Experimental Psychology

HOGREFE

Experimental Psychology

HOGREFE

Inferences From Counterfactual Threats and Promises

Suzanne M. Egan[1] and Ruth M. J. Byrne[2]

[1]Mary Immaculate College, University of Limerick, Ireland, [2]School of Psychology and Institute of Neuroscience, Trinity College Dublin, University of Dublin, Ireland

Abstract. We examine how people understand and reason from counterfactual threats, for example, "if you had hit your sister, I would have grounded you" and counterfactual promises, for example, "if you had tidied your room, I would have given you ice-cream." The first experiment shows that people consider counterfactual threats, but not counterfactual promises, to have the illocutionary force of an inducement. They also make the immediate inference that the action mentioned in the "if" part of the counterfactual threat and promise did not occur. The second experiment shows that people make more negative inferences (modus tollens and denial of the antecedent) than affirmative inferences (modus ponens and affirmation of the consequent) from counterfactual threats and promises, unlike indicative threats and promises. We discuss the implications of the results for theories of the mental representations and cognitive processes that underlie conditional inducements.

Keywords: inducements, promises, threats, reasoning, conditionals, counterfactuals

The Irish Times on March 11, 2011 reported under the headline "Man acquitted in case where alleged threat to kill was made in past tense," that a man was acquitted at the special criminal court because his defence counsel argued successfully that he had not threatened to kill or cause serious harm to the defendant's family because the "assertion that the accused man told him he 'could have had' his father shot and his sister 'done' was a statement in the past tense and could not amount to words in law of a threat to future conduct." Yet people often assert counterfactual inducements, for example, "if you had hit your brother then I would have sent you to your room," or "if you had cleared away the dishes like I asked you to then I would have given you ice-cream." What then do people think counterfactual threats and promises mean and what inferences do they make from them? The answer is in fact unknown and our aim in this paper is to report the first experimental evidence on how people reason from counterfactual inducements.

Indicative Promises and Threats

People make inferences about inducements of various sorts – promises, threats, tips, and warnings. Inducements, for example, "if you clear away the dishes then I will give you ice-cream" are usually in the indicative mood and they have implications for an actor's goals (Bonnefon, 2009; see also De Neys, Vartanian, & Goel, 2008); they are speech acts, that is, they contain the commission of an action within the utterance (Searle, 1969; see also Athanasiadou & Dirven, 1997). The hearer must understand the illocutionary force of the utterance, that is, the speaker's intention, which for a promise is to *encourage* an action by offering an *incen-*

tive, and for a threat is to *discourage* an action by offering a *disincentive*. The encouraged or discouraged action referred to in the "if" part of a promise or threat must be one over which the hearer has some choice, and the incentive or disincentive referred to in the "then" part must be one over which the speaker has control and which has a value (Beller, Bender, & Kuhnmunch, 2005; Bonnefon & Hilton, 2004). Threats may be considered promises to do something negative, but one way they differ from promises is that promises mention the action they wish to induce (e.g., clear away the dishes) whereas threats mention the *opposite* of the action they wish to induce (e.g., hit your brother). The difference, as we will see, may be important for the sorts of mental representations that people construct most readily of promises and threats.

No consistent pattern of inferences has been established for indicative inducements, perhaps because of the influence of different contents, as Table 1 shows. From a basic conditional, "if A then B," most reasoners make the modus ponens (MP) inference "A therefore B," and some also make the affirmation of the consequent (AC) inference "B therefore A" (Evans, Newstead, & Byrne, 1993). Only about half of the participants in experiments tend to make the modus tollens (MT) inference "not-A therefore not-B," and some also make the denial of the antecedent (DA) inference "not-A therefore not-B." For conditional inducements, reasoners make more inferences when the speaker has a high degree of control over the outcome than when the speaker does not (Evans & Twyman-Musgrove, 1998). The value of the outcome has to be high enough to elicit the action, but not so high as to render the inducement implausible (Fillenbaum, 1976). Reasoners may consider the cost of the action and weigh it against the cost or benefit of the

Table 1. Percentages of inferences endorsed in several previous inference tasks using inducements

	MP	AC	MT	DA	*M*
Threats					
Newstead, Ellis, Evans, and Dennis (1997, Exp. 4)	93	71	84	84	*83*
Ohm and Thompson (2004)	63	58	52	42	*54*
Carriedo et al. (1999)	92	72	80	72	*79*
Promises					
Newstead et al. (1997, Exp. 4)	89	78	54	74	*73*
Ohm and Thompson (2004)	60	67	45	60	*58*

outcome (Evans, Neilans, Handley, & Over, 2008). They endorse fewer of the affirmative (MP and AC) inferences from promises and threats that are low in credibility compared to those high in credibility (Verbrugge, Dieussaert, & Schaeken, 2004). They appear to make the negative inferences (MT and DA) readily from threats.

Counterfactual Inducements

Conditional promises and threats are uttered by a speaker in an attempt to affect the future behavior of the hearer. A threat in the indicative mood, for example, "if you place that package in the bin I will ground you for a week," becomes a mere description of a past event when it is cast in the past tense, "if you placed that package in the bin I grounded you for a week." A promise or threat cannot be about a past event when it is in the indicative mood (Searle, 1969). But *counterfactual* conditionals about promises and threats may be somewhat different. Counterfactuals have long been considered to be special, for example, their logical meaning is unlike that of indicative conditionals – it cannot be captured by a standard truth functional account in which a conditional is true if its consequent is true or its antecedent is false, because the antecedent of every counterfactual is false (Lewis, 1973; Stalnaker, 1968).

Counterfactual thoughts are pervasive in everyday thinking (e.g., Byrne, 2005, 2007; Kahneman & Tversky, 1982). They help people to learn from their past mistakes (e.g., Markman, Klein, & Suhr, 2009; Roese & Olson, 1995), and they amplify social ascriptions such as blame and fault, and emotions such as guilt and regret (e.g., Mandel, Hilton, & Catellani, 2005). When people read a counterfactual conditional "if Linda had placed the package in the bin, Noel would have put the box on the shelf" they are primed to read a negative conjunction "Linda did not place the package in the bin and Noel did not put the box on the shelf" faster compared to when they first read an indicative conditional (Santamaria, Espino, & Byrne, 2005; see also De Vega, Urrutia, & Riffo, 2007; Ferguson & Stanford, 2008). They read quickly the affirmative conjunction "Linda placed the package in the bin and Noel put the

box on the shelf" when they are primed by either the counterfactual or the indicative conditional. They may understand the counterfactual by thinking about two possibilities (Byrne, 2005). They consider the conjecture:

Linda placed the package in the bin and Noel put the box on the shelf.

And also the presupposed facts:

Linda did not place the package in the bin and Noel did not put the box on the shelf.

People judge that someone who utters the counterfactual meant to convey "Linda did not place the package in the bin" and "Noel did not put the box on the shelf" (Thompson & Byrne, 2002) and they mistakenly believe they were told so when they are given a surprise memory task (Fillenbaum, 1974). They also make more of the negative inferences (the DA inference, "Linda did not place the package in the bin therefore Noel did not put the box on the shelf" and the MT inference, "Noel did not put the box on the shelf therefore Linda did not place the package in the bin") from a counterfactual compared to an indicative conditional (e.g., Byrne & Egan, 2004; Byrne & Tasso, 1999); they make the same frequency of the affirmative inferences (the MP inference, "Linda placed the package in the bin therefore Noel put the box on the shelf" and the AC inference, "Noel put the box on the shelf therefore Linda placed the package in the bin"). The inferences people make from counterfactuals are influenced by linguistic form, for example, "even if" and "only if" (Egan, García-Madruga, & Byrne, 2009; Moreno-Rios, García-Madruga, & Byrne, 2008), and by content, for example, causal or deontic (Quelhas & Byrne, 2003; Thompson & Byrne, 2002). The results have been interpreted to mean that people understand a counterfactual, "if A had happened then B would have happened" by thinking about the presupposed facts (Byrne, 1997, 2002),

not-A and not-B

as well as the conjectured events mentioned in the counterfactual,

A and B.

How do people interpret and reason about a counterfactual promise or threat? In two experiments we examine whether people judge that a counterfactual threat and counterfactual promise have the same illocutionary force to encourage or discourage an action as an indicative threat and indicative promise. We also examine whether they judge that the encouraged or discouraged action did or did not occur for a counterfactual threat and promise as they do for an indicative threat and promise. Finally we examine the frequency of affirmative and negative inferences that people make from counterfactual promises and threats and from indicative promises and threats.

Experiment 1

The first aim of the experiment was to test whether people understand counterfactual promises and threats as active inducements or as inactive descriptions of the past. If people understand counterfactuals by thinking about two possibilities, the presupposed facts and the counterfactual conjecture, then a counterfactual threat is consistent with the illocutionary force of an indicative threat, that is, with the speaker's intention. The *indicative* threat "if you place that package in the bin then I will ground you" must draw the hearer's attention to the *opposite* of the action it mentions (i.e., do *not* place the package in the bin) because the speaker's intention is to discourage the action and encourage its opposite. Likewise, the counterfactual threat "if you had placed that package in the bin then I would have grounded you" draws the hearer's attention to the conjecture:

> You placed the package in the bin and I grounded you

But it also draws their attention to the presupposed facts:

> You did not place the package in the bin and I did not ground you

Hence the counterfactual threat, like the indicative threat, draws the hearer's attention to the *opposite* of the action it mentions in the presupposed facts (the package was not placed in the bin). As a result the counterfactual threat conveys the intention of the speaker, to discourage the action and encourage its opposite.

In contrast, a counterfactual promise may be inconsistent with the illocutionary force of an indicative promise. The indicative promise "if you place that package in the bin then I will give you extra pocket money" must draw the hearer's attention to the action it mentions (place the package in the bin) because the speaker's intention is to encourage the action. In contrast, the counterfactual promise "if you had placed that package in the bin then I would have given you extra pocket money" draws the hearer's attention to the conjecture:

> You placed the package in the bin and I gave you extra pocket money

But it also draws their attention to the presupposed facts:

> You did not place the package in the bin and I did not give you extra pocket money

Hence the counterfactual promise, unlike the indicative promise, draws the hearer's attention to the *opposite* of the action it mentions, in the presupposed facts (the package was not placed in the bin). As a result the counterfactual promise may not convey the intention of the speaker, to encourage the action. In the experiment participants were asked to classify indicative and counterfactual promises and threats as a promise, a threat, or neither a promise nor a threat.

The second aim of the experiment was to examine the immediate inferences that participants make from indicative and counterfactual promises and threats. When participants are given an indicative conditional, for example, "if Cedric takes up this new job his life will improve in every respect" and they are asked what they think follows, most people make the immediate inference that they think Cedric will take up the new job. Immediate inferences are not valid deductions but many people make them (Bonnefon & Hilton, 2004). We adapted this task for counterfactual inducements (see Egan, 2004) and participants were asked to choose one of the five options:

Jason's mother said to him "if you had put that package in the bin then I would have grounded you." What, if anything, do you think happened?

(a) Jason definitely put the package in the bin.
(b) Jason probably put the package in the bin.
(c) Nothing follows.
(d) Jason probably did not put the package in the bin.
(e) Jason definitely did not put the package in the bin.

A conditional does not support the inference of either of its components categorically. For example, when you are given the conditional, "if there is a circle drawn on the blackboard in the next room then there is a square," you cannot make the categorical conclusion, "therefore there is a circle" or "therefore there is a square." Hence, the valid response is "Nothing follows" (see also Thompson & Byrne, 2002). People tend to believe that for indicative threats the antecedent is unlikely to be carried out but for indicative promises the antecedent is likely to be carried out, when there is a strong link between the antecedent and consequent and the inducements are credible (Evans et al., 2008). Given the presuppositions conveyed by a counterfactual, we expect that they will infer that the action did not occur for both counterfactual threats and promises. The counterfactual threat, "if you had put that package in the bin then I would have grounded you," draws attention to the presupposed facts "you did not put the package in the bin and I did not ground you." Hence we expect that participants will infer the actor did not put the package in the bin. Likewise, the counterfactual promise, "if you had put that package in the bin then I would have given you an ice-cream," draws participants' attention to the presupposed facts "you did not put the package in the bin and I did not give you an ice-cream." Once again, we expect that participants will infer the actor did not put the package in the bin.

Method

Participants

The participants were 64 undergraduate psychology students from Dublin Business School's School of Arts who participated voluntarily. They were 51 women and 13 men and their average age was 27 years, ranging from 18 to 41 years.

Materials and Design

The participants acted as their own controls and the materials were based on three sorts of conditionals – promises, threats, and neutral conditionals, and they each depended on four different types of content – a parent speaking to a child, a schoolteacher speaking to a pupil, a boss speaking to an employee, and a nightclub owner speaking to a customer. The antecedent of the three conditionals was the same within each content (as in Bonnefon & Hilton, 2004), for example, if you put that package in the bin then ... I will double your pocket money this week (promise), ... I will ground you for the week (threat), ... I will put this box on the shelf (neutral). Half of the resulting 12 problems were phrased in the indicative mood and the other half in the subjunctive mood. We used materials in which the speaker has control over the outcome and its value was set at an appropriate level to ensure the credibility of the inducements.

After each conditional participants carried out two tasks. One task was to classify the conditional as (a) a promise, (b) a threat, or (c) neither. The other task was the immediate inference task, and participants were asked to say what, if anything, they thought happened (counterfactual) or would happen (indicative). They chose one option from five response options, (a)–(e), as illustrated earlier. They completed the two tasks for a problem before they moved on to the next problem. They completed the immediate inference task first and the classification task second, although we report them in the opposite order for clarity and ease of comprehension. The 12 problems were each presented on a separate page and in a different random order for each participant.

Procedure

The participants were tested in large groups. The experimenter read aloud instructions which asked participants to complete the problems in the order they were presented in their booklet and not to change any answers. Participants were advised that they could take as long as they needed to complete the task. Once all participants had completed the booklet, they were given a debriefing sheet explaining the main aims of the study.

Results and Discussion

Participants classified indicative promises as promises (95%), $\chi^2(2, N = 64) = 110.66$, $p < .01$, and indicative threats as threats (93%), $\chi^2(2, N = 64) = 105.22$, $p < .01$, rather than as the opposite or as neither, as Table 2 shows. Most participants judged counterfactual threats to be threats (67%), $\chi^2(2, N = 64) = 33.03$, $p < .01$; they judged counterfactual promises to be promises (45%), $\chi^2(2, N = 64) = 32.46$, $p < .01$, but no more often than they judged them to be neither promises nor threats (52%). The results suggest that counterfactual threats continue to exert an illocutionary force, similar to indicative threats, and there is no significant difference between them (93% vs. 67%), $\chi^2 < 1$; counterfac-

Table 2. Percentages of judgments that conditionals were promises, threats, or neither in Experiment 1

	Participants' classification		
	Promise	Threat	Neither
Counterfactual			
Promise	45	3	52
Threat	12	67	22
Neutral	11	6	83
Indicative			
Promise	95	2	3
Threat	6	93	1
Neutral	16	8	76

tual promises appear to exert a far weaker illocutionary force compared to indicative promises (95% vs. 45%), $\chi^2(2, N = 64) = 59.04$, $p < .01$.

In the immediate inference task, we constructed a composite score that the actor would tend to do the action based on the judgments that the actor "definitely" or "probably" would do the action, and a composite score that the actor would tend not to do the action based on the judgments that the actor "definitely" or "probably" would not do the action; we did so for simplicity, given that the same pattern occurs when these responses are analyzed separately. Participants inferred that the action would occur for indicative promises and they inferred it would *not* occur for indicative threats, as Table 3 shows. They inferred the action occurred for indicative promises (95%) more often than for indicative neutral conditionals (40%), $t(63) = 12.81$, $p < .01$, $\eta^2 = .72$, or indicative threats (6%), $t(63) = 32.46$, $p < .01$, $\eta^2 = .94$. In contrast, they inferred the action would *not* occur for indicative threats (92%), more often than indicative neutral conditionals (13%), $t(63) = 25.00$, $p < .01$, $\eta^2 = .91$, or indicative promises (3%), $t(63) = 28.49$, $p < .01$, $\eta^2 = .93$. They judged that nothing followed from indicative neutral conditionals more often than for indicative promises (45% vs. 1%), $t(63) = 8.90$, $p < .01$, $\eta^2 = .56$, or indicative threats (45% vs. 3%), $t(63) = 8.63$, $p < .01$, $\eta^2 = .54$.

Participants inferred that the antecedent had not occurred for counterfactual threats, counterfactual promises, and neu-

Table 3. The percentages of choices of the antecedent, A (definitely or probably), the negation of the antecedent, not-A (definitely or probably), or that "nothing follows" in Experiment 1

	A	not-A	Nothing
Counterfactual			
Promise	13	84	2
Threat	6	90	4
Neutral	5	80	15
Indicative			
Promise	95	3	1
Threat	6	92	3
Neutral	40	13	45

tral counterfactuals, as Table 3 shows. They judged that the action had not happened more often for counterfactual threats than neutral counterfactuals (90% vs. 80%), $t(63) = 2.05$, $p < .03$, $\eta^2 = .06$. However, there were no differences between counterfactual promises and neutral counterfactuals (84% vs. 80%), $t(63) = 1.04$, $p < .15$, $\eta^2 = .02$, or between counterfactual promises and threats (84% vs. 90%), $t(63) = 1.15$, $p < .17$, $\eta^2 = .02$.

The experiment provides two novel findings about counterfactual inducements: First, counterfactual threats continue to exert an illocutionary force, similar to indicative threats, but counterfactual promises exert a far weaker illocutionary force, if any, compared to indicative promises. Second, participants make an immediate inference that the antecedent action did not occur when they understand a counterfactual threat or promise, they make this inference even more so for a counterfactual threat than for a neutral counterfactual.

Immediate inferences provide us with an insight into participants' initial comprehension of counterfactual promises and counterfactual threats. Given a counterfactual promise, such as Laura's mother said to her "If you had mowed the lawn then I would have paid you 10 euro," participants' understanding of the utterance is revealed by uncovering that they make the immediate inference "Laura did not mow the lawn." Our next experiment moves on to examine participants' more considered inferences. Given a counterfactual promise, such as Laura's mother said to her "If you had mowed the lawn then I would have paid you 10 euro" and then subsequently given additional information, for example, "Laura did not mow the lawn," what if anything follows?, it is instructive to discover whether or not participants judge that it is valid to infer that therefore "Laura's mother did not pay her 10 euro." In the next experiment we examine such conditional inferences from counterfactual promises and threats. Conditional inferences can provide us with an insight into participants' deliberative reasoning from counterfactual promises and counterfactual threats.

Experiment 2

The aim of the experiment was to examine the conditional inferences reasoners make from counterfactual promises and threats (Egan & Byrne, 2006). Immediate inferences provide us with information about participants' initial comprehension of the inducement, whereas conditional inferences provide us with information about participants' reasoned judgments about what can be inferred validly from counterfactual inducements. We presented reasoners with a task of the following sort:

Laura's mother said to her "If you had mowed the lawn then I would have paid you 10 euro"
Laura did not mow the lawn.

Therefore:

(a) Laura's mother paid her 10 euro.
(b) Laura's mother did not pay her 10 euro.
(c) Laura's mother may or may not have paid her 10 euro.

They made inferences corresponding to MP, MT, DC, and AC, for counterfactual promises and threats and indicative promises and threats.

We examined their endorsements of conclusions. Given "Laura's mother said to her 'If you had mowed the lawn then I would have paid you 10 euro,' " and the MP premise "Laura mowed the lawn" the endorsement of the MP conclusion is "Laura's mother paid her 10 euro." Given the AC premise "Laura's mother paid her 10 euro" the endorsement of the AC conclusion is "Laura mowed the lawn." For the MT premise "Laura's mother did not pay her 10 euro," the endorsement of the MT conclusion is "Laura did not mow the lawn." And for the DA premise, "Laura did not mow the lawn" the endorsement of the DA conclusion is "Laura's mother did not pay her 10 euro."

Method

Participants

Thirty-five participants were recruited from the Trinity College Dublin psychology school's participant panel (members of the general public recruited through newspaper advertisements). They were paid 8 euro for their participation. There were 23 women and 12 men. Their average age was 50 years (ranging from 27 to 74 years). They were assigned at random to the indicative group ($n = 16$) or the counterfactual group ($n = 19$).

Materials and Design

Participants were assigned at random to one of two groups, indicative or counterfactual. The inference task consisted of 24 problems, 12 promises and 12 threats. The promises and threats were presented in two blocks to control for order effects, half the participants received promises first and half the participants received threats first. Each type of inference (MP, MT, DA, AC) was presented three times for promises and threats (4 Inference Types × 2 Inducement Types × 3 Contents = 24 problems). The materials were based on things a parent might say to a child to induce them to help with housework. The content was designed to ensure that the speaker had a high level of control to carry out the consequence of the action (e.g., to give a child money). It was also chosen to ensure that the inducement would not be misinterpreted as deontic and no conditionals contained phrases that might indicate obligation or permission such as "you may," "I will let you," or "you can." To control for content effects, the contents were assigned to the problems at random twice to make two sets and participants received one or other set at random. A different child's name – familiar boys' and girls' names – was used for each of the 24 problems.

Procedure

Participants were tested individually. The 24 problems for the inference task were presented on Macintosh computers

Table 4. Percentages of inferences endorsed in Experiment 2

	Affirmative			Negative		
	MP	AC	*M*	MT	DA	*M*
Promises						
Indicative	94	96	*95*	69	69	*69*
Counterfactual	67	65	*66*	89	82	*86*
Threats						
Indicative	88	77	*83*	75	79	*77*
Counterfactual	65	68	*67*	84	88	*86*

Notes. MP = modus ponens; AC = affirmation of the consequent; MT = modus tollens; DA = denial of the antecedent.

using SuperLab 1.75. The instructions were presented on the computer and included a sample problem and three practice problems (based on conjunctions and disjunctions of shapes) to familiarize participants with the task presentation and keyboard response options. Participants were advised that they could take as long as they needed to complete the task. They pressed the space bar to view each new piece of information (the conditional, the minor premise, the conclusion set), and each remained on screen to be joined by the additional information. The participants pressed one of the keys labeled "a," "b," or "c" to select a conclusion. These keys were in the center of the keyboard and corresponded to the T, G, and B keys.

Results and Discussion

A $2 \times 2 \times 2$ mixed ANOVA was carried out on the number of conclusions endorsed by participants. The between-participant factor was mood (indicative and subjunctive) and the within-participant factors were inducement (threat and promise) and inference (affirmative and negative). There were no main effects of mood, inducement, or inference ($F < 1$ in each case). Mood did not interact with inducement, $F < 1$, but it did interact with inference, $F(1, 33) = 19.80$, $p < .01$, $\eta_p^2 = .38$. There was also an interaction between inducement and inference, $F(1, 33) = 4.18$, $p < .05$, $\eta_p^2 = .112$, and between inducement, inference, and mood, $F(1, 33) = 5.60$, $p < .02$, $\eta_p^2 = .15$. We conducted four pairwise comparisons to investigate further the three-way interaction, using Bonferroni-adjusted alpha levels of .0125 per test (.05/4).

For counterfactual promises participants made more negative than affirmative inferences (86% vs. 66%), $t(18) = 2.96$, $p < .01$, $\eta^2 = .33$, whereas for indicative promises, participants made more affirmative than negative inferences (95% vs. 69%), $t(15) = 4.28$, $p < .01$, $\eta^2 = .55$. For counterfactual threats, they made more negative than affirmative inferences, 86% versus 67%, $t(18) = 2.90$, $p < .01$, $\eta^2 = .32$, whereas for indicative threats, they made as many negative as affirmative inferences, 77% versus 83%, $t(15) = 1.00$, $p < .16$, $\eta^2 = .06$ (see Table 4).

The experiment provides two novel results about counterfactual inducements: First, for counterfactual promises participants made more negative than affirmative

inferences, indicative promises show the opposite pattern. Second, for counterfactual threats, but not indicative ones, participants made more negative than affirmative inferences.

General Discussion

The experiments provide novel findings about counterfactual inducements: First, counterfactual threats exert an illocutionary force, similar to indicative threats, whereas counterfactual promises do not compared to indicative promises, as the first experiment shows. Second, participants make an immediate inference that the antecedent action did not occur when they understand a counterfactual threat or promise, they make this inference even more so for a counterfactual threat than for a neutral counterfactual. Third, participants make fewer affirmative MP and AC inferences than negative MT and DA inferences from counterfactual threats and promises. The experiments also reveal important differences between indicative promises and threats: People infer that the encouraged action occurred for a promise, and that it did not occur for a threat. We have examined indicative and counterfactual promises and threats that contained plausible content. We anticipate that different results would be found for counterfactual promises and threats with implausible content, just as has been found for indicative promises and threats with implausible content, as outlined in the Introduction (Fillenbaum, 1976; Verbrugge et al., 2004).

Inducements and Possibilities

One explanation that is consistent with this pattern of results is that when people understand the indicative promise, for example, "if you mow the lawn I will give you 5 euro" they think about a single possibility "you mow the lawn and I pay you 5 euro" (A and B) (Johnson-Laird & Byrne, 1991; Johnson-Laird, Byrne, & Schaeken, 1992). In contrast when they understand an indicative threat, "if you hit your sister I will make you wash the dishes," they think about two possibilities, "you hit your sister and I make you wash the dishes" (A and B), and "you do not hit your sister and I do not make you wash the dishes" (not-A and not-B).

The second possibility is envisaged because the threat draws attention to the opposite of the action it mentions (do *not* hit your sister). Hence they infer the action will occur for promises (A) but not for threats (not-A), and they make as many negative inferences as affirmative ones for indicative threats, but more affirmative inferences than negatives ones for indicative promises. When people understand a counterfactual threat, for example, "if you *had* hit your sister I *would* have made you wash the dishes" they think about two possibilities: the presupposed facts, "you did not hit your sister and I did not make you wash the dishes" (not-A and not-B) and the conjecture, "you hit your sister and I make you wash the dishes" (A and B). Likewise for counterfactual promises, for example, "if you had mowed the lawn I would have given you 5 euro," they think about two possibilities: the presupposed facts, "you did not mow the lawn and I did not give you 5 euro" (not-A and not-B) as well as the conjecture, "you mow the lawn and I give you 5 euro" (A and B). Hence counterfactual threats continue to exert illocutionary force because they draw attention to the same two possibilities as indicative threats, whereas counterfactual promises do not. People infer the action did not happen for both counterfactual threats and promises. Of course people may think about two possibilities for all four sorts of inducement, but one possibility appears to be more salient or accessible than the other for some of them.

Alternative Explanations

Alternative explanations have been proposed of how people understand and reason from conditionals such as "if A then B" (for a review see Byrne & Johnson-Laird, 2009). One view is that people access a set of abstract, general-purpose formal inference rules, for example, "if A then B, A, therefore B" and construct mental proofs (Braine & O'Brien, 1998; O'Brien, 2009; Rips, 2008). The logic-like rules can be applied to any area of knowledge. Given that the form of promises and threats is the same (i.e., if A then B), the difference in the rates of inferences from promises and threats cannot be explained purely in terms of their syntax; it instead requires an explanation of the effect of inducement content, which has yet to be provided by rule theories. Given that the conditional expresses a hypothetical relation, it is also not valid to infer one of the components categorically, and so on a formal inference rule account, reasoners should not make an immediate inference.

Another view is that people access content-sensitive rules, for example, "if a benefit is to be taken then a cost must be paid" (Cheng & Holyoak, 2008; Cosmides, Tooby, Fiddick, & Bryant, 2005). On this view promises could be considered to be a type of social contract in that both the hearer and the speaker gain from the exchange and they both agree to it. A promise could activate a permission or obligation schema (Cheng & Holyoak, 2008), depending on whether the perspective of the reasoner is as hearer or speaker (Politzer & Nguyen-Xuan, 1992). If reasoners rely on a permission or obligation schema they should make the MP and MT inferences readily and they should refrain

from the DA and AC inferences. But this pattern does not correspond to the one found in Experiment 2: The rates of the DA and AC inferences were high.

Another view is that people calculate probabilities based on their beliefs when they reason from conditionals (Oaksford & Chater, 2007). People make choices that maximize the information available to them (e.g., Oaksford & Chater, 1994, 2007). But probability does not predict the pattern of inferences reasoners make from promises and threats (Ohm & Thompson, 2004). For example, when participants were asked to rate the probability of B occurring given A, on a 7-point scale, for conditionals such as "if you attract new clients, you will get promoted," they judged it very likely ($M = 5.55$). They also judged that there was a high probability of B given A for conditional threats ($M = 5.26$). But despite the high ratings of probability, it was not a predictor of the inferences reasoners drew in a subsequent task; moreover, the difference in the probability ratings of promises and threats was small (Ohm & Thompson, 2004).

One probabilistic view suggests that people may suppose the "if" part of a conditional to be true, and add the "then" clause to their beliefs and judge how probable the result is (Evans, 2007; Evans & Over, 2004). On this view, people think about only one possibility at a time (the singularity principle). A counterfactual such as, "if Paul had won the lottery then he would have bought a Ferrari," is represented in a single mental model (Evans, 2007, p. 74) which combines numerical values to represent both a strength of belief in the counterfactual consequent given the counterfactual antecedent, and strengths of belief in each of the implied facts – that Paul did not win the lottery, and that he did not buy a Ferrari. According to this account, "Johnson-Laird and Byrne (2002) propose two models rather than one for the initial representation of counterfactual conditionals. The suppositional theory assumes that implicature may be readily added to the representation of all conditionals, and hence there is nothing intrinsically different in adding the pq case here" (Evans, Over, & Handley, 2005, p. 1049). However, given the proposal that people think about the antecedent occurring with the consequent (A and B) and compare it to the antecedent occurring without the consequent (A and not-B); and they do not think about the negated antecedent (not-A), it is not apparent that the theory could account for the evidence that reasoners think readily about the negated antecedent for counterfactual and indicative threats.

Conclusions

When a speaker utters a counterfactual threat, for example, "if you had been disobedient then I would have grounded you" they may wish to draw the hearer's attention to what *nearly* happened, for example, being grounded. The speaker, however, may be uncertain about whether or not the hearer will continue to carry out the discouraged action in other situations (e.g., to be disobedient) and may utter the counterfactual threat to focus attention on the just-missed negative consequence of the action. Our results suggest that a

counterfactual threat, "I could have had your father killed …" is interpreted by most ordinary people as threatening for the future, contrary to recent legal argument. In contrast, counterfactual promises appear to exert only a weak illocutionary force; they may serve as a general guide to future behavior (e.g., be good and you will be rewarded) rather than as a specific instruction (e.g., if you mow the grass tomorrow then I will pay you 10 euro). It may be that a speaker's intention when uttering a counterfactual promise is to draw the hearer's attention to the missed opportunity, perhaps amplifying the experience of regret or guilt about the missed action. Many of our decisions to act in everyday life are based on whether or not the outcome will be beneficial or detrimental and these findings may have important implications for how people reason on a daily basis.

References

Athanasiadou, A., & Dirven, R. (Eds). (1997). *On conditionals again*. Amsterdam, The Netherlands: John Benjamins.

Beller, S., Bender, A., & Kuhnmunch, G. (2005). Understand conditional promises and threats. *Thinking and Reasoning, 11*, 209–238.

Bonnefon, J. F. (2009). A theory of utility conditionals: Paralogical reasoning from decision-theoretic leakage. *Psychological Review, 116*, 888–907.

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., & O'Brien, D. (1998). *Mental logic*. Mahwah, NJ: Erlbaum.

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

Byrne, R. M. J. (2002). Mental models and counterfactual thinking about what might have been. *Trends in Cognitive Sciences, 6*, 426–431.

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

Byrne, R. M. J. (2007). Precis of the rational imagination: How people create alternatives to reality. *Behavioral and Brain Sciences, 30*, 439–453.

Byrne, R. M. J., & Egan, S. M. (2004). Counterfactual and prefactual conditionals. *Canadian Journal of Experimental Psychology, 58*, 113–120.

Byrne, R. M. J., & Johnson-Laird, P. N. (2009). 'If' and the problems of conditional reasoning. *Trends in Cognitive Sciences, 13*, 282–287.

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

Carriedo, N., García-Madruga, J. A., Gutiérrez, F., & Moreno, S. (1999). How does content affect 'unless' conditional reasoning. In S. Bagnara (Ed.), *Proceedings of the European Conference on Cognitive Science* (pp. 271–277). Sienna, Italy.

Cheng, P. C., & Holyoak, K. J. (2008). Pragmatic reasoning schemas. In J. E. Adler & L. J. Rips (Ed.) (Eds,), *Reasoning: Studies of human inference and its foundations* (pp. 827–842). Cambridge, MA: Cambridge University Press.

Cosmides, L., Tooby, J., Fiddick, L., & Bryant, G. (2005). Detecting cheaters. *Trends in Cognitive Sciences, 9*, 505–506.

De Neys, W., Vartanian, O., & Goel, V. (2008). Smarter than we think: When our brains detect that we are biased. *Psychological Science, 19*, 483–489.

De Vega, M., Urrutia, M., & Riffo, B. (2007). Cancelling updating in the comprehension of counterfactuals embedded in narrative. *Memory & Cognition, 35*, 1410–1421.

Egan, S. (2004). *Reasoning about counterfactual conditionals: The role of content, tense and linguistic form*. (Unpublished doctoral thesis). Trinity College Dublin, University of Dublin, Ireland.

Egan, S., & Byrne, R. M. J. (2006). Counterfactual promises and threats. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 1257–1262). Hillsdale, NJ: Erlbaum.

Egan, S., García-Madruga, J. A., & Byrne, R. M. J. (2009). 'Only if' counterfactuals. *Acta Psychologica, 132*, 240–249.

Evans, J. S. B. T. (2007). *Hypothetical thinking: Dual processes in reasoning and judgment*. Hove, UK: Psychology Press.

Evans, J. St. B. T., Neilans, H., Handley, S. J., & Over, D. E. (2008). When can we say 'if'? *Cognition, 108*, 100–116.

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

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

Evans, J. St. B. T., Over, D. E., & Handley, S. J. (2005). Suppositions, extensionality, and conditionals: A critique of the mental model theory of Johnson-Laird and Byrne. *Psychological Review, 112*, 1040–1052.

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

Ferguson, H. J., & Sanford, A. J. (2008). Anomalies in real and counterfactual worlds: An eye-movement investigation. *Journal of Memory and Language, 58*, 609–626.

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

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

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

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

Johnson-Laird, P. N., Byrne, R. M. J., & Schaeken, W. (1992). Propositional reasoning by model. *Psychological Review, 99*, 418–439.

Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgement under uncertainty: Heuristics and biases* (pp. 201–208). New York, NY: Cambridge University Press.

Lewis, D. (1973). *Counterfactuals*. Oxford, UK: Basil Blackwell.

Mandel, D. R., Hilton, D. J., & Catellani, P. (2005). *The psychology of counterfactual thinking*. New York, NY: Routledge.

Markman, K. D., Klein, W. M. P., & Suhr, J. A. (2009). *The handbook of imagination and mental simulation*. New York, NY: Psychology Press.

Moreno-Ríos, S., García-Madruga, J. A., & Byrne, R. M. J. (2008). Inferences from semifactual 'even if' conditionals. *Acta Psychologica, 128*, 197–209.

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.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review, 101*, 608–631.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality the probabilistic approach to human reasoning*. Oxford, UK: Oxford University Press.

O'Brien, D. P. (2009). Human reasoning includes a mental logic. *Behavioral and Brain Sciences, 32*, 96–97.

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

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

Quelhas, A. C., & Byrne, R. M. J. (2003). Reasoning with deontic and counterfactual conditionals. *Thinking and Reasoning, 9*, 43–46.

Rips, L. J. (2008). Causal thinking. In J. E. Adler & L. J. Rips (Eds.), *Reasoning: Studies of human inference and its foundation* (pp. 597–631). Cambridge, UK: Cambridge University Press.

Roese, N. J., & Olson, J. M. (Eds). (1995). *What might have been: The social psychology of counterfactual thinking*. Mahwah, NJ: Erlbaum.

Santamaria, C., Espino, O., & Byrne, R. M. J. (2005). Counterfactual and semifactual conditionals prime alternative possibilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 1149–1154.

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. London, UK: Cambridge University Press.

Stalnaker, R. C. (1968). A theory of conditionals. In N. Rescher (Ed.), *Studies in logical theory* (pp. 98–112). Oxford, UK: Basil Blackwell.

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

Verbrugge, S., Dieussaert, K., & Schaeken, 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*, 106–112.

Suzanne M. Egan

Department of Psychology
Mary Immaculate College
University of Limerick
Ireland
E-mail suzanne.egan@mic.ul.ie