find more friends".
4. Paula is thinking of buying a car. Her friend tells her "If you buy a Ford, you will be satisfied with your purchase".
5. Monica has decided to buy a laptop computer that will help her with her studies. Her boyfriend tells her "If you go to Staples, you will find a reasonably priced laptop".
6. Several law students have decided to ask the professor, Dr. Wilson, about the upcoming exam. When asked what to read, Dr. Wilson says "If you study the section on international law, you will do well on the exam".
7. Paul is very eager to ask a girl out on a date, but doesn't know if she wants to. So Paul asks his friend's opinion, who replies by saying "If you send her flowers, she will go out with you".

8. Judy is feeling stressed. When hearing this, her brother suggests “If you take yoga classes, your tension will go away”.
9. Steven is having problems with his math homework. When he tells his father about these problems, his father says “If you take remedial lessons, you will understand the topic better”.
10. Victoria is seeing her doctor for the pain she has experienced lately. Her doctor says “If you take these pills, your pain will go away”.

**Warnings**:

1. Simon is a young salesperson in a new job. He has been wearing clothes that are less formal than the other employees. An older colleague has recently told Simon “If you wear jeans to work, you will be fired”.
2. Robbie has just started a new job as a mailman. His colleague tells him “If you carry too much mail at one time, your back will be sore”.
3. A basketball player is suffering from a loss of form. One of his teammates tells him “If you continue missing shots, you will be dropped from the team”.
4. Veronica and her mother are shopping in a very crowded shopping mall. Her mother tells Veronica “If you wander away from me, you will get lost”.
5. At a local council meeting, politicians are discussing the proposed plan to close the railway line. One politician makes the point that “If you close the railway line, many people will move to a bigger city”.
6. Warren is driving to town with his girlfriend, and Warren is trying to find a place to park. His girlfriend points out that “If you park too close to the fire hydrant, you will get a ticket”.

7. Mark is playing with the family's new puppy. His father says to him "If you tease the dog, it will bite you".
8. Rick is planning to move to a bigger city. His friend tells him "If you find an apartment close to the freeway, the traffic will annoy you".
9. Irene has recently been very busy at work. As a result, she has stopped going to the gym. Her friend tells her "If you stop exercising, you will gain weight".
10. The drummer in a band is not showing up for practice. The other band members are discussing what to do with him. The bass player tells the others "If we fire the drummer, we will have to cancel the next show".

**Promises**:

1. Brian wants to watch a TV program that is shown after the hour he usually goes to bed. His mother tells him "If you wash the dishes first, you can watch the TV program".
2. A family is spending their vacation at the beach. The son asks for some ice cream, so his mother says "If you take your sister for a swim, you can have ice cream".
3. Trevor wants to go outside to play with his friends. When asking his mother if he can go outside, the mother replies "If you tidy your room first, you can go out to play".
4. Mark is a new employee in a law firm. He is very ambitious and wants to climb the corporate ladder. After being hired, his employers told him "If you attract new clients, you will get promoted".

5. It's Kevin's job to mow the lawn, but he doesn't feel like doing it tonight. So he tells his brother "If you mow the lawn, I will give you $10".
6. John goes shopping after work every Friday. However, this Friday he has no money left in his account. So he calls his wife and says "If you go shopping today, I will cook dinner".
7. Patrick needs to wash the family car, but he'd rather stay in to watch a baseball game. However, his son wants to borrow the car tonight. So Patrick tells his son "If you wash the car, you can borrow it tonight".
8. Bill and his friend are traveling through the countryside. They are discussing eating arrangements when Bill says "If you buy me breakfast, I will buy you lunch".
9. Sheila wants to buy a new jacket, but she is currently broke. She tells her sister "If you lend me $50, I will do your chores for a month".
10. Russell is having problems finding a date for the school dance. He comes up with a plan to ask Charlotte, a pretty girl who happens to be struggling at school. Russell tells this girl "If you go with me to the dance, I will help you study".

**Threats**:

1. Peter is working in a shoe store. Lately he has been arriving late for work. His boss tells him "If you show up late for work again, you will be fired".
2. Zoe is pulled over by a police officer after driving too fast. The officer decides to let her off with a warning, but tells her "If you speed again, you will get a fine".
3. Richard and his father disagree about the suitability of Richard's fiancée. His father tells Richard "If you marry that girl, I will disown you".

4. A baseball coach disagrees with the umpire's latest call. He has stepped out of the dugout, and is demonstrating with the umpire. The umpire tells the coach "If you swear at me again, you will be ejected from the game".
5. The star player of a basketball team is not passing to his teammates. His coach has told him "If you continue playing selfishly, you will be dropped from the team".
6. A bank is being robbed. To maintain order, the gunman tells the bank clerk "If you shout, I will shoot you".
7. A group of students are protesting outside city hall. A policeman who is watching over the protest says to the students "If you enter the building, you will be arrested".
8. Tommy has recently been stealing chocolate bars from the nearby grocery store. One day the cashier catches him stealing, and tells him "If you are caught stealing again, we will call the police".
9. Maria and Ulrich have been married for 10 years. In the last few years, Ulrich's drinking problem has become very taxing on their marriage. Maria tells him "If you continue drinking, I will file for a divorce".
10. Geraldine has been arrested for illegal drug use. The judge decides to let her off with a fine, but tells Geraldine "If you are caught with drugs again, you will be sentenced to jail".

Appendix C

*Conditional statements used in Experiment 3*

(note: there are three different consequents for each conditional)

**Tips:**

1. Rachel has recently started in a new job. While talking to her colleague, she is told that "If you show up early for work, you will get fresh coffee/ impress your boss/ be promoted".
2. Tim has lately encountered problems at school. He is falling behind his classmates in a few classes. While talking to his mother, he is told that "If you study harder, your teacher will be pleased/ your grades will improve/ you will be the best student in the class".
3. Local politicians are discussing the budget. One politician says that "If nurses' salaries are improved, the nurses will be happy/ the recruitment of nurses will increase/ the problems in our health system will be solved".

**Promises:**

1. It is Kevin's turn to wash the car, but he doesn't feel like doing it tonight. Instead, he wonders whether he can get his brother to wash the car. So Kevin tells his brother that "If you wash the car, I will give you $1/ $10/ $100".
2. Paul is nervous about an upcoming exam, as he feels he doesn't understand the material. He is hoping to get some help from his friend Julian who is a good student. So he tells Julian that "If you help me study, I will save you a seat in class/ buy you lunch/ buy your textbooks next term".

3. The manager of a consulting company is working hard to finish a report that must be ready by next Monday. She hopes that Sarah, one of her employees, will assist her this Saturday. So she tells Sarah that "If you work this Saturday, you can leave an hour early on Monday/ you can take Monday off/ you will get a raise".

**Warnings:**

1. At a local council meeting, politicians are discussing the proposed plan to close the museum in order to cut costs. One politician makes the point that "If you close the museum, people will be upset/ tourism will decrease/ people will move to another town".
2. Peter has started in a new job in a shoe store. A colleague tells him that "If you show up late for work, you will have to park your car far away/ you will have to work overtime/ you will be fired".
3. John is not attending all of his classes at university. While talking to his roommate, he is told that "If you skip classes, you will disappoint your parents/ fail your courses/ drop out of school".

**Threats:**

1. A baseball coach disagrees with the umpire's latest call. He has stepped out of the dugout and is arguing with the umpire. The umpire tells the coach that "If you swear at me, I will walk away/ eject you from the game/ give the other team the win".
2. The members of a rock band are considering hiring a sax player, but their manager is opposed to this idea. So he tells the band members that "If you hire a sax player,

you will each receive less money from your shows/ I will resign as manager/ I will cancel your next tour".

3. Tracy, 16, has been invited to a party by her friends. Before going to the party, her dad tells her that "If you come home after 11, I will take $5 off your allowance this week/ take away your allowance this week/ take away your allowance for a year".

Appendix D

*Computed conditional probabilities and mean inference rates for all conditional statements in Experiment 3*

| | Conditional probability | | | | Inference rate | | | |
|---|---|---|---|---|---|---|---|---|
| Conditional statement | q/p | ~q/~p | p/q | ~p/~q | MP | DA | AC | MT |
| *Tips* | | | | | | | | |
| If you show up early for work | | | | | | | | |
| You will get fresh coffee | .81 | .63 | .80 | .65 | .61 | .26 | .45 | .32 |
| You will impress your boss | .73 | .69 | .81 | .59 | .31 | .31 | .22 | .00 |
| You will be promoted | .53 | .64 | .65 | .52 | .16 | .19 | .26 | .16 |
| If you study harder | | | | | | | | |
| Your teacher will be pleased | .85 | .79 | .84 | .80 | .53 | .37 | .47 | .40 |
| Your grades will improve | .82 | .85 | .89 | .76 | .65 | .55 | .52 | .42 |
| You will be the best student | .48 | .92 | .90 | .55 | .16 | .75 | .69 | .34 |
| If nurses salaries are improved | | | | | | | | |
| The nurses will be happy | .77 | .83 | .86 | .73 | .41 | .63 | .31 | .19 |
| Recruitment will increase | .70 | .71 | .75 | .65 | .48 | .23 | .35 | .32 |
| Problems will be solved | .52 | .69 | .65 | .56 | .26 | .35 | .23 | .32 |
| Total | .69 | .75 | .79 | .65 | .40 | .40 | .39 | .28 |
| *Promises* | | | | | | | | |
| If you wash the car | | | | | | | | |
| I will give you $1 | .77 | .90 | .88 | .80 | .59 | .78 | .59 | .47 |
| I will give you $10 | .73 | .94 | .94 | .74 | .38 | .94 | .81 | .31 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| I will give you $100 | .34 | .86 | .73 | .54 | .23 | .74 | .58 | .42 |
| If you help me study | | | | | | | | |
| I will save you a seat | .83 | .61 | .75 | .72 | .52 | .26 | .35 | .19 |
| I will buy you lunch | .77 | .79 | .88 | .65 | .77 | .81 | .77 | .74 |
| I will buy your textbooks | .60 | .84 | .86 | .57 | .22 | .78 | .75 | .31 |
| If you work this Saturday | | | | | | | | |
| You can leave 1 hour early | .82 | .84 | .87 | .77 | .69 | .69 | .72 | .63 |
| You can take Monday off | .72 | .82 | .88 | .61 | .55 | .45 | .55 | .42 |
| You will get a raise | .80 | .85 | .91 | .71 | .77 | .71 | .61 | .68 |
| Total | .71 | .83 | .86 | .68 | .52 | .68 | .64 | .46 |
| *Warnings* | | | | | | | | |
| If you close the museum | | | | | | | | |
| People will be upset | .77 | .76 | .83 | .68 | .66 | .41 | .44 | .34 |
| Tourism will decrease | .71 | .74 | .75 | .69 | .56 | .31 | .25 | .44 |
| People will move | .43 | .70 | .63 | .52 | .33 | .10 | .20 | .27 |
| If you show up late for work | | | | | | | | |
| You will park car far away | .71 | .75 | .66 | .79 | .43 | .23 | .30 | .50 |
| You will work overtime | .73 | .83 | .76 | .80 | .59 | .41 | .31 | .47 |
| You will be fired | .69 | .89 | .82 | .81 | .50 | .56 | .44 | .69 |
| If you skip classes | | | | | | | | |
| You will disappoint parents | .76 | .82 | .82 | .76 | .66 | .41 | .50 | .44 |
| You will fail your classes | .72 | .64 | .67 | .70 | .37 | .03 | .27 | .23 |
| You will drop out of school | .63 | .74 | .76 | .60 | .34 | .28 | .38 | .25 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Total | .68 | .76 | .74 | .71 | .49 | .30 | .34 | .40 |
| *Threats* | | | | | | | | |
| If you swear at me | | | | | | | | |
| I will walk away | .65 | .64 | .72 | .57 | .44 | .28 | .31 | .34 |
| I will eject you | .87 | .77 | .80 | .85 | .94 | .31 | .34 | .81 |
| I will give the team the win | .64 | .83 | .76 | .73 | .40 | .33 | .40 | .43 |
| If you hire a sax player | | | | | | | | |
| You will receive less money | .55 | .71 | .68 | .59 | .40 | .33 | .33 | .33 |
| I will resign as manager | .65 | .76 | .76 | .64 | .26 | .58 | .35 | .42 |
| I will cancel your next tour | .58 | .85 | .81 | .65 | .48 | .55 | .58 | .58 |
| If you come home after 11 | | | | | | | | |
| I will take $5 off | .77 | .92 | .93 | .74 | .59 | .78 | .81 | .56 |
| Your allowance this week | .74 | .89 | .85 | .81 | .50 | .67 | .70 | .60 |
| Your allowance for 1 year | .57 | .94 | .90 | .71 | .31 | .66 | .66 | .47 |
| Total | .67 | .81 | .80 | .70 | .48 | .50 | .50 | .51 |
| Overall | .69 | .79 | .80 | .68 | .47 | .47 | .47 | .41 |

Appendix E

*Conditional statements used in Experiment 4, across manipulations of credibility and alternative antecedent*

**Tips**:

1. Rachel has recently started in a new job. One of Rachel's friends (low credibility)/ a senior colleague who has worked in the company for more than 20 years (high credibility) tells her that "If you show up early for work, you will impress the boss".

Alternative antecedent: "You can also impress the boss by working late"

2. The owner of a toy factory is concerned about a recent decline in productivity. The secretary (low credibility)/ an organizational relations expert (high credibility) tells the owner that "If you build an exercise facility for the employees, worker productivity will increase".

Alternative antecedent: "Worker productivity can also increase by providing performance bonuses"

3. Jane and Tim want to move out of their apartment and buy a house downtown. However, the houses they have looked at so far have been too expensive. One of their friends (low credibility)/ their real estate agent (high credibility) tells them that "If you wait until next year, you will find an affordable house".

Alternative antecedent: "They may also find an affordable house on the outskirts of town"

4. The relationship between Sheila and her parents has become strained in the last few months. A friend of the family (low credibility)/ their family counselor (high credibility)

tells the mother that "If you give your daughter more independence, your relationship will improve".

Alternative antecedent: "Your relationship may also improve by taking a family holiday"

**Promises**:

1. Brian's academic work has suffered since he started hanging out with a new group of friends. (Low credibility: Although the father expects to have little financial flexibility in the near future)… The parents tell Brian that "If your grades improve, will we buy you a DVD player".

Alternative antecedent: "You can also get a DVD player by quitting smoking"

2. A newspaper is putting together a special election issue for Monday's deadline. (Low credibility: Although the whole month is a very busy time for the newspaper, and the owner wants full productivity right now)… The manager tells the employees that "If you work overtime this Saturday, you can take Monday off".

Alternative antecedent: "You can also take Monday off by working overtime on Sunday"

3. Julia wants to watch a TV program that is shown after her regular bed-time. (Low credibility: Although Julia's mother is opposed to extending the children's bed-time)… Her father tells Julia that "If you wash the dishes tonight, you can stay up late".

Alternative antecedent: "You can also stay up late by cleaning your room"

4. A patient in the psychiatric unit of a hospital is resisting treatment. (Low credibility: Although it is the attending psychiatrist who determines whether patients are allowed to

go home)… A nurse tells the patient that "If you take your medication this week, you can go home for the weekend".

Alternative antecedent: "You can also go home for the weekend by attending group therapy this week"

Appendix F

*Conditional statements used in Experiment 5, across high and low credibility conditions*

**Tips**:

1. Rachel has recently started in a new job. One of Rachel's colleagues tells her that "If you show up early for work, you will impress your boss".

Low credibility: "If you show up early for work, you will be promoted"

2. Tim has lately encountered problems at school. He is falling behind his classmates in a few classes. Tim's mother tells him that "If you study harder, your grades will improve".

Low credibility: "If you study harder, you will be the best student in the class"

3. Local politicians are discussing the budget. One politician says that "If nurses' salaries are improved, the recruitment of nurses will improve".

Low credibility: "If nurses' salaries are improved, the problems in our health system will be solved"

4. Frank has been arrested for assault. One of Frank's friends tells him that "If you plead guilty, the judge will be lenient when sentencing you".

Low credibility: "If you plead guilty, the judge will let you off with a warning"

**Promises**:

1. It is Kevin's turn to wash the car, but he doesn't feel like doing it tonight. So Kevin tells his brother that "If you wash the car, I will give you $10".

Low credibility: "If you wash the car, I will give you $200"

2. Sheila wants to buy a new jacket, but is currently broke. So Sheila tells her sister that "If you lend me $50, I will do your chores for a month".

Low credibility: "If you lend me $50, I will do your chores for a year"

3. Brian's grades have deteriorated over the last year. Brian's parents tell him before his upcoming exam that "If you get an A on the exam, we will buy you a DVD player".

Low credibility: "If you get an A on the exam, we will buy you a car"

4. The workers in a factory are on strike due to financial grievances. The management tells the union representative that "If you go back to work, we will give you a 5% salary raise".

Low credibility: "If you go back to work, we will give you a 25% salary raise"

Appendix G

*Conditional statements used in Experiment 6, with "reversed" condition*

**Tips**:

1. The owner of a factory is concerned about a recent decline in worker productivity. One of the workers tells the owner that "If you provide performance bonuses, worker productivity will increase".

Reversed condition: "If you send the workers on a Caribbean cruise, worker productivity will increase"

2. Rachel has recently started in a new job. One of Rachel's colleagues tells her that "If you arrive 10 minutes early, you will get a good parking spot".

Reversed condition: "If you arrive 1 hour early, you will get a good parking spot"

3. Dan, an enthusiastic hockey player, has recently been feeling some back pain. One of his teammates tells him that "If you take painkillers, your pain will go away".

Reversed condition: "If you stop playing, your pain will go away"

4. Robert and his girlfriend recently had a fight. One of Robert's friends tells him that "If you apologise, she will forgive you".

Reversed condition: "If you propose to her, she will forgive you"

**Promises**:

1. It's Kevin's turn to wash the family car, but he doesn't feel like doing it tonight. So he tells his brother that "If you wash the car, I will give you $10".

Reversed condition: "If you wash the car, I will give you $1"

2. Julia wants to watch a TV program that is shown after her regular bed-time. Julia's mother tells her that "If you wash the dishes tonight, you can stay up an extra hour".

Reversed condition: “If you wash the dishes tonight, you can stay up an extra 10 minutes”

3. The star player of a basketball team is unhappy and has requested a trade. The general manager tells the player that “If you withdraw your request, you will get a salary raise”.

Reversed condition: “If you withdraw your request, you will get a contract extension”

4. A newspaper is putting together a special election issue for Monday’s deadline. The owner tells the employees that “If you work overtime this weekend, you can take next Friday off”.

Reversed condition: “If you work overtime this weekend, you can go home an hour early on Monday”