This is a section of doi:10.7551/mitpress/11252.001.0001

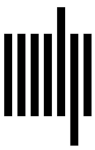
The Handbook of Rationality

Edited by: Markus Knauff, Wolfgang Spohn

Citation:

The Handbook of Rationality
Edited by: Markus Knauff, Wolfgang Spohn
DOI: 10.7551/mitpress/11252.001.0001
ISBN (electronic): 9780262366175
Publisher: The MIT Press
Published: 2021

Funding for the open access edition was provided by the MIT Libraries Open Monograph Fund.



The MIT Press

6.4 Utility Conditionals

Jean-François Bonnefon

Summary

Utility conditionals are statements like "If you update this app, your phone will crash" or "If I let you do this, my boss will fire me," where antecedents, consequents, or both are actions or outcomes that are desirable or undesirable for some agents. Because they describe desirable and undesirable states of the world, utility conditionals support peculiar inferences that draw on both theoretical rationality and practical rationality. They offer a convenient device for studying how people reason about the beliefs, preferences, and decisions of others. As a consequence, they are a natural point of contact between disparate fields such as theory of mind, Bayesian approaches to reasoning, and behavioral economics.

1. Three Ways We Reason with Utility Conditionals

Utility conditionals (Bonnefon, 2009) are statements of the form "if *p*, *q*" where *p* or *q* (or both) are actions that increase or decrease the utility (in the economic sense) of some agents. For example:

(1) a. If you wash the car, I'll let you borrow it tonight.
b. If you update this app, your phone will crash.
c. If you testify against me, you will have an accident.
d. If I let you do this, my boss will fire me.
e. If Peyton moves to Paris, Alex will be unhappy.
f. If I study instead of partying, I will get good grades.

Your washing the car is useful to me, you would not like your phone to crash, having accidents is bad, I don't want to be fired, Alex would be unhappy to be unhappy, and having good grades is good. Thus, all these examples are utility conditionals.

Some utility conditionals are speech acts, statements that perform a social function, for which we have a label: conditional (1a) is a *promise*, and conditional (1b) is a *warning*. Other utility conditionals are speech acts for which we do not have as clear a label: conditional (1c) would be something like a veiled threat, and conditional (1d) is something like an apologetic interdiction. Other utility conditionals yet, such as (1e) and (1f), do not seem to be speech acts at all.

One line of research about utility conditionals is concerned with how people sort them out into categories such as promises, threats, tips, and warnings (these categories are also of interest for deontic reasoning; see chapter 11.2 by Elqayam, this handbook). For example, in a promise such as conditional (1a), the consequent of the conditional is an event under the control of the speaker, which has positive valence for the listener. Change the valence to something negative, and you get a threat; change the event such that it is no longer under the control of the speaker, and you get a tip; change both, as in conditional (1b), and you get a warning (Evans, 2005; López-Rousseau & Ketelaar, 2004, 2006).

The frontiers between these four categories can be blurry, though. For example, conditional (1c) has the

Figure 6.4.1
Distance between promises, threats, tips, and warnings along the axes of valence and control. Some "warnings" are really veiled threats, and some threats are issued as "promises," perhaps for dramatic effect.

properties of a warning but sounds like a threat. And indeed, eye-tracking studies have shown that readers are not too perturbed when such a statement is referred to as a threat (Wray, Wood, Haigh, & Stewart, 2016)—suggesting that the speaker's explicit control over the realization of *q* is not a strong constraint for distinguishing between threats and warnings. Altogether, other eye-tracking studies (Haigh, Ferguson, & Stewart, 2014; Haigh, Stewart, Wood, & Connell, 2011; Stewart, Haigh, & Ferguson, 2013; Wood, Haigh, & Stewart, 2016) paint a picture not unlike figure 6.4.1: the frontiers around threats are blurry (both warnings and promises can be disguised or "indirect" threats), but tips are clearly identifiable as conditionals "if *p*, *q*" where *q* is something positive for the listener over which the speaker has no control.

Regardless of the way we sort utility conditionals into linguistic or pragmatic categories, there are three ways to reason about them, which involve three different forms or mixtures of rationality.

The first way, which I will not address in this chapter because it only concerns a subset of utility conditionals, is to consider their use as persuasion tools. Indeed, some utility conditionals can be used as *arguments* for or against a given course of action or policy (Schellens & De Jong, 2004). More precisely, these utility conditionals are equivalent to *arguments from consequences*, in which the good or bad consequences of some action are pointed out in order to argue that this action should or should not be taken (Bonnefon, 2016). The persuasive use of utility conditionals appeals to practical rationality. Indeed, a core question within this field of research is to assess whether people are convinced by "good" arguments—arguments that would convince a rational agent who wants actions to maximize expected utility (Corner & Hahn, 2007; Corner, Hahn, & Oaksford, 2011; Feteris, 2002; Hahn & Oaksford, 2006, 2007; Hoeken, Šorm, & Schellens, 2014; Hoeken, Timmers, & Schellens, 2012; Thompson, Evans, & Handley, 2005).

The second way one can reason about utility conditionals is to use them to derive standard syllogistic inferences (modus ponens, modus tollens, etc.), for example,

(2) a. If Peyton moves to Paris, Alex will be unhappy.
    b. Alex is happy.
    c. Therefore, Peyton has not moved to Paris.

This form of reasoning about utility conditionals primarily draws on theoretical rationality, that is, it mostly aims to maintain belief consistency. We will see, though, that practical rationality can sneak into this kind of reasoning, because what we believe of the practical rationality of others can affect the confidence we have in utility conditionals and thus the confidence we have in the conclusions of syllogisms that use them as main premises. For example, consider

(3) a. Lucy, if you don't eat your salad, you will be grounded for 5 years.
    b. Lucy does not eat her salad.
    c. Therefore, Lucy is grounded for 5 years.

As I will discuss in section 2, the reason why this conclusion sounds doubtful has to do with the lack of credibility of conditional (3a), and this lack of credibility derives from our assumptions about whether it would be practically rational to enforce the threat expressed in conditional (3a).

The third way one can reason about utility conditionals is to use them to derive nonstandard inferences about what agents want, believe, and do. For example:

(4) a. If Alex leaves early, she will be fired.
    b. Therefore, Alex won't leave early.

I call this inference "nonstandard" because it is not licensed by a logical reading of (4a). Indeed, there are no logical grounds to infer not-*p* from "if *p*, *q*." Instead, the inference that Alex won't leave early is based on the assumption that Alex is a practically rational agent who would not want to be fired and thus will not take an action that would get her fired. As I will discuss in section 3, such nonstandard inferences can concern actions, as in (4b), but also preferences and beliefs.

2. Standard Inferences

For the sake of simplicity, I will focus in this section on the most basic conditional inference, modus ponens (for other phenomena, see, e.g., Couto, Quelhas, & Byrne, 2017; Demeure, Bonnefon, & Raufaste, 2009; Hilton, Kemmelmeier, & Bonnefon, 2005). Given a utility conditional "if *p*, *q*" and the observation that *p* is true, how likely are we to conclude that *q* is true?

First, we need to observe that some utility conditionals are more credible than others. For example, I have already discussed the distinction between promises and tips. When speakers make promises, they usually are in a position to fulfill these promises:

(5) If you tidy up your room, I'll make your favorite dessert.

But when speakers give tips, they cannot actually guarantee that the tip will actually result in the expected consequence:

(6) If you work hard, your teacher will give you a good grade.

Accordingly, we can expect reasoners to be relatively confident that promises will be carried out but less confident in the consequences of tips. As a consequence, reasoners are more likely to carry out modus ponens inferences from promises than from tips (Couto et al., 2017; Evans, Neilens, Handley, & Over, 2008), and *mutatis mutandis*, they are more likely to carry out modus ponens inferences from threats than from warnings.

Even promises and threats, though, can be more or less credible. Compare, for example, the following two conditionals:

(7) a. If you don't eat your salad, you will be grounded tonight.
b. If you don't eat your salad, you will be grounded for 5 years.

Threat (7b) is so disproportionate to the offense that it loses its credibility (Amgoud, Bonnefon, & Prade, 2005; López-Rousseau, Diesendruck, & Benozio, 2011; Verbrugge, Dieussaert, Schaeken, & Van Belle, 2004). Parents who would unleash such punishment onto their children for mere peccadilloes would soon find themselves out of options—not to mention the effort that would be required from the parents to actually maintain the punishment for 5 years. Assuming that the parents have some modicum of practical rationality, it seems unlikely that they would actually carry out the threat.

Just like the vast majority of conditionals, utility conditionals have exceptions, that is, circumstances that weaken the inferential connection from their antecedent to their consequent. For example, Gazzo Castañeda and Knauff (2016b) have investigated "if crime, then punishment" utility conditionals, such as

(8) a. If a person downloads music from the Internet without allowance, then this person will be punished for breaching the copyright law.
b. If a person kills another human, then this person will be punished for manslaughter.

Because reasoners find it easier to imagine exceptions to (8a) than to (8b), they are less likely to derive modus ponens inferences from (8a) than from (8b). In other words, utility conditionals behave just like regular conditionals in this respect. However, the magnitude of the crime (i.e., the disutility inflicted by whomever took the action in the antecedent clause) affects modus ponens performance when the modal "will" is replaced with the modal "should" (Gazzo Castañeda & Knauff, 2016a):

(9) a. If a person gains admission to an event without paying, then this person should be punished for obtaining benefits by devious means.
b. If a person maltreats a minor in their charge, then this person should be punished for maltreatment of wards.

Reasoners take into account exceptions and exculpatory circumstances when they apply modus ponens to (9a), but they ignore these exceptions and exculpatory circumstances when dealing with (9b), because of the moral outrage they feel about maltreatment. In other words, the utility component of the conditional moderates the effect of exceptions on the application of modus ponens (see also Gazzo Castañeda, Richter, & Knauff, 2016).

3. Nonstandard Inferences

The most important feature of utility conditionals is that they support nonstandard inferences that are based on assumptions of practical rationality. Consider this example:

(10) If Alex writes an application, she'll get the award.

Conditional (10) suggests that Alex will write an application, based on three assumptions: that she wants the award enough to take the time to write an application, that she knows she'll get the award if she does, and that she's a rational agent who takes actions that increase her net utility. People are very well attuned to this calculus (Bonnefon & Hilton, 2004; Bonnefon & Sloman, 2013; Elqayam, Thompson, Wilkinson, Evans, & Over, 2015; Evans et al., 2008; Jara-Ettinger, Hyowon, Schulz, & Tenenbaum, 2016). Given a utility conditional "if p, q" where p is an action of an agent and q a positive (or negative, respectively) consequence of that action, reasoners will infer that the agent will or will not (and even *should* or should not) do q. They are more likely to make that inference when the magnitude of q increases, they are less likely to make that inference when the probability of q given p decreases, and they are less likely to make that inference when they can think of a way to obtain q that is less costly than p. Furthermore, eye-tracking evidence suggests that they make these inferences very quickly, in real time, as they read the conditional (Haigh & Bonnefon, 2015a, 2015b).

Not only can people predict actions based on utility conditionals, but they can also make inferences about the beliefs and preferences of other agents:

(11) a. If Brenda uses her loyalty card, she will save $20.
b. Brenda does not use her loyalty card.

Conditional (11a) would usually encourage the inference that Brenda uses her loyalty card. Given that this inference is falsified by (11b), reasoners look for a reason to explain the behavior of Brenda. Their expectation that Brenda would use the loyalty card was based on at least four assumptions: (i) Brenda cares about saving $20, (ii) she knows she can save the $20 by using her loyalty card, (iii) she's a rational agent, and (iv) she is able to use her loyalty card. When they have no reason to believe that (iii) and (iv) may be the case, they decide to revise their assumptions about Brenda's beliefs or preferences: that is, they conclude *either* that Brenda did not know about the opportunity to save money *or* that she could not be bothered to save $20 (Bonnefon, Girotto, & Legrenzi, 2012).

Utility conditionals can express more complex situations than just one action and its consequences for the agent that takes it. Consider, for example,

(12) If you hurt me, I will help her.

This conditional encapsulates a complex, three-agent situation: if *a* hurts *b*, *b* will help *c*. Is *a* going to hurt *b*? Reasoners try to make sense of such a complex situation by fitting it into an easier template, such as a social contract in which people trade favors or exchange threats (Bonnefon, 2012; Bonnefon, Haigh, & Stewart, 2013; see also chapter 10.6 by Cosmides & Tooby, this handbook). More precisely, they hesitate between a good-for-good interpretation (for some reason, *b* wants to be hurt by *a*, and *b* wants *a* to help *c*; therefore, *a* will hurt *b*) and a bad-for-bad interpretation (*a* does not want *b* to help *c*, so *b* is threatening to help *c* if *a* hurts *b*; therefore, *a* will not hurt *b*).

Reasoners make similar template-fitting inferences about the preferences of agents when they do not have intuitions about these preferences. Consider the following example, where "tymping" and "yorbing" are nonverbs whose interpretation is open:

(13) If Peyton tymps Jesse, then Jesse will yorb Peyton.

When told that Jesse enjoys being tymped, reasoners apply a social-contract template and infer that Peyton likes to be yorbed (and *mutatis mutandis* if told that Jesse does not enjoy being yorbed). Reasoners can even revise their interpretation of nonambiguous verbs such as "hurt":

(14) If Peyton hurts Jesse, then Jesse will zim Peyton.

Here reasoners spontaneously assume that Peyton would not like to be zimmed. But if told that Peyton would actually like to be zimmed, they revise their assumptions about the preferences of Jesse, coming to the conclusion that Jesse likes to be hurt (for other such template effects, see Bonnefon et al., 2013).

4. Future Directions

Utility conditionals bridge a gap between research on reasoning and research on decision making, between theoretical rationality and practical rationality. They offer a convenient device for studying how people reason about the beliefs, preferences, and decisions of others. As a consequence, they are a natural point of contact between disparate fields such as theory of mind, Bayesian approaches to reasoning, moral judgment (see chapter 12.3 by Wiegmann & Sauer, this handbook), and behavioral economics (De Vito & Bonnefon, 2014; Jara-Ettinger et al., 2016). The greater challenge ahead will be to tackle what economists call "other-regarding preferences," that is, preferences over the outcomes of other individuals (Cooper & Kagel, 2015). There is currently no well-defined theory about how people weight the utilities of others when making their decisions and least of all a theory of how people think *others* weight the utilities of others when making their decisions. Utility conditionals offer a useful tool to explore this uncharted territory and to discover the naïve axioms of reasoners' folk theory of decision.

References

Amgoud, L., Bonnefon, J.-F., & Prade, H. (2005). An argumentation-based approach for multiple criteria decision. *Lecture Notes in Computer Science*, *3571*, 269–280.

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

Bonnefon, J.-F. (2012). Utility conditionals as consequential arguments: A random sampling experiment. *Thinking & Reasoning*, *18*, 379–393.

Bonnefon, J.-F. (2016). Predicting behavior on the basis of arguments from consequences. In F. Paglieri, L. Bonelli, & S. Felletti (Eds.), *The psychology of argument: Cognitive approaches to argumentation and persuasion* (pp. 1–15). London, England: College Publications.

Bonnefon, J.-F., Girotto, V., & Legrenzi, P. (2012). The psychology of reasoning about preferences and unconsequential decisions. *Synthese*, *187*, 27–41.

Bonnefon, J.-F., Haigh, M., & Stewart, A. J. (2013). Utility templates for the interpretation of conditional statements. *Journal of Memory and Language*, *68*, 350–361.

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.

Bonnefon, J.-F., & Sloman, S. A. (2013). The causal structure of utility conditionals. *Cognitive Science*, *37*, 193–209.

Cooper, D. J., & Kagel, J. H. (2015). Other-regarding preferences: A selective survey of experimental results. In J. H. Kagel & A. E. Roth (Eds.), *The handbook of experimental economics* (Vol. 2, pp. 217–288). Princeton, NJ: Princeton University Press.

Corner, A., & Hahn, U. (2007). Evaluating the meta-slope: Is there a slippery slope argument against slippery slope arguments? *Argumentation*, *21*, 349–359.

Corner, A., Hahn, U., & Oaksford, M. (2011). The psychological mechanism of the slippery slope argument. *Journal of Memory and Language*, *64*, 153–170.

Couto, M., Quelhas, A. C., & Byrne, R. M. J. (2017). Advice conditionals about tips and warnings: Interpretations and inferences. *Journal of Cognitive Psychology*, *29*, 364–380.

Demeure, V., Bonnefon, J.-F., & Raufaste, E. (2009). Politeness and conditional reasoning: Interpersonal cues to the indirect suppression of deductive inferences. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *35*, 260–266.

De Vito, S., & Bonnefon, J.-F. (2014). People believe each other to be selfish hedonic maximizers. *Psychonomic Bulletin & Review*, *21*, 1331–1338.

Elqayam, S., Thompson, V. A., Wilkinson, M. R., Evans, J. St. B. T., & Over, D. E. (2015). Deontic introduction: A theory of inference from is to ought. *Journal of Experimental Psychology: General*, *41*, 1516–1532.

Evans, J. St. B. T. (2005). The social and communicative function of conditional statements. *Mind & Society*, *4*, 97–113.

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

Feteris, E. T. (2002). A pragma-dialectical approach of the analysis and evaluation of pragmatic argumentation in a legal context. *Argumentation*, *16*, 349–367.

Gazzo Castañeda, L. E., & Knauff, M. (2016a). Defeasible reasoning with legal conditionals. *Memory & Cognition*, *44*, 499–517.

Gazzo Castañeda, L. E., & Knauff, M. (2016b). When *will* is not the same as *should*: The role of modals in reasoning with legal conditionals. *Quarterly Journal of Experimental Psychology*, *69*, 1480–1497.

Gazzo Castañeda, L. E., Richter, B., & Knauff, M. (2016). Negativity bias in defeasible reasoning. *Thinking & Reasoning*, *22*, 209–220.

Hahn, U., & Oaksford, M. (2006). A Bayesian approach to informal argument fallacies. *Synthese*, *152*, 207–236.

Hahn, U., & Oaksford, M. (2007). The rationality of informal argumentation: A Bayesian approach to argument fallacies. *Psychological Review*, *114*, 704–732.

Haigh, M., & Bonnefon, J.-F. (2015a). Conditional sentences create a blind spot in theory of mind during narrative comprehension. *Acta Psychologica*, *160*, 194–201.

Haigh, M., & Bonnefon, J.-F. (2015b). Eye movements reveal how readers infer intentions from the beliefs and desires of others. *Experimental Psychology*, *62*, 206–213.

Haigh, M., Ferguson, H. J., & Stewart, A. J. (2014). An eye-tracking investigation into readers' sensitivity to actual versus expected utility in the comprehension of conditionals. *Quarterly Journal of Experimental Psychology*, *67*, 166–185.

Haigh, M., Stewart, A. J., Wood, J. S., & Connell, L. (2011). Conditional advice and inducements: Are readers sensitive to implicit speech acts during comprehension? *Acta Psychologica*, *136*, 419–424.

Hilton, D. J., Kemmelmeier, M., & Bonnefon, J.-F. (2005). Putting *if*s to work: Goal-based relevance in conditional directives. *Journal of Experimental Psychology: General*, *135*, 388–405.

Hoeken, H., Šorm, E., & Schellens, P. J. (2014). Arguing about the likelihood of consequences: Laypeople's criteria to distinguish strong arguments from weak ones. *Thinking & Reasoning*, *20*, 77–98.

Hoeken, H., Timmers, R., & Schellens, P. J. (2012). Arguing about desirable consequences: What constitutes a convincing argument? *Thinking & Reasoning*, *18*, 394–416.

Jara-Ettinger, J., Hyowon, G., Schulz, L. E., & Tenenbaum, J. B. (2016). The naïve utility calculus: Computational principles underlying commonsense psychology. *Trends in Cognitive Sciences*, *20*, 589–604.

López-Rousseau, A., Diesendruck, G., & Benozio, A. (2011). My kingdom for a horse: On incredible promises and unpersuasive warnings. *Pragmatics and Cognition*, *19*, 399–421.

López-Rousseau, A., & Ketelaar, T. (2004). "If . . .": Satisficing algorithms for mapping conditional statements onto social domains. *European Journal of Cognitive Psychology*, *16*, 807–823.

López-Rousseau, A., & Ketelaar, T. (2006). Juliet: If they do see thee, they will murder thee: A satisficing algorithm for pragmatic conditionals. *Mind & Society*, *5*, 71–77.

Schellens, P. J., & De Jong, M. (2004). Argumentation schemes in persuasive brochures. *Argumentation*, *18*, 295–323.

Stewart, A. J., Haigh, M., & Ferguson, H. J. (2013). Sensitivity to speaker control in the online comprehension of conditional tips and promises: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *39*, 1022–1036.

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.

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*, 106–112.

Wood, J. S., Haigh, M., & Stewart, A. J. (2016). This isn't a promise, it's a threat. *Experimental Psychology*, *63*, 89–97.

Wray, H., Wood, J. S., Haigh, M., & Stewart, A. J. (2016). Threats may be negative promises (but warnings are more than negative tips). *Journal of Cognitive Psychology*, *28*, 593–600.

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Stone Serif and Stone Sans by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data

Names: Knauff, Markus, editor. | Spohn, Wolfgang, editor.
Title: The handbook of rationality / edited by Markus Knauff and Wolfgang Spohn.
Description: Cambridge : The MIT Press, 2021. | Includes bibliographical references and index.
Identifiers: LCCN 2020048455 | ISBN 9780262045070 (hardcover)
Subjects: LCSH: Reasoning (Psychology) | Reason. | Cognitive psychology. | Logic. | Philosophy of mind.
Classification: LCC BF442 .H36 2021 | DDC 153.4/3—dc23
LC record available at https://lccn.loc.gov/2020048455