Justification for a Probabilistic Account of Conditionals

Vanessa Streisel

A Dissertation submitted to the Faculty of Graduate Studies in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

Graduate program in Philosophy, York University, Toronto, Ontario

December, 2016

Abstract

In this dissertation I argue that a probabilistic account of conditionals similar to the one proposed by Robert Stalnaker in 1968 is the logical account of conditionals that most aptly models conditional use in natural language. I argue that a probabilistic account of conditionals is best able to account for the most systematic and widespread uses of conditionals in natural language as is evidenced by both its compatibility with the descriptively accurate psychological account, as well as its ability to take into account expert intuitions that diverge from the material conditional interpretation. I provide expert support for Stalnaker's account by describing the ways that a probabilistic conditional can avoid the paradoxes of the material conditional. I argue that the predictive accuracy of the alternative mental models account provides support for the claim that Stalnaker's logical account of conditionals is descriptively accurate. I conclude that both expert and naive reasoners uses of conditional statements are most accurately modelled by a probabilistic account of conditionals similar to that proposed by Stalnaker.

Acknowledgements

This dissertation could not have been completed without the steadfast assistance I received from the following people. I would like to heartily thank my supervisor Judy Pelham for her support and guidance. I would also like to offer my profound gratitude to my committee members who, due to the disparate topics covered in this dissertation, did far more work than typically required of readers and acted as additional advisors: Henry Jackman, Muhammad Ali Khalidi, and Verena Gottschling.

Table of Contents

List of Tables

List of Figures

Introduction: Justification for a Probabilistic Account of Conditionals

The question motivating this dissertation is: “which logical account of conditionals most aptly models conditional use in natural language, i.e., provides the best representational model of the data of conditional use in natural language?” As this question concerns itself with natural deduction, the aspects of conditional use that will be attended to in this dissertation are: the truth conditions for conditional statements, which sentences are considered tautological, and which inference forms are considered valid.

The account of conditionals that I propose is one that provides a logical description of competent conditional use in natural language that is the most broadly descriptive. By “most broadly descriptive” I mean an account that describes: uses of conditionals across different contexts, conditionals represented in different grammatical tenses, and the uses of conditionals by both tutored and untutored subjects. Given the goal of providing a general description of competent conditional use in natural language, this logical account should be able to describe most of the examples of competent uses of the conditional in natural language. For example, the standard truth functional conditional is able to describe conditional use only in certain contexts— it cannot accurately describe natural language use of counterfactuals. My goal is to outline a logical account of conditionals that can describe conditional use in the broadest sense, accounting for as much of the data of competent conditional use (from both expert and non-expert reasoners) in natural language as possible. For a broad sense of conditional use in natural language I include data from experimental psychologists studying conditional reasoning in addition to more traditional examples from philosophical

sources. My main purpose in including data from experimental psychologists studying conditional reasoning is to get a clearer sense of the phenomenon— conditional use in natural language— than we can with the study of grammar and logic alone. Examining psychological theories of conditionals provides a sense of conditional use across different contexts and can show which tense and word choice changes effect subject interpretation. I will defend a particular psychological interpretation of conditional use, then I will argue that this psychological theory is most consistent with a probabilistic logical account of conditionals. Due to having a shared structure with the most descriptively accurate psychological account, a logical account much like the one Stalnaker introduced in 1968 is the best answer to the question: “which logical account of conditionals most aptly models conditional use in natural language?”

I make use of the following constraints in evaluating different psychological accounts of conditionals. The ideal psychological account of conditionals will be one that: corresponds as much as possible to the most systematic and widespread phenomena of competent natural language use of conditionals, that can outline the uses of conditionals in natural language that are taken to be performance errors and describe consistent reasons for these errors, and that allows the normative theory underpinning the account to retain the conventional logical virtues (in order of importance) of consistency, simplicity, inferential power, and completeness. The most broadly descriptively accurate psychological account of conditionals will be the account that is able to simplify the greatest proportion of conditional use in natural language, and give a consistent account of performance errors. The psychological theory must also be consistent with a normative logical theory that maintains the validity of inference forms

that have strong support in the form of expert intuitions, and that rejects as invalid those inference forms that expert intuitions find strongly counter-intuitive.

I ultimately argue for a logical account of conditionals that is probabilistic on the basis that a probabilistic account is the most broadly descriptively accurate, i.e., a probabilistic account of conditionals is best able to account for the most systematic and widespread uses of conditionals in natural language as is evidenced by both its compatibility with the descriptively accurate psychological account (as seen in both predictive power and its ability to coherently describe performance errors), as well as its ability to take into account expert intuitions that diverge from the material conditional interpretation. As the previous paragraphs have made obvious, I diverge from traditional ways of examining the logical question of conditionals by taking into account both formal considerations and data from experimental psychology. As relying upon data from experimental psychology is an unconventional choice, I will take a moment to outline my method and respond to some concerns that may arise from it.

First, it should be noted that I am simply reviewing these psychological studies for their relevant philosophical content and am not involved in an extensive study of their methods. I will be examining a range of studies that involve mostly English language speakers living in the United States or Britain. These studies monitor the behaviour of subjects in conditions designed to elicit their intuitions about conditionals and what makes conditional sentences true. None of these studies observe speakers in natural, unprompted situations. Historically, experimental psychologists studying conditionals presupposed that the standard truth functional interpretation of conditionals is the normatively correct way to reason and designed tasks specifically to see if they

could elicit responses in line with the material conditional. These tasks were designed to test whether the standard truth functional account of conditionals was descriptively accurate (I will refer to this as the "top-down approach"). Most modern psychologists studying reasoning use a more bottom-up approach where the normative theory underpinning the phenomenon are not presupposed when designing the studies. As will be seen in chapter four, I believe this slightly different approach has led to more accurate theories that rely less on *ad hoc* additions.

The introduction of data from experimental psychology complicates the question of which logical account of conditionals is most descriptively accurate since some of the responses on particular reasoning tasks are likely to simply be performance errors. For the purpose of developing a psychological theory of natural language use of conditionals we cannot include every single response. We need a way to determine which responses are descriptively relevant and which are simply the result of processing errors on the part of subjects. My own determinations about which results are relevant appeals to psychologists' own method of sorting between competence and performance error as well as the additional requirement that the most descriptively accurate psychological account will be one that has a normative underpinning that retains as many of the logical virtues as possible.

The Competence Performance Distinction

Experimental psychologists studying reasoning rely upon the competence/ performance distinction introduced first by Chomsky[1] and elaborated on by Cohen[2] as a means of sorting responses into those that are performance errors and those that can be used as evidence of the underlying competence. I will first outline the distinction as drawn by Cohen since this is the understanding that most psychologists have historically replied upon. I will then outline my slightly different use of the performance/ competence distinction that is closer to that introduced by Chomsky. The basic idea for Cohen was that we take a subject's given performance, which we then interpret in an idealized way to account for processing errors caused by the time, attention, and memory limitations of human cognition, to arrive at the underlying competence. Cohen argues that the collective set of human intuitions must be rational by definition. This is because any argument of irrationality presupposes a normative standard that is failing to be met. However, this normative standard is determined by a process of reflective equilibrium, which takes this collective set of intuitions as data. In other words, it is these intuitions that form the basis for a theory of epistemic competence. Cohen's distinction between competence and performance arises out of his explanation of how we determine what the theory of epistemic competence is. He argues that competence can't be identical to performance as it is obviously the case that people can make mistakes. So, we need a way to develop a theory of competence that factors out

---

[1] Chomsky, Noam. (1965).

[2] Cohen, L. Jonathan, (1981).

performance errors such as lack of attention, lack of time etc as these errors do not show competent conditional use. According to Cohen,

> "to ascribe a cognitive competence, in this sense, within a given community is to characterise the content of a culturally or genetically inherited ability which, under ideal conditions, every member of the community would exercise in appropriate circumstances."[3]

Experimental psychologists working on human reasoning use a process like reflective equilibrium to determine what counts as experimental evidence that should be taken as evidence of competence and which can be disregarded as performance error. The process of reflective equilibrium was first outlined by Goodman in *Fact, Fiction and Forecast*[4] but was not labelled with the term 'reflective equilibrium' until Rawls coined the term in *A Theory of Justice*.[5] The basic idea behind reflective equilibrium is that principles of inference can be justified in a circular manner[6] in the sense that "a rule is amended if it yields an inference that we are unwilling to accept; an inference is rejected if it violates a rule we are unwilling to amend."[7] We take the accounts of actual performance, which are the results obtained by experiments on reasoning, and compare

---

[3] Ibid, p. 321.

[4] Goodman, N., (1955).

[5] Rawls, J. (1971).

[6] I take reflective equilibrium to provide evidence for justification and not to be constitutive of justification. For example, being in reflective equilibrium is good evidence for the claim that an inference rule is justified but reflective equilibrium isn't all there is to justification. At least, the relatively narrow sense of reflective equilibrium here isn't constitutive of justification— a truly wide reflective equilibrium might be. See Stich, *(1990)*, p. 71.

[7] Stich, *(1990)*, p. 77.

these to normative theories of reasoning. We consider something to be a performance error if it conflicts with a normative rule that we are unwilling to amend (because of strong support for the rule in the form of other intuitions or for purely logical reasons such as inconsistency or triviality) and we consider the competence to be whichever system best accounts for the observational and experimental data.

In a slight deviation from the method described by Cohen above, the responses that I consider processing errors are errors of attention and not normative performance errors. For example, when I argue that a group of subjects have given responses that can be considered performance errors it is not motivated by a commitment to a particular normative theory (the top-down approach that psychologists used to rely on), rather these errors must be described in a way that does not already presuppose the normative theory that they will support. In other words, when I describe certain responses on reasoning tasks as performance errors I do not mean that these are errors because they fail to accord with a particular normative theory, rather these responses are errors of performance due to things like misreading the question and insufficient processing time. The types of errors that I will be considering performance errors are those outlined by Chomsky in *Aspects of a Theory of Syntax*, "memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic)."[8] In order for a set of responses to be considered a performance error, for my purposes, there must be more motivation than simply that a certain response is inconsistent with a particular normative theory.

---

[8] Chomsky, Noam. (1965), p. 3.

The method experimental psychologists use to determine underlying competence is fallible, meaning that they can be wrong about the system at the heart of human reasoning since whichever system they propose as the underlying competence is a best guess. The best guess is not always reducible to the already normatively accepted system. For example, psychologists can examine which normative theory violates subjects' responses the least and infer that this is the system that best describes those responses. Essentially, they look for a description of subjects' responses that does not require inconsistencies of performance errors[9] and that can accommodate subjects' responses without assuming that the vast majority of responses are performance errors.

I will now summarize the above conclusions about how to distinguish between experimental responses that should be included into a descriptive psychological account of conditionals and those that can be regarded as performance errors. Firstly, responses that are systematic in the sense of being seen across a broad range of different studies by many different groups of subjects are less likely to be performance errors than responses that are seen only in one type of study. The motivation behind this is that responses seen only in one type of study that disappear with minor changes to the question, etc., are most likely to reflect the individual study in question rather than a general notion of inference for conditionals. Secondly, responses that are seen broadly in the sense of being selected by the overwhelming majority of subjects (for example the approximately 90 percent acceptance rate for *modus ponens*) are less likely to be performance errors than responses that are seen by only a small number of

---

[9] By inconsistent performance errors I mean performance errors that are seen intermittently or irregularly. For example, if subjects have particular difficulty in analyzing negated sentences, then this should apply to all negated sentences and not simply negated sentences in one task and not others.

subjects (less than 20 percent). The motivation for this is the notion that it is less likely that a majority of subjects are misreading or failing to read the question than it is that a small number of subjects are doing so.

The general method described above will indicate the most accurate descriptive psychological account of conditionals which shows natural language use of conditionals across different contexts. It should be noted that this dissertation focuses on purely descriptive rather than explanatory psychological accounts. A descriptive psychological account of conditional reasoning describes the phenomena of natural language use of conditionals in the sense of outlining the different responses to conditional reasoning tasks, describing which of these responses are examples of competent conditional use, and describing the reasons for performance errors made by subjects in these tasks. An explanatory account would provide a causal explanation of the mechanisms involved in subject selections, for example, a specific explanation of how subjects select their models.

The constraints supplied by the most descriptively accurate psychological account of conditionals will be used to determine the logical account of conditionals that best fits with this psychological account. This will show that the proposed logical account has the ability to describe the broadest sense of conditional use in natural language. In other words, I will take the logical account that proves to be consistent with the most descriptively accurate psychological account to be in general agreement with the competent use of conditionals in natural language.

Chapter Breakdown

Chapter one is a discussion of the standard truth functional conditional and the paradoxes of the material conditional as outlined by C. I. Lewis. The paradoxes of the material conditional show ways in which the standard truth functional account of the conditional is inadequate to describe conditional use in natural language. I outline some early responses to the paradoxes of the material conditional— C. I. Lewis's fishhook and connection accounts. I argue that these early responses to the paradoxes do not resolve the problems with the material conditional in a satisfactory way. Lewis's fishhook results in strengthened paradoxes and connection accounts fail to block counter-intuitive inferences and do not allow for some inferences that are intuitively valid.

In chapter two I describe a more modern response to the paradoxes of the material conditional in the form of accounts based on assertibility. I outline Quine and von Wright's early description of assertibility and describe Jackson's attempt to retain the standard truth functional account of conditionals for indicatives by relying upon assertibility as a way to explain the paradoxes. I argue that this attempt ultimately fails to correctly describe natural language use of conditionals since empirical evidence shows that intuitions against the material conditional are more widespread than Jackson claims. Jackson argues that the intuitions that conflict with his account can be disregarded since, he claims, David Lewis's triviality proof shows that a probabilistic conditional must result in contradiction or absurdity. I argue that the triviality proof does not in fact discount the possibility of a probabilistic conditional.

In chapter three I describe possible world semantics and outline what I believe is the most promising probabilistic account of conditionals in the form of Stalnaker's 1968

account. I outline some philosophical arguments against this account. Firstly, Jackson's arguments that indicative and subjunctive conditionals should admit of different analyses, and secondly, David Lewis's argument that Stalnaker's selection function is too vague. I defend Stalnaker's account against these criticisms arguing that an account that is unified and that can accommodate semantic underdetermination is in fact better than the alternatives.

In chapter four I outline the experimental support for a probabilistic psychological theory of conditionals, specifically, those proposed by Oaksford and Chater, Schroyens and Schaeken. I begin by describing some biases of reasoning outlined in the psychological literature— the suppression effect, matching bias, negative conclusion bias and confirmation bias. I argue that a probabilistic psychological account offers a more convincing description of these responses than do the alternative theories. I then describe the probabilistic psychological theories of conditional reasoning outlined by Oaksford and Chater, Schroyens and Schaeken. I conclude that most natural language use of conditionals is probabilistic as evidenced by the fact that these probabilistic theories are able to predict a far greater percentage of the responses on conditional inference tasks. I then argue that the alternative mental models account of conditional use shares a basic structure with the logical account of conditionals outlined by Stalnaker. This provides evidence for the claim that Stalnaker's account (or similar probabilistic account) is the most descriptively accurate logical account.

In chapter five I describe some experimental evidence that may challenge a probabilistic account of conditionals. I outline Braine and O'Brien's criticisms of Stalnaker's account and argue that they have only shown that Stalnaker's account fails

to be an adequate *psychological* theory of conditionals. Specifically, Braine and O'Brien argue that the logical theory outlined by Stalnaker is too complex to represent natural language users understanding of conditionals (as natural language users could not reasonably be said to hold information about possible worlds in their minds). The issue with this argument is that Braine and O'Brien examine only the formal description of conditionals found in Stalnaker's account (which is designed to give the truth-conditions for conditional sentences) and not the aspect of his account that was intended to describe actual natural language interpretations of conditionals (the epistemic analysis which gives conditions for belief in a conditional).[10] I argue that the aspect of Stalnaker's account which outlines belief conditions for conditionals includes a sense of possible worlds that is very similar to mental models. Next I describe an experiment that Johnson-Laird and Byrne take as evidence against Stalnaker's hypothesis and Adams's thesis. I argue that the experiment is structured in such a way that it fails to actually test either Stalnaker's hypothesis or Adams's thesis. Lastly, I describe research from Douven and Verbrugge who conclude that Adams's thesis is not descriptively accurate. I argue that this conclusion is stronger than what is warranted by their studies.

I conclude that both philosophical and empirical evidence strongly support the claim that probabilistic accounts of conditionals are the most descriptively accurate. Evidence in the form of expert intuitions such as those examined in chapter one and evidence in the form of naive intuitions such as those examined in chapter four both support the conclusion that most natural language use of conditionals is probabilistic. This I take as strong support for the claim that the logical account of conditionals that

---

[10] See page 68-69 for further discussion of this distinction.

most broadly describes conditional use in natural language will be a probabilistic account.

Chapter One: The Problem of Conditionals

The general idea of seeing classical propositional logic as a model of arguments in English rests on some agreement of form between a formal propositional language and the English language. The atomic sentences stand for basic sentences of English, and the connective symbols stand for logical words in English. Specifically '^' stands for 'and,' '~' for 'not,' '∨' for 'or,' and '→' stands for "if…, then ... ". The most common account of conditionals is what I will refer to as the standard truth-functional account also called the material conditional. By 'account of conditionals' here I mean a formal interpretation of sentences with the 'if…, then …' and similar forms. On the standard truth functional account of conditionals, the truth or falsity of the antecedent and the consequent entirely determine the truth or falsity of the conditional statement itself. The conditional is true in case the antecedent is false or the consequent is true, and is otherwise false. This account contrasts with non-truth-functional (for example, intuitionist) accounts of conditionals where a conditional with a false antecedent may have an indeterminate truth value and accounts where the truth value is probabilistic. Symbolically the material conditional is typically represented with a ⊃ whereas any other conditional is represented with an →— a convention that I will follow.

The material conditional is the account first taught to philosophy and logic students and is the account that most philosophers are familiar with. This popularity has been maintained in the past century in spite of the fact that the material conditional doesn't represent conditional inferences in natural language very well. This is a problem because it means that the system we have designed around this particular account of

conditionals is not necessarily describing conditional inference in natural language. If we want to be able to draw parallels between the rules in our formal system and arguments in natural language (by saying for example that conditionals are transitive), then there has to be shared meaning between the logical particle in natural language and in the formal language.

The counter-intuitive aspects of the material conditional (called the paradoxes of the material conditional) were discovered almost immediately after its creation. In this chapter I describe the development of the material conditional and outline the counter-intuitive aspects of this account, the paradoxes of the material conditional, and some of the controversial statements that this account considers tautologies. I describe some of the early responses to the paradoxes including C. I. Lewis's attempt to resolve the paradoxes by developing an account based on strict implication. My intention is to show that these early attempts to dismiss or deal with the paradoxes fail. In subsequent chapters I outline modern responses to the paradoxes of the material conditional arguing that the best solution to the paradoxes of the material conditional is to adopt a probabilistic conditional similar to that put forward by Robert Stalnaker.

The Material Conditional

Historically, the notion of a truth functional conditional first arose from Frege's new system of logic.[11] In the *Begriffsschrift* a conditional is represented as a function that maps a pair of objects to The False if the first is true and the second false but otherwise maps the pair into The True. The similarities between this function and the

[11] Frege, G. *(*1879)..

material conditional we use now can easily be seen if we compare their truth conditions. The horseshoe symbol was introduced by Russell and Whitehead in *Principia Mathematica* in 1910. This type of conditional was called the material conditional and played a role in the system designed to provide a foundation for mathematics. For the logicists, the purpose of developing an account of conditionals was not to explain or describe the use of conditionals in natural language. Nor was the explicit purpose to represent or reflect correct reasoning in natural language. However, a large part of why the system outlined in *Principia* was so compelling to logicians and philosophers is that it seems to reflect the structures of ordinary language so well.

The material conditional is taken often to apply to all grammatical categories of conditional— i. e., to both indicative and subjunctive conditionals. Indicative conditionals are conditionals where the antecedent value is true, which are almost always expressed in the indicative mood. Subjunctive conditionals are those where the antecedent value is either known to be false or is unknown because of the future tense, which are often expressed in the subjunctive mood. With a few notable exceptions[12] the material conditional is taken to apply to both forms of conditional–- giving us a unified account. Important features of the truth functional account are that it makes the classical inferences of *modus ponens* and *modus tollens* valid. Conditional contrapositive, If (P ⊃ Q) is true, then so is (~Q ⊃ ~P); hypothetical syllogism, If (P ⊃ R) and (R ⊃ S) are true then so is (P ⊃ S); and strengthening of the antecedent, If (P ⊃ Q) is true then so is ((P ^ R) ⊃ Q), are all also valid on this account of conditionals. Counter-intuitive inferences such as denial of the antecedent and affirmation of the consequent are invalid.

[12] Lewis and Jackson both believe that only indicative conditionals should be interpreted using the material conditional.

P	Q	⊃
T	T	T
T	F	F
F	T	T
F	F	T

According to the truth table of the material conditional as shown above, a conditional statement is false only in case the antecedent is true and the consequent false but is otherwise true. This truth table not only preserves many of the intuitive inference forms and rejects the counter-intuitive inference forms but also in many cases matches ordinary language use of the conditional— if we assume the constraint of a bivalent system. For example, if we are forced to select a T/F value for the lines on the truth table where the antecedent is false, we would likely arrive at the truth table for the material conditional since selecting F for these lines seems too strong.[13] The truth functional conditional also has truth-functionally equivalent statements in disjunction and conjunction form. For example, (P ⊃ Q) is equivalent to both (~P v Q) and ~(P ^ ~Q). One of the most positive aspects of this account is that the features of the material conditional allow for conditional proofs with which we can infer a conditional statement if the consequent has been derived from the assumption of the antecedent. For example, if we assume P and are able to derive Q based on this assumption then we can infer P ⊃ Q.

---

[13] In the construction task subjects are given a conditional statement and asked to fill in a truth table. When subjects are given only two possible truth values – T and F – their truth tables mostly mimic the truth table of the standard truth functional conditional. It is only when subjects are given a third value – Unknown – that they change the value for the lines with a false antecedent value. Evans, J. St. B. T., Newstead, S. E., and Byrne, R. M. J. (1993).

Problems with the Standard Truth-Functional Account

The problematic aspects of Russell's material conditional were recognized almost immediately after the publication of the second volume of the *Principia*.[14] In a 1912 article in *Mind*, C. I. Lewis challenged Russell's account of conditionals by arguing that it failed to capture the common notion of implication. Lewis's criticism concentrated on the tautologies ~P ⊃ (P ⊃ Q) and Q ⊃ (P ⊃ Q) arguing that they are not in themselves problematic but that they fail to capture ordinary notions of implication. Lewis claimed that,

> "in themselves, they are neither mysterious sayings, nor great discoveries, nor gross absurdities. They exhibit only, in sharp outline, the meaning of 'implies' which has been incorporated into the algebra."[15]

This sense of implication, according to Lewis, differs from the meaning of implies in "ordinary inference and proof."[16] To show this conflict C. I. Lewis's article discussed two counter-intuitive tautologies that arose from Russell's account of conditionals— the paradoxes associated with material implication.[17]

The first problem outlined by C. I. Lewis is that all conditionals with a true consequent are taken to be true by the material conditional. This means that a conditional like "if I bring my umbrella to campus, then Lake Ontario is a Great Lake" is

---

[14] Russell, B and Whitehead, A. N., (1910, 1912, 1913).

[15] Lewis, C. I. (1912).

[16] Ibid.

[17] These are referred to as "the paradoxes of material implication"— they aren't actual paradoxes, but rather ways that the material conditional conflicts with ordinary language use.

true. According to the material conditional, if a statement is true, then it is implied by any proposition. The other conditional paradox outlined by Lewis is associated with the fact that all conditionals with a false antecedent are true on Russell's account. This would make all counterfactuals like “If the USSR had not been dismantled, then the sun would have disappeared from the sky” true.

Subsequent to C. I. Lewis’s publication, many consequences of the paradoxes of the material conditional have been outlined.[18] One is the tautology ~P ⊃ (P ⊃ ~P). It seems clearly counter-intuitive to imagine that, “my cat is not old” can imply “If my cat is old, then my cat is not old” yet, given the truth-functional account of conditionals, statements of this form are always true.

Strengthening of the antecedent is also a controversial inference form.[19] The idea is that if something implies P, then that something conjoined with any other statement must also imply P. Symbolically we represent this as (P ⊃ Q) ⊃ ((P ^ R) ⊃ Q). This only seems controversial when we interpret the content of P and R in particular ways. For example, the sentence, “if I strike this match, then I can light the candle” if true should not lead to the truth of, “If I strike this match and the candle is under water, then I can light the candle”.

Another problem with the material conditional is that some other formulae that are tautologies under this interpretation seem certainly false when translated into

---

[18] For a description see: Johnson, D. K. (1996), especially pages 94-95, Brandom, R. (1981).

[19] See: Johnson, D. K. (1996).

natural language sentences.[20] For example, (P ⊃ Q) v (Q ⊃ R)[21] can be translated as “It is either the case that if the sun rises tomorrow, cows will talk, or it is the case that if cows talk, then we will all move to a colony on the moon” and ~(P ⊃ Q) ⊃ (P ^ ~Q)[22] can be translated as “it is not the case that if the moon is made of cheese, then pigs have curly tails, therefore, the moon is made of cheese and pigs don’t have curly tails”. One of the most widely questioned consequences of the material conditional is called the paradox of entailment, which is that (P ^ ~P) ⊃ Q.[23] This conditional means that any sentence can be drawn from a contradiction and is related to one form of the classical inference of *reductio ad absurdum*. This sentence is controversial because it seems in some cases to be clearly counter-intuitive–- for example, “my neighbour is both bald and not bald, thus the moon is the centre of the universe.” However, for practical reasons, many argue that this particular sentence should be taken in favour of the material conditional as many want to retain the intuitive inference form of *reductio ad absurdum*. For example, it may seem correct that a contradiction implies anything. Alternately we may be unwilling to reject one of the inferences that leads to the truth of the paradox of entailment (simplification, addition, and disjunctive syllogism).

---

[20] For a list of the conditional tautologies see: Kalish, Montague and Mar (1980). pp. 107-112.

[21] This is theorem 58 in Kalish, Montague and Mar (1980), p. 109.

[22] This can be derived from theorem 40 in Kalish, Montague and Mar (1980), p. 109.

[23] This is the formalization of *ex contradictione quodlibet*. For a discussion see: Kapsner, A. (2013), pp. 77-94.

The inference of conditional excluded middle, represented by the tautology (P ⊃ Q) v (P ⊃ ~Q), is also controversial.[24] It is of course motivated by traditional excluded middle–- the assumption that Q is either true or false–- meaning that at least one of the conditionals represented in the disjunction must have a true consequent. For example, let P: Jeremiah is 50, and Q: Jeremiah is a farmer. P⊃(Q∨~Q) is true in this case, so (P⊃Q) ∨ (P⊃~Q) seems to be true as well. However, we don't have sufficient reason to believe that either of P⊃Q or P⊃~Q on their own is true. For example, we know that if Jeremiah is 50, then Jeremiah is either a farmer or not a farmer but we do not know that, "if Jeremiah is 50, then Jeremiah is/is not a farmer" is true or false. For those who argue that a disjunction is only true if we know the truth of at least one of the disjuncts, we cannot infer, "(if Jeremiah is 50, then Jeremiah is a farmer) or (if Jeremiah is 50, then Jeremiah is not a farmer)" since we don't have enough information to assert that either of the disjuncts is true.

Some argue that concerns with examples such as the Jeremiah conditional above do not show an issue with conditional excluded middle but are caused by the fact that the law of excluded middle itself fails for vague terms. However, we may feel that a sentence like "Tyrone is either bald or not bald" may be true and relevant even if we don't know which of the disjuncts is true without somehow precisifying 'bald'. Whereas we may want to reject a sentence like "If Tyrone is Canadian, then Tyrone is bald; or If Tyrone is Canadian, then Tyrone is not bald" as arbitrary since the antecedent statements are irrelevant to the content.

---

[24] For a discussion of the arguments against CEM see Cross, C. B. (2009), especially pages 174-176.

The problems with the material conditional are not merely that one can come up with statements that have a false antecedent or true consequent yet are clearly false. The counter-examples to the material conditional show something much deeper that is missing from this account. Namely, that the conditional as used in natural language has a meaning that is different from what is represented by the material conditional. The material conditional does not capture our intuitive notions of implication since the connective fails to correctly match the truth or falsity of many statements in natural language. For example, we almost never take conditional statements to be true *merely* on the basis of the truth of the consequent or falsity of the antecedent. It is possible that the conditional in natural language is used to describe states of affairs that are contingent and uncertain whereas the logical conditional is used to express states of affairs that are necessary and certain (this is roughly what C. I. Lewis proposes)[25] so we have two different notions of implication— a regular one and a much stronger logical one. Another interpretation is that the material conditional doesn't capture either the logical or the regular notions of inference since true implication requires that there be some connection between the antecedent or consequent that is not represented by the material conditional (this is roughly the position taken by those who support a connection account).[26] What these accounts have in common is the belief that it is not enough for the consequent to be true in every case where the antecedent is true— there must be something about the antecedent truth that leads to the truth of the consequent. The tricky part is capturing this connection using purely formal elements.

---

[25] Lewis, C. I. (1912).

[26] Dunn, J. M. (1986).

Early Dismissals of the Paradoxes

Many of the initial responses to Lewis' publication took it that these “paradoxes” are not of significant concern since they clearly misunderstand the purpose of implication. The central issue is the fact that material implication as introduced by Russell and Whitehead was not intended to capture ordinary language intuitions about implication. Both Russell and Moore argued that ‘implication’ as used by logicians is a formal term and should not be confused with colloquial uses of this term.[27] One reason why these counter-intuitive elements of the material conditional were not seen as relevant is because the logicists were not attempting to provide a system that modelled natural language use. Thus, its failure to intuitively capture natural language use of conditionals wasn’t seen as a problem. However, philosophers and logicians agreed that the truth functional, material conditional does not seem to be how we analyze conditionals in natural language because in many cases where we believe that the antecedent is false or the consequent is true, we still do not take the conditional to be true. The problem with the material conditional is, as pointed out by C. I. Lewis, that it fails to capture the ordinary notions of implication that justify our logical and mathematical inferences.

Another problem with these early dismissals of C. I. Lewis’s paradoxes, from a contemporary perspective, is that philosophers and others take logical notions (specifically and perhaps most importantly implication) as expressions of norms of human reasoning. Philosophers are interested in a logical theory of conditionals that

---

[27] For a description of Russell’s responses to the paradoxes see: Jager, R. (1972), pp. 126-134. For a description of both Russell and Moore’s responses to the paradoxes see: Haaparanta, L., and Heikki J. K. (eds.), (2012), p. 12.

demonstrates the use of conditional statements in natural language. We are also interested in a theory that can outline valid and invalid inferences associated with the conditional as used in natural language and we are looking for a theory that can describe when it is appropriate or correct to make particular inferences. Given these goals, if this account fails to capture intuitive notions of inference or fails to represent conditional use in natural language, we must look for a better account.

Alternate Conditional Accounts

Since the publication of Lewis's paradoxes, philosophers of logic have been attempting to meet this challenge and develop a better account. Lewis argued that the problem elucidated by his description of the paradoxes is that the standard truth functional conditional isn't strong enough to account for true logical implication. Lewis's proposed solution was to introduce a stronger sense of implication that he called strict implication. In order to define strict implication Lewis makes use of a possibility operator. This is particularly important as this led to the development of modal logic and possible worlds semantics, which are still embraced as solutions to the paradoxes. The possibility operator introduced by Lewis's system is the diamond – ◊ – which can be read as "is possible." So strict implication, represented by the fishhook ⥽, can be defined as ~◊ (P ^ ~Q) or, it is not possible that p is true and q is false. In other words, $A \strictif B$ is true at a world *w* if and only if for all *w'* such that *w'* is accessible to *w*, either *A* fails in *w'* or *B* obtains there. Accessibility here refers to the accessibility relation signified in modal logic by 'R'. w^1 is accessible to *w* if and only if w^1 is possible given the facts of *w*. Thus, this system has two different notions of the conditional, one that is

extensional–- material implication–- and one that is intensional— strict implication as represented by the fishhook. One of the philosophical implications of having two different notions of implication is that this resulted in two different notions of truth. For example, with material implication a conditional statement is true, but a statement is only logically true with strict implication. So a contingent conditional such as “If you drop that glass, then it will break” is true in only the worlds where either you do not drop the glass or the glass breaks. However, a necessary conditional such as “if you drop the glass, then you drop the glass” is true in every possible world.

Lewis’s proposed solution avoids many of the problems of material implication, for example A → (B → A) and ~B → (B → A) are not logical truths (when represented with Lewis’s strict implication) since the connections between B and A may be contingent. In other words, a conditional statement is not logically true or true under strict implication simply in virtue of the consequent truth or antecedent falsity. Since the connection represented by the material conditional is a contingent one, then the truth of the consequent in one possible world does not guarantee that the two propositions will in fact have this connection. If the consequent is true in every possible world or if the antecedent is false in every possible world, then the connection represented by the fishhook becomes true. Thus, if A is necessary or if ~B is necessary, then B ⪡ *A* is true. One consequence of Lewis’s notion of strict implication is that it leads to the strengthened paradoxes of material implication elaborated on below. Lewis’s system also fails to avoid some of the other counter-intuitive results of the material conditional. For example, conditional transitivity is still valid in Lewis’s system in virtue of axiom B6, ((P ⪡ Q) ^ (Q ⪡ R)) ⪡ (P ⪡ R).

Lewis's system is primarily criticized for the fact that it leads to strengthened paradoxes of strict implication that parallel those of material implication. Since strict implication is defined as ~◊ (P ^ ~Q), an impossible proposition implies anything and a necessary proposition is implied by anything. There are many examples where these two consequences lead to counterintuitive propositions such as:

> If 2+2=5, then cats are actually aliens from another planet. (Cases where the antecedent is a necessary falsehood.)
>
> If the conservatives win the next election, then 3 is a prime number. (Cases where the consequent is a necessary truth.)

Lewis argued that these conditionals are not counter-intuitive and that it is appropriate that an impossibility implies anything, since a logical impossibility is a contradiction, however those who went on to develop a connection account of the conditional claim that Lewis's fishhook still fails to correctly describe logical inference. Lewis's solution, doesn't correct for every counter-intuitive result of material implication.For example, the fishhook is also unsatisfactory since it allows for both strengthening of the antecedent and conditional transitivity.

Connection Accounts

These strengthened paradoxes led some[28] to conclude that counterintuitive examples of the material conditional are counterintuitive because there is some relevant connection needed between the antecedent and the consequent. In other words, what is truly going wrong with the paradoxical conditionals is not that the material implication

---

[28] For example: Anderson, A. R., and N. D. Belnap, Jr. (1975), and Dunn, J. M., (1986), pp. 117–124.

of Russell and Whitehead isn't strong enough; what is going wrong is that there is no connection between the antecedent and the consequent in these conditionals. The idea is the common sense notion that something can only follow from a particular argument if there is some kind of thematic connection between the conclusion and what has previously been stated. The intuitions of natural language users do tend to lean toward the idea that conditionals of the form "If Jack is two meters tall, then Jack is an odd number of meters tall" are false in cases where the antecedent is false because of a missing connection between the antecedent and consequent and not potentially true, according to the material conditional, on the basis of Jack not being two meters tall. The difficulty then is trying to account for what seems entirely semantic (the particular content of a proposition) in logical terms. Attempts to capture this content led to connection accounts of implication and relevance logic. The area of relevance logic is so vast and varied that doing justice to the work in this area would require a lengthy section in this dissertation. Further, many of the issues relevant to relevance logic are not specifically relevant to the topic of this dissertation. Since I am interested in an account of conditionals that retains many of the classically valid inference forms, I have decided to concentrate here on the objections to connection accounts in general.

One type of connection account is the covering law account. The covering law account states that there must be some connection between the antecedent and the consequent of a conditional, specifically a causal law. So a conditional like "If I wake up this morning, then the sun will rise" is false because there is no empirical connection between my rising and the sun rising. The covering law account of conditionals is problematic because, while there are cases in natural language where a conditional is

clearly being used to express a law-like connection between the antecedent and the consequent, there are many more uses of conditional statements that are not making use of such a connection. For example, “If you buy me an ice cream, then I will wash your car.” The issue here is that conditional statements are used in a variety of different ways— promises, rules, causal statements, counterfactuals, etc.–- and the type of connection between the antecedent and consequent may be different for each of these different types of conditional, or for some of them there may be no explicit connection between the antecedent or consequent at all.

One objection to connection accounts is that they leave us with the impracticable task of defining what we mean by this connection and then determining or discovering the metaphysical status of this connection. It is also unclear on the covering law account how counterfactual conditionals that do not represent a clear natural law would be evaluated given that the existence of a law-like connection between the antecedent and consequent in these cases could be impossible to establish. It is obvious to most philosophers of logic that the covering law account and the connection account cannot save the truth functional account of conditionals from C. I. Lewis’s paradoxes. This is because there is no satisfying description of what the connection between the antecedent and consequent would be, and no satisfying logical analysis of how to take such a connection into account without relying upon the semantic features of the statement.

Another problem with adopting a connection account, as a replacement for the material conditional, is that there are many aspects of this account that are also counter-intuitive. There are some cases when a conditional with no relevant connection

between the antecedent and the consequent seem intuitively true. For example, Stalnaker argues[29] that connection accounts do not actually leave us with an analysis of conditionals that satisfy the intuitions of natural language users. He gives a counter to connection accounts by claiming that anyone who believes that the United States will use nuclear weapons regardless of what the Chinese do clearly also believes that the conditional "If the Chinese enter the Vietnam conflict, the United States will use nuclear weapons" is true. So it is clear, Stalnaker argues, that there are cases where a person can believe that there is no relevant connection between the antecedent and consequent in a conditional and still take the conditional to be true; meaning that a connection of relevance between the antecedent and the consequent in a conditional statement is sometimes present for natural language users, but sometimes is not.[30] Stalnaker argues that the best way to capture the fact that there is sometimes a required connection between the antecedent and the consequent of a conditional and sometimes not is to have an account of conditionals that uses possible worlds.

I have outlined in this chapter some of the expert intuitions that challenge the standard truth functional account. There are many ways in which philosophers of logic have argued that the material conditional is counterintuitive, but the paradoxes of the material conditional are likely the most compelling. I also examined two early responses to the paradoxes of the material conditional— strict implication and connection accounts — arguing that neither attempt is satisfactory as a descriptive logical account of conditionals.

---

[29] Stalnaker, R., (1968).

[30]This claim seems intuitively correct but, so far as I am aware, has not been empirically tested.

In the following chapters I look at the modern attempts to solve the problem including Stalnaker's possible worlds account and an account based on conversational implicature outlined by Jackson. I argue that Stalnaker's probabilistic conditional captures the intuitions of natural language users better than Jackson's account does and that the adoption of a probabilistic conditional is the best way to respond to the counter-intuitive results of the material conditional.

Chapter Two: Conditional Accounts Based on Assertibility

One response to the paradoxes is to explain the counter-intuitive results outlined by C. I. Lewis by relying upon a distinction between assertibility and truth. This strategy aims to retain the standard truth functional account of the conditional for indicatives. The most successful of these attempts is that put forward by Frank Jackson.[31] Jackson argues that the standard truth functional analysis of the conditional is correct–- i.e., the conditional as used in natural language does in fact have the truth conditions of standard propositional logic–- but that there is some factor aside from simple truth conditions that affects our assent to certain conditional statements. I outline Jackson's view, and describe his attempt to explain the counter-intuitive aspects of his account. I argue that Jackson's attempt to use Lewis's triviality results to explain the counter-intuitive aspects of his account is unsatisfying since the proof does not in fact show what Jackson takes it to show. I conclude by outlining some data from tasks in experimental psychology that show that the intuitions Jackson considers mistakes are widespread. This data challenges Jackson's claim that his theory is descriptively accurate. Alternative accounts of conditionals are able to explain more of the available linguistic/psychological data than Jackson's account and his account is therefore not the most descriptively accurate logical account as he claims.

Quine and von Wright

Some philosophers argue that the paradoxes of the material conditional as outlined by C. I. Lewis do not indicate a problem with the truth table of the material

---

[31] Jackson, F. (1979), p. 573.

conditional but simply show that some uses of conditionals violate conversational norms associated with hypothetical statements. Advocates of an account of conditionals based on assertibility[32] argue that our hearers assume that we would not use a conditional unless there were some connection between the antecedent and the consequent. So, claiming that a conditional is true simply on the basis of a false antecedent doesn't violate the logical rules of conditionals but may violate conversational propriety. Specifically, it would violate conversational rules if the truth of the antecedent wouldn't lead to our acceptance of the conditional. The idea is that if we knew that a false antecedent would make the consequent false, then by uttering the conditional we are violating an unwritten conversational agreement to give the most and most accurate information possible. So a conditional like, "if it rains, I will bring an umbrella" is perfectly acceptable if the antecedent is false, since I would remain committed to the truth of this statement as the utterer even if the antecedent were true. Whereas a conditional like, "if the moon falls out of the sky, then there is life on Mars" violates conversational rules as I would not endorse this statement if the antecedent were true. In the case of the Mars conditional, I'm not giving any information at all except for the fact that I believe the antecedent to be false. This belief would more succinctly be expressed by simply denying the antecedent.

One of the first explanations of the paradoxes in terms of assertibility comes from Quine[33] and von Wright[34] who argued that, in addition to having truth values,

---

[32] Such as Jackson, von Wright and Quine.

[33] Quine, W.V.O., (1960).

[34] Von Wright, G. H. (1957).

conditionals also have conditions of assertibility. Conditions of assertibility are the conditions under which it is appropriate to utter a particular statement. This would make the conditional statements outlined by C. I. Lewis true, and our intuition that they should not be uttered can be explained non-semantically. Quine's idea was that hypothetical statements themselves aren't taken to be true or false but are merely conditional assertions of the consequent. For example, if the antecedent were shown to be false, the conditional would be withdrawn. The obvious problem with this attempt is that it fails to make sense of counterfactuals where the antecedent is known to be false. Natural language users often use conditionals with clearly false antecedents that are taken to be potentially true. For example "If America had entered WWII later than 1942, then the Nazis would have won the war" is a conditional that most would take to be potentially assertible. Most people would not withdraw this conditional due to the fact that we know that the antecedent is clearly false. This is a significant challenge to Quine and von Wright's proposal since people frequently use counterfactuals with obviously false antecedents in meaningful and important ways. We use counterfactuals to examine possible personal futures such as, "If John and Mark had not gotten married …," to examine alternate histories as with the WWII example, and we often use counterfactuals in scientific contexts such as, "If we didn't release so much CO^2 into the atmosphere, global warming would not have occurred." An acceptance of Quine and von Wright's proposal would force us either to claim that people do not realize that the antecedent of these statements is false, or to develop some alternate account of conditionals specifically for counterfactuals. This last option is the one that Jackson adopts in his attempt to use assertibility to explain the paradoxes of material implication.

Jackson's Account of Conditionals

Frank Jackson develops an account of conditionals[35] wherein indicative conditionals are analyzed according to the standard truth functional account and the paradoxes of the material conditional are explained away by appealing to assertion conditions. He gets around the issue faced by Quine and von Wright (that their proposal cannot explain the use of counterfactuals) by offering an account of conditionals that has different analyses for indicative and subjunctive conditionals.[36] Indicative conditionals are analyzed according to the material conditional, but subjunctive conditionals are analyzed probabilistically. The counter-intuitive aspects of indicative conditionals are explained by arguing that indicative conditionals have both rules of truth and rules of assertibility. According to Jackson,

> "the meaning of the indicative conditional 'If P, then Q' is given by a rule of truth and a rule of assertibility. The rule of truth is: 'If P, then Q' is true if and only if 'P ⊃ Q' is true. The rule of assertibility is: 'If P, then Q' is assertible for S at t to the extent that the closest P-probability function to S's at t is a Q-probability function."[37]

By "assertibility" Jackson is referring to the degree to which a sentence is epistemologically warranted for a particular subject. Jackson's rule of assertibility above is a formalization of the idea that a conditional is assertible for a subject if the consequent is robust with respect to the antecedent. For Jackson, robustness means

---

[35] Jackson, F. (1980), p. 568.

[36] Note: for Jackson only counterfactuals are subjunctives.

[37] Jackson, F. (1980), p.136.

that the subject (S) accords a high degree of probability to the consequent being true (the Q-probability function), in the nearest possible case (either the actual world, or the situation most like the actual world) where the antecedent is true (the closest P-probability function). Jackson argues that, the main benefit of describing conditionals in terms of assertibility is that it accounts for the intuition that something like Ramsey's test[38] (where we determine the assertibility of a conditional by hypothetically adding the antecedent to our stock of beliefs) is used to determine the assertibility of a conditional while still retaining the simplicity of the standard truth functional account of conditionals. For Jackson, we determine the truth of an indicative conditional by using the standard truth table, and we determine the assertibility of an indicative conditional based on subjective probability.

There are two distinctions that are necessary to understand the determination of subjective probability for conditionals. The relative/absolute probability distinction refers to whether the assignment of probability is to a single proposition considered in itself (absolute) such as the probability that, "it will rain tomorrow," or whether the assignment of probability is to the probability of a proposition in relation to another proposition (relative) such as the probability that, "it will rain tomorrow" given that, "a high pressure system is moving in." When we refer to conditional probability, we are of course referring to relative probability.

---

[38]The idea of Ramsey's test evolved from a passage in Ramsey F. P. (1929), p. 143 where Ramsey describes a disagreement between two people where one believes that "if P, then Q," and the other believes that "if P, then not Q." Ramsey describes the source of the disagreement, in cases where P is false, as due to each person having a different degree of belief in Q after hypothetically adding the antecedent to their stock of beliefs.

The subjective/epistemological probability distribution versus the objective/ statistical probability distribution distinction is about whether the probability is determined statistically (objective) such as the probability that a dice throw will turn up one pip, or whether we are assigning probability based upon the degree of belief for a person (subjective) such as the probability that the Conservative party will win the next federal election.[39] Subjective/epistemological probability means that a particular subject assesses the assertibility of an indicative conditional based upon their personal assessment of the likelihood of the antecedent and the consequent. For example, if we are analyzing the relative conditional “If it rains tomorrow, then the ballgame will be cancelled” according to subjective probability assignments, then subject (S) assigns a certain probability (say .8) to the conditional based upon how likely they think it is that the ballgame will be cancelled if it rains tomorrow. The objective or subjective probability distribution distinction here refers only to how the probability of the conditional in a particular case is determined and should not be taken to imply subjectivist or frequentist *accounts* of probability. The term “subjective probability” so used, describes epistemic conditions for conditionals and does not have general implications about what probabilities are (as the probability logic proposed by Stalnaker is compatible with either subjectivist or frequentist accounts). A probability function (also called a probability distribution) is a function from propositions to numbers between 0 and 1 that assigns to each proposition a probability.

A counterfactual conditional is assertible, according to Jackson, if the consequent is robust with respect to the antecedent. Similar to the theory of assertibility outlined by

---

[39] Hacking distinguishes these two types of probability by calling the first "statistical" and the second "epistemological". Hacking, I. (1984), p. 12.

Quine and von Wright, robustness means that we would still hold the conditional to be true even if the antecedent were true. Jackson uses David Lewis's poison mushroom example[40] to attempt to show the importance of robustness. In this example, two friends (Jane and Ike) are hiking in the forest and come across some mushrooms. Jane turns to Ike and asks if the mushrooms are poisonous to which Ike replies, "If you eat that mushroom, then you will die." In this example Ike knows that the mushrooms in question aren't poisonous but utters the conditional knowing that it is true since Jane won't eat the mushrooms. Jackson argues that the reason why we feel that there is some deception occurring on the part of Ike is that the hearer assumes that the speaker is giving them information that is robust with respect to the antecedent. Jane assumes that the conditional would remain true even if the antecedent were true. Therefore, Jackson argues, "when considering propriety of assertion we should take account of robustness *as well* as high probability, relevance, informativeness, and so on."[41] Jackson defines his rule of assertibility as follows: "'If P, then Q' is assertible for S at t to the extent that the closest P-probability function to S at t is a Q-probability function."[42] Here we are referring to the things that S knows or believes at time t (the time of her uttering the conditional). P-probability function refers to the probability assigned to P and Q-probability function refers to the probability assigned to Q. This is one formalization of the Ramsey test— the assertibility of a conditional for Jackson is based upon how likely the subject thinks the consequent is, given the truth of the antecedent.

---

[40]Lewis, D. (1976), p. 143.

[41]Jackson, F. (1979), p 573.

[42]Jackson, F. (1980), p.136.

Jackson argues that there is linguistic evidence, by which he means "the linguistic practices and intuitions of speakers concerning conditionals and related construction,"[43] in support of the claim that the conditional is not the only logical term to have both truth conditions and conditions of assertibility. Jackson argues that, "we use the 'P or anyway Q' construction to indicate that 'P v Q' is robust with respect to ~P, but not with respect to ~Q."[44] According to Jackson, there are no specific linguistic cues signalling the robustness of "P → Q" with respect to P because the very fact that we have used a conditional statement indicates that the consequent is robust with respect to the antecedent.

Since there are no specific semantic cues signalling robustness, Jackson's assertion that common intuitions that conflict with the standard truth functional view of conditionals are explainable in terms of this robustness seems ad hoc. Jackson tries to avoid this criticism by describing the observations of Dummett[45] which he takes to lend support to his claim that conditionals have both truth conditions and assertibility conditions. Dummett argues that for many logical constructions in ordinary language, a perceived difference in meaning does not necessarily have to be explained as a change in truth conditions. For example, according to Dummett, the sentences "P and Q" and "P but Q" have the same truth conditions,

---

[43] Jackson, F. (1984), p. 67.

[44] Jackson, F. (1979), p 575.

[45] Dummett, M. (1973), *Frege*, Duckworth: London.

> ‘‘the difference in meaning between 'but' and 'and' is just as objective a feature, requiring to be grasped by anyone who wishes to speak English, as that between 'and' and 'or'.’’[46]

Jackson intends for Dummett’s argument about changes in meaning that are not reflected in a change of truth values to lend support to his claims about assertibility. The main problem here is that no one denies that there are nuances of meaning in natural language that cannot be captured by truth conditions. The question is whether the change in meaning that is reflected by the paradoxes of the material conditional can be reflected in terms of truth conditions or not. To this end, the example of ‘but’ above is a poor example to use as many uses of ‘but’ do indicate a change to the truth value of the first conjunct. We interpret ‘but’ as a straightforward conjunction because it is easier (i.e., attempting to model the cases where ‘but’ signals a slight change in the truth value of the first conjunct is extremely difficult)–- not because it is the most accurate interpretation of the meaning of ‘but.’ An obvious difference between ‘but’ and ‘if …, then …’ is that we have a purely semantic explanation of conditionals (in terms of probability) that can accommodate these differences in meaning whereas we have no such alternate account to semantically explain the nuances of the meaning of ‘but.’

Aside from explaining away the paradoxes of the material conditional, Jackson also believes that a major benefit of his account is that it explains the intuition that conditionals are in some sense based on subjective probability while still retaining the standard truth conditions for indicative conditionals. Some of the other purported benefits of Jackson's theory are that strengthening the antecedent, hypothetical

---

[46] Ibid., pp. 85-86.

syllogism, and contraposition are all still valid. As described in the previous chapter, these inference rules have counter-intuitive results that are similar to the paradoxes of the material conditional so some may not consider the validity of these inference forms a benefit. Further, the characteristics cited by Jackson in support of his theory are only benefits if his account has enough support in the form of intuitions and natural language evidence.

Jackson himself acknowledges that conditionals are in some sense about probability. However, accepting a probabilistic account for indicatives may result in violating logical norms, e.g., contradiction, or severely limiting the usefulness of conditional statements, e.g., triviality.[47] For Jackson, the simplicity of the material conditional and the logical issues that arise from combining conditional logic with probability logic justify an account of conditionals that considers more of the intuitions about conditionals to be defective. Jackson argues that the intuitions that run counter to his account are atypical. If he were correct in asserting that there are relatively few cases where his account fails to match up with intuitions about conditional use, then it would make sense to adopt his account where we get to retain the simplicity of the material conditional. On the other hand, if it can be shown that Jackson's account conflicts with many widespread intuitions about conditional use, then this account may simplify conditional use too drastically to be considered descriptively accurate. In the final sections of this chapter I will show that the linguistic/psychological evidence overwhelmingly conflicts with Jackson's account of conditionals. Natural language users rarely rely on a standard truth functional interpretation of conditionals even with

---

[47] Discussion of the triviality proofs begins on page 47.

straightforward conditionals (i.e., not the conditionals that are recognized as problems like those referred to in the paradoxes of the material conditional). Therefore, the truth conditions that Jackson's account outlines for indicative conditionals does not reflect natural language use of conditionals and the assertibility conditions that Jackson outlines for conditionals cannot entirely reconcile those cases where intuitions fail to correspond to the standard truth functional account. Given the fact that Jackson's account conflicts in significant and widespread ways with natural language use of conditionals, the only motivation for accepting this account would be if the alternative failed to accord with logical norms in some significant way. Jackson argues that the intuitions that run counter to his account should be disregarded because a probabilistic account of conditionals would result in absurdity or triviality. I outline a simplified version of the triviality results to show that the conclusion "all probabilistic accounts of conditionals will result in triviality or absurdity" cannot be drawn from it. Since the triviality results do not show that any probabilistic conditional will lead to absurdity or triviality, there is no motivation for accepting the less descriptively accurate account of conditionals outlined by Jackson.

How Jackson's Account Deals with the Paradoxes

The essay ‘‘On Assertion and Indicative Conditionals"[48] is ostensibly Jackson's attempt to show how robustness can explain away the paradoxes of the material conditional. However, the real purpose of this essay is to explain why the intuitions of natural language users are frequently mistaken when it comes to conditional

---

[48]Jackson, F. (1979), pp. 566 - 567.

statements. It is important to note that rather than offering an account that explains these intuitions, Jackson is attempting to argue that the intuitions drawn out by the conditional paradoxes can be ignored. Jackson's argument gives (fallacious) reasons to reject natural language users intuitions and doesn't take linguistic/psychological evidence seriously enough. Specifically, Jackson argues that the intuitions that conflict with his account are due to a mistaken commitment to a probabilistic conditional (where the truth of a conditional is determined by the probability of the consequent given the antecedent, i.e., P(C/A)).

Jackson argues that one way to explain away what he calls counterexamples to the material conditional (the paradoxes of the material conditional) is to rely upon the Gricean maxim "assert the stronger instead of the weaker." Jackson's explanation of the paradoxes also relies upon the equivalence thesis— $(P \supset Q) \equiv (\sim P \vee Q)$ meaning that Jackson's explanation of the paradoxes of the material conditional require intuitions about conditionals and disjunctions to be analogous. Jackson claims that C. I. Lewis's first two paradoxes— $(Q \supset (P \supset Q)$ and $(\sim P \supset (P \supset Q)$— can be explained away by arguing that uttering "If P, then Q" on the basis of knowing either ~P or knowing Q is logically reliable but violates conversational norms. This is because the statements ~P or Q alone are stronger than the conditional itself. As Jackson puts it,

> "the Equivalence theorist explains away the impropriety of asserting (P ⊃ Q) when one of ~P or Q is highly probable by saying that in such a

> case you should come right out and assert the logically stronger statement, namely, either ~P or Q, as the case may be."[49]

If I know that my friend will not eat the mushrooms or if I know that my friend will die, I should utter those stronger, simpler sentences and not utter the conditional "If you eat those mushrooms, then you will die" at all. According to Jackson, our intuitions about the conditional paradoxes outlined by C. I. Lewis have nothing to do with the truth or falsity of these statements but are actually intuitions about their low assertibility.

Jackson also attempts to explain the counter-examples that are associated with certain tautologies of the material conditional, for example, ((P ⊃ Q) v (Q ⊃ R)). As discussed in chapter one, according to the standard truth functional account of conditionals, statements of this form are always true; however, most natural language speakers would not assent to a statement like, "It is either the case that if I wake up tomorrow, then I will be out of coffee, or it is the case that if I am out of coffee, giant monsters will attack Toronto." Jackson argues that these tautological disjunctions are also logically true but not highly assertible. This disjunction, in spite of being logically true, is not assertible because for neither of the conditional statements included in the disjunction is the conditional robust with respect to its antecedent. For example, the disjunct on the left, "if I wake up tomorrow, then I will be out of coffee" is not robust with respect to its antecedent because I would not assert it if the antecedent were true.[50] Likewise, according to Jackson, the conditional on the right, "If I am out of coffee, then giant monsters will attack" is not robust with respect to its antecedent. When one utters

---

[49] Ibid., pp. 566 - 567.

[50] Since, as it happens, I am not out of coffee.

a conditional based on the disjunction above, they do not have reason to assert either one of the disjuncts. So, the disjunction itself, while being logically true, is not assertible. According to Jackson, "if the equivalence thesis is true, then ((P ⊃ Q) v (Q ⊃ R)) is a logical truth. But evidently it is not in general highly assertible."[51]

Jackson addresses one of the more obvious ways that his account conflicts with natural language intuitions about conditionals which is the fact that our intuitions about the assertibility of disjunctions and conditionals are not analogous. He claims that,

> "if the standard way of trying to explain away the paradoxes is right, 'or' and '⊃' are on a par. It would, for instance, be just as wrong, and just as right, to assert 'P or Q' merely on the basis of knowing P as to assert (P ⊃ Q) merely on the basis of knowing not P." [52]

According to Jackson's view, a conditional is equivalent to it's corresponding disjunction, i.e., (P ⊃ Q) is equivalent to (~P ∨ Q). Jackson argues that, the conversational rules for the assertibility of conditionals mirror that of disjunctions.

Conversational rules for disjunctions and conditionals, contra the claims of Jackson, are not parallel since people's intuitions about the wrongness of uttering a disjunction on the basis of knowing one of its disjuncts is far less strong than the wrongness associated with uttering a conditional on the basis of knowing that its antecedent is false. It seems reasonable in the case of disjunctions to say that the difference in meaning is due to assertibility alone, but in the case of conditionals the

---

[51] Jackson, F. (1979), p. 568.

[52] Ibid., pp. 568 - 569.

difference in meaning seems clearly to reflect on the truth value. For example, a disjunction is not always assertible simply on the basis of knowing one of the disjuncts to be true. If I say to you, ‘‘Madeleine is either in Paris or in Berlin’’ on the basis of knowing that she is in Paris you would feel misled because I have asserted the weaker statement instead of the stronger— I have given you less information than I had available and implied that I knew less than I did. The difference between the disjunction and the conditional is not a difference in mere assertibility. Assertibility alone cannot account for the strength of natural language intuitions about conditional truth. Natural language speakers are willing to (perhaps grudgingly) accept the fact that what I have said to you when I say, “either it is not raining in Ottawa or the world will end’’ is true on the basis of my knowledge that it is not raining in Ottawa, whereas they are loath to accept that a statement like, ‘‘If it is raining in Ottawa, then the world will end’’ is true because it is not raining in Ottawa. There is not merely a difference in degree— these intuitions are radically different.[53] A conditional that is asserted merely on the basis of a false antecedent is not simply misleading or keeping back some information from the hearer. The statements outlined by C. I. Lewis are problematic because they seem, by both expert and naive reasoners, to be clearly false. In other words, our intuitions about conditionals are about their truth conditions— not merely about the propriety of asserting them as Jackson claims.

---

[53] To the best of my knowledge there are no experimental studies about these intuitions but the contrasting intuitions about these cases are frequently written about by those working in the area— for examples see: Bennett, J. (2003), p. 38, and Edgington, D., Woods, M., and Wiggins, D. (1997a), pp. 103-104.

Jackson claims that these problematic conditionals have low or zero assertibility but are still true, whereas most people would argue that conditionals such as those outlined by C. I. Lewis are both not assertible and false. One example of such a conditional outlined by Jackson is the conditional, ‘‘If Carter weighs 100 kilograms, then he weighs an odd number of kilograms.’’ Jackson claims that,

> ‘‘I must allow that 'If Carter weighs 100 kilograms, then he weighs an odd number of kilograms' is true because its antecedent is false, and say that what is wrong with it is a very low (zero) assertibility.’’[54]

Jackson’s response to this problem is to claim that people’s intuitions to the Carter conditional above are simply mistaken because they think that conditional truth is governed by conditional probability. Jackson argues that the linguistic/psychological data that conflicts with his account (our intuitions that these conditionals are simply false) can be ignored since it relies upon what he thinks is a mistaken account of conditionals. Jackson’s claim is that the triviality proofs of David Lewis show that there can be no non-trivial account of conditionals based on subjective probability— so these (prevalent) intuitions must be simply wrong. However, recent work by Alan Hájek makes a good case that the triviality proofs do not show that a probabilistic account of conditionals will always result in triviality. In the section that follows I will give a brief description of the first triviality proof presented by David Lewis and then I will outline the reasons why the conclusion drawn by Lewis (that probabilistic conditionals lead to triviality) should be questioned.

---

[54]Jackson, F. (1980), p.132.

Triviality and Conditionals

In his paper “Probabilities of Conditionals and Conditional Probabilities”[55] David Lewis, in the minds of many philosophers, ended the debate about whether the conditional connective can be appropriately expressed as a probability function. In this paper, Lewis attempts to show that the assumption that probabilities of conditionals equal conditional probability results in absurdity or triviality. Lewis argues in this paper that, “there is no way to interpret a conditional connective so that, with sufficient generality, the probabilities of conditionals will equal the appropriate conditional probabilities.”[56] Specifically this proof attempted to show that Stalnaker's hypothesis, $P(A \rightarrow C) = P(C/A)$ if $A > 0$, is not a feasible definition of indicative conditionals, and neither is any alternative probabilistic account of conditionals. Many took the proof to show that any attempt to develop an account of conditionals based on conditional probability would either require a trivial language (the language would need to have less than three independent propositions) or would lead to a contradiction. The proof was considered a reductio of the notion that we can model natural language use of conditionals using conditional probability. The proof aims to show that the adoption of Stalnaker's hypothesis leads to triviality results unless we adjust our system in some pretty serious ways, i.e., rejecting certain axioms of probability theory, accepting the claim that conditionals do not really have truth values (a la Edgington[57] who argues that conditionals with false antecedents simply have no truth value), or rejecting

---

[55]Lewis, D. (1976).

[56] Ibid., p. 298.

[57]Edgington, D. (1995).

conditionalization. I will argue that the first option, rejecting certain axioms of probability theory, possibly what Hájek refers to as RATIO[58] (the ratio analysis of conditional probability, i.e., P(A/B) = P(A&B)/P(B), given that P(B) > 0), is the most promising of the above three suggestions.

Lewis's proof relies upon some of the basic axioms of probability logic:

(1) $1 \geq P(A) \geq 0$

(2) If A=B, then P(A) = P(B)

(3) If A and B are incompatible, then P(A v B) = P(A) + P(B)

(4) If A is necessary, then P(A) = 1

Lewis's proof also relies upon some derived axioms:

(5) P(C/A)=P(A&C) ÷ P(A) (Ratio)

(6) P(A) = P(A&C) + P(A&~C) (Addition theorem)

(7) P(A&C) = P(C/A) × P(A) (MVRF)

(8) P(A → C) = P(C/A) (Stalnaker's hypothesis)

(9) P((C→D)/A) = P(D/C&A) (if-and theorem)

We start with:[59]

(10) P(Q→R)

(11) P(R&(Q→R)) + P(~R&(Q→R)) (line 10 and line 6)

(12) (P((Q→R)/R) × P(R)) + (P((Q→R)/~R) × P(~R)) (line 11 and line 7)

---

[58] Hájek, A. (2003).

[59] I am reproducing here a simplified version of Lewis's proof roughly based on Blackburn, S. (1986), pp. 201–232.

(13) (P(R/(Q&R)) × P(R)) + (P(~R/(Q&R)) × P(~R)) (applying line 9 to the 1st and 3rd parts of line 12)

(14) (1× P(R)) + (0 × P(~R)) (simplifying line 13 to represent that P(R/(Q&R) is 1 and P(~R/(Q&R) is 0.

(15) P(R) (simplifying line 14)

We have gone from P(Q→R) to P(R) meaning that, given the assumptions above, the probability of Q→R is the same as the probability of R alone.

Lewis brings up an example that shows the absurdity of this conclusion. Suppose that P is our subjective probability assignment of a fair dice toss. Q is the statement that an even number turns up and R is the statement that a 6 comes up. The probability of Q and R is roughly .1666 and the probability of Q and ~R is roughly .333 (so both values are positive). However, the probability of R given Q is roughly .333 and the probability of R is roughly .1666— the values are, contra the proof above, not equal.

Problems arise from this conclusion even when looking at subjective probabilities, for example if P(A→ C) reduces to P(C), then when you say, “if it rains tomorrow, then the baseball game will be cancelled” you are merely asserting the probability that the consequent is true. Even without the contradiction shown by the dice example, this seems absurd, i.e., even when we are dealing with real life probabilities that are not certain or known.

The upshot of Lewis’s proof is that values for P(C/A) and P(A→C) differ, Lewis concludes that P(C/A) does not equal P(A→C) and the assumption that they are equivalent results in absurdity (the conditional becomes meaningless since it is the same as uttering the consequent alone) or triviality (we have less than the three

independent propositions needed to arrive at the proof).

The main problem with Lewis's proof is that it assumes that conditional probability can be analyzed in terms of unconditional probability. In other words, Lewis assumes that the axiom P(C/A) = P(C&A)/P(A) (which Hájek has called RATIO[60]) is appropriate. What this axiom means is that the probability of C given A is the probability of C and A divided by the probability of A. The implication of this is that conditional probabilities can be determined by unconditional probabilities. Natural language users, both expert and naive, rarely determine conditional probabilities by relying upon unconditional probabilities. For example, my assessment of the probability that, “If it rains tomorrow, then I will carry an umbrella” is not determined by calculating the absolute probabilities that make up the antecedent and the consequent. More concerning for this proposed definition, we are often able to calculate the probability of C given A without having any idea about the individual probabilities of C or A. Many authors have argued that this axiom is only appropriate when moving from left to right and not when moving from right to left.[61]

The upshot of Lewis’ proof is supposed to be that certain intuitions about conditional meaning when combined with certain laws of probability theory so conceived result in triviality. Lewis favours a rejection of the intuitions rather than a re-examination of probability theory, as this would “burden us with too much work still to be done.”[62] I argue instead that these intuitions should be taken seriously. Firstly, because there are

---

[60] Hájek, A. (2003).

[61] For examples see Bennett, J., (2003), p. 53, and Edgington, D., (1997b).

[62]Lewis, D. (1976), p. 305.

reasons to be suspicious of (5)[63] that are independent of Lewis's proof, as I will show below in the discussion of Hájek's article. Secondly, because expert intuitions support the descriptive accuracy of Adams's thesis.[64] Thirdly, because untrained intuitions support the notion that conditionals are interpreted probabilistically, i.e., there is good empirical support for the claim that untrained reasoners employ something like Stalnaker's hypothesis when analyzing conditional statements.

In "What Conditional Probability Could Not Be"[65] Alan Hájek attempts to cast doubt on the notion that RATIO is an adequate analysis of conditional probability. The difficulties with RATIO arise when the assigned values for A or B are vague (Hájek gives values for A or B that fall within a certain range), or when there is a probability gap for either A or B (when A or B has no assigned value).[66] Hájek argues that an analysis of conditional probability that does not allow for vague or unknown values does not reflect scientific or ordinary language use. Hájek argues that, unlike in the situation where our intuitions tell us that the earth is flat, our intuitions regarding conditional probability are actually pretty reliable. Furthermore, and more importantly, the fact that RATIO cannot give conditional probability for conditionals where the antecedent or consequent is unknown or vague does not reflect conditional use where we often assign values to

---

[63] $P(C/A)=P(A\&C) \div P(A)$

[64]Adams's thesis is roughly that the assertibility of a conditional correlates or moderately correlates with the probability of the consequent given the antecedent. $As(A \rightarrow C) \approx P(C/A)$. Adams, E. (1975), p. 3.

[65]Hájek, A. (2003).

[66]Hájek also discusses issues with infinitesimal probabilities in this article but I have left this out of the discussion as there are no common sense uses of conditionals where our probability assignments are infinitesimal.

P(A→C) when A or C is vague or unknown. In other words “there are cases in which the probabilities of conjunctions required by the right-hand side... are vague or undefined, and yet the corresponding conditional probabilities are defined.”[67]

Hájek brings up a number of examples that show why calculating conditional probability in terms of unconditional probabilities is problematic. In the first example we are asked to give our subjective probability that the Democrats will win the next election. Hájek argues that most likely we all assign a vague value (we have some idea of what the probability is within a range) say .6 to .8. We are then invited to give our subjective probability that the Democrats will win the next election, given that the Democrats win the next election. Clearly the probability that this conditional is true is 1.[68] We have a case where the unconditional probability values are vague but the conditional probability is sharp. Another compelling example is 'if the Democrats win the next election, then this fair coin will land on heads'. The antecedent is assigned a vague value but the conditional probability is clearly .5.

Hájek gives another example where the antecedent probability is undefined and yet the conditional probability is clearly defined. For example, what is the probability that this fair coin will turn up heads, given that I fairly toss it? Hájek argues that we have no ability to assign a probability to the antecedent as we are given no information about my inclination to toss the coin, nor about my disposition to toss such a coin fairly–- we haven't the ability even to assign a vague value to the antecedent. Yet the probability that the coin lands on heads, given that I toss it fairly is clearly .5.

---

[67] Hájek, A. (2003), p. 291.

[68] Hájek, A. (2011), p. 10.

Hájek argues also that scientific uses of probability (for example, quantum mechanics) conflict with RATIO as frequently conditional probability is considered to be determinate in these fields even in cases where one of the unconditional probabilities is vague or unknown. Hájek concludes then that scientific uses of probabilities and everyday uses of probabilities both conflict with an account of conditional probability that is calculated on the basis of unconditional probabilities. Even if Lewis is correct and coming up with a new account of subjective probability is too much work, this still means that the triviality proof actually shows either that there is a problem with the axiom used in line 5 of Lewis's proof,[69] or the probabilistic account of conditionals given in line 8 of the proof. Jackson relies on this proof to show that natural language users intuitions that conditional truth is probabilistic must be mistaken and can simply be ignored. However, it is not the case that these intuitions will ultimately lead to triviality— it may be that the best account we have of subjective probability is simply flawed.[70] Jackson is wrong to discount these intuitions and ignore the fact that there is an account of conditionals that can better explain them.

Further, even though Jackson claims that, ‘‘our data are the linguistic practices and intuitions of speakers concerning conditionals and related construction’’[71] his argument relies upon the rejection of most of this data. His claim that these intuitions

---

[69] P(C/A)=P(A&C) ÷ P(A)

[70] I show that Lewis's initial triviality results may speak against using RATIO as an axiom in probability logic. This means that Stalnaker's theory needs to be revised in order to avoid the triviality results. Specifically, Kolmogorov's calculus cannot be used if these triviality results are to be avoided. Thus, the account of conditionals offered by Stalnaker in 1968 requires some minor revisions in order to avoid triviality.

[71] Jackson, F. (1984), p. 67.

can be ignored as they are based on a mistaken understanding of conditionals (a probabilistic one) is unconvincing as the only evidence offered in support of this claim is Lewis's triviality proof. As shown above, Lewis's proof does not in fact show that all probabilistic conditionals result in absurdity or triviality. In the following section I aim to show that natural language users intuitions and practices clash in numerous and widespread ways with a standard truth functional view of conditionals. The intuitions of natural language users conflict with the standard truth functional account of conditionals across many different contexts.

Experimental Data Against Jackson's Account

In the following section I will be examining some of the experimental data that challenges Jackson's account of conditionals. Subject responses on a multitude of different tasks challenge the view that the conditional as used by natural language users is the material conditional— unless we take the subjects across all of these very different studies to be radically misinterpreting the tasks. A philosopher of logic like Jackson does not have such an explanation available to him since his account of conditionals is an attempt to capture natural language use of conditionals. Jackson himself does not refer to psychological studies, but his attempt to defend his account of conditionals by arguing that it accords with natural language use and intuitions about inference (what he calls the linguistic/psychological evidence) means that his account must be able to explain these uses of conditionals. Nowhere does Jackson argue that his account is meant to apply to only uses of conditionals in certain contexts, nor does he specify that, "linguistic/psychological evidence" does not include evidence from

studies on reasoning. He maintains that there are certain types of conditional (such as the Carter example discussed above),[72] which people mistakenly attempt to interpret probabilistically but argues that cases that conflict with the material conditional interpretation are rare. In this section I present data that strongly suggests that natural language users intuitions diverge from the standard truth functional account of conditionals most of the time. Jackson's claim that his account of conditionals explains the linguistic/psychological evidence is false. Not only does most of the data challenge Jackson's account of conditionals, but even more damning, there is an alternative theory that is consistent with far more of the data.[73]

The first type of experimental task that shows how deeply subject intuitions conflict with the standard truth functional account of conditionals is the conditional inference or acceptance task.[74] In this task the subject is given a conditional statement along with another premise and is asked either what follows from the two statements, whether a particular conclusion is valid, or to choose from a list of possible conclusions. For example, the subject is given the statement, "if I have an apple then I have a pear" and the statement, "I do not have a pear" and is asked what, if anything, follows from the two statements. In another common form of this task the subject is also given the statements "I do not have an apple," "I do have a pear," "I have a pineapple" and, "I do have an apple" as a list of possible conclusions and is asked to select all or any of the

---

[72] See page 46 of this chapter.

[73] This would be Stalnaker's probabilistic account of conditionals which will be outlined in chapter three.

[74] For a detailed description and examination of these types of task see: Oaksford, M., Chater N., and Larkin J., (2000).

conclusions that can be validly drawn. In the acceptance task the subjects are given a list of possible conclusions representing the valid inferences of *modus ponens (MP)* and *modus tollens (MT)* and the common invalid inferences of affirmation of the consequent (AC) and denial of the antecedent (DA) and is asked which, if any, of the conclusions follow. For example, a subject may be given the text:

> "If you have apples, then you don't have pears. But you do have pears.
>
> What conclusions can you draw from this?"

In their meta-analysis of a number of different conditional inference task studies, Evans, Newstead, and Byrne[75] conclude that subjects only appear to adopt a truth-functional inference pattern approximately fifty percent of the time. They define a truth functional inference pattern to be a subject who has accepted both *modus ponens* and *modus tollens* and rejected both denial of the antecedent and affirmation of the consequent.[76] This is particularly notable as the conditionals used in these tasks are simple conditionals with only one antecedent and one consequent. These experiments are not testing intuitions about the embedded conditionals that C. I. Lewis outlines or the other complex conditionals that are recognized as examples of conditionals where our intuitions conflict with the standard truth functional account. If any conditionals should be straightforwardly interpreted by subjects as material conditionals, it is these simple conditionals.

---

[75] Evans, J. St. B. T., Newstead, S. E., Byrne, R. M. J. (1993), p. 37.

[76] The studies Evans et al use for this meta-analysis are: Evans (1977), Taplin (1971), Wildman and Fletcher (1977), Marcus and Rips (1979), Kern, Mirels, and Hinshaw (1983), Rumain et al (1983) and Markovits (1988).

Another result that challenges the standard truth functional account is the discovery that there is a significant difference between the acceptance of *modus ponens* versus the acceptance of *modus tollens*. The average acceptance of *modus ponens* across the different studies analyzed by Evans, Newstead, and Byrne is 97.36%, whereas the average *modus tollens* acceptance rate is only 62.18%. If subjects were interpreting the conditional given in accordance with the standard truth functional view, we would expect the acceptance rates for these two forms of inference to be much closer. The large difference in the acceptance rates suggests that the speakers are interpreting the conditional to be non-standardly-truth-functional.

Additional data from these psychological studies that challenge the standard truth functional account of conditionals are the acceptance rates for the invalid inferences of denial of the antecedent and affirmation of the consequent. Firstly, the acceptance rates for these inferences that are considered invalid on the standard truth functional view are extremely high (an average of around 42% for each of them[77]). What this means is that nearly half of the subjects tested across these different studies picked as true the two conclusions that the standard truth functional account deems false. The other problematic aspect of these studies is that the variability across the different studies for the acceptance of the invalid inference forms is extremely high. For denial of the antecedent the acceptance across the different studies examined by Evans, Newstead, and Byrne ranges from 17% to 73% and the acceptance for affirmation of the consequent ranges from 23% to 75%.

---

[77] This average is taken from the following studies: Evans (1977), Taplin (1971), Wildman and Fletcher (1977), Marcus and Rips (1979), Kern, Mirels, and Hinshaw (1983), Rumain et al (1983) and Markovits (1988). There is considerable variation in the acceptance of invalid inference forms in these studies ranging from 17% to 75%.

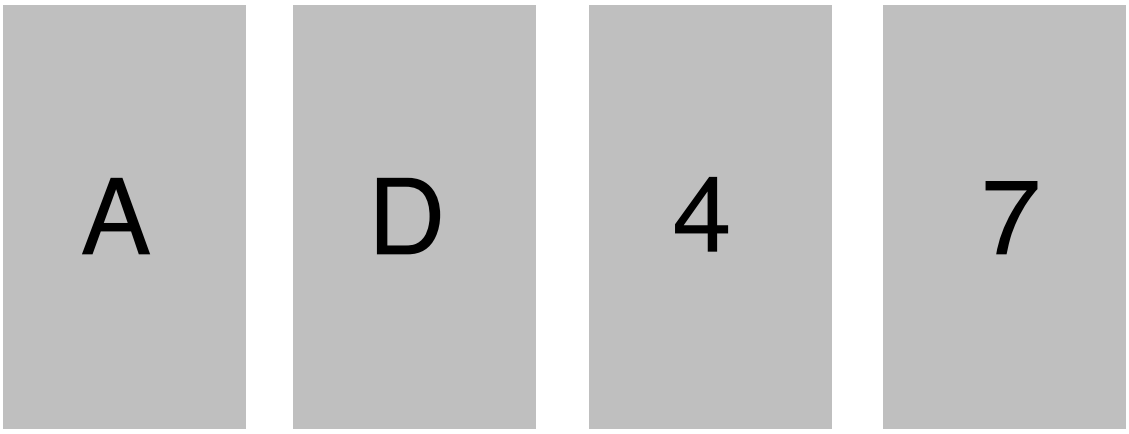
The factors influencing subjects' acceptance or rejection of these inference forms are things like whether the antecedent or consequent themselves are positive or negative, which type of conditional is used,[78] the content of irrelevant additional conditionals given to the subject, whether there is a duration of time between the antecedent and consequent conditions,[79] etc. What is most striking about this data is that most subjects' assessment of validity is based on factors that are not relevant according to the standard truth functional view. The fact that these factors are relevant to which inferences subjects accept cannot be explained on Jackson's view as most of these factors are also irrelevant from the point of view of assertibility. For example, Jackson's account cannot explain the difference between the acceptance of inference patterns between a conditional where the antecedent is negative versus one where the antecedent is positive since these features are not relevant to whether a conditional is robust with respect to its antecedent. The difference between a positive or negatively stated antecedent (when the content is abstract and the subject has no inclination as to which of these formulations is true or false) does not have an effect on whether it is proper to assert a conditional or not.

---

[78] For example, whether the conditional is causal as in "If you turn your key in the ignition, then the car will start," social rules as in "If you put proper postage on your letter, then it will be accepted by the post office," or promises as in "If you give me a hundred dollars, then I will watch your dog".

[79] For example, how long does it take for the antecedent condition to be met. "If you are in the library after the demonstration, then you will be arrested" yields very different interpretations depending on whether you are in the library five minutes after the demonstration or five days after.

The Wason selection task is the most well known task for studying conditional reasoning. In this type of task the subjects are given four cards (or pictures of cards) with a letter on one side and a number on the other. For example,



Subjects are given a conditional statement and asked which cards they need to flip over to falsify or prove the rule expressed by the conditional. For example, the subject may be given the statement "if a card has an A on one side then it has a 4 on the other" and asked which cards they need to flip over in order to falsify or prove this rule. In studies of this form subjects rarely choose to select the cards that will falsify the rule according to the standard truth functional account (the A card and the 7 card). Most commonly either the antecedent alone, or both the antecedent and the consequent jointly are selected (the A card and the 4 card).[80] Wason selection task was developed to study content effects on reasoning and is still used primarily for this purpose. These content effects, as will be discussed in Chapter Four,[81] overwhelmingly support a probabilistic rather than material interpretation of conditionals.

---

[80] Wason, P. C. (1968), and Wason, P. C. (1969), and Wason, P. C., and Johnson-Laird, P. N. (1970), and Wason, P. C., and Johnson-Laird, P. N. (1972).

[81] pp. 92-102 and pp. 115-120.

The divergence between the proscribed selections according to the standard truth functional view and the actual responses on the Wason selection task does not speak against the standard truth functional view as strongly as the results on the other tasks. This is because the Wason selection task is far less straightforward and, due to this, it is possible that subjects are misinterpreting the task itself. In fact a common interpretation of the ‘‘errors’’ on this particular task is to assert that subjects are misinterpreting a closed-scenario problem (where it is appropriate to rely on deductive inference) as an open-scenario problem (where it is appropriate to use inductive inference). One suggestion presented by Oaksford and Chater[82] is that subjects mistake the Wason selection task for an open-scenario problem because they are used to reasoning in real-world situations that involve verification and not falsification. This experience primarily with real-world situations leads subjects to flip over the cards that will verify rather than falsify the conditional given. This could be why nearly all subjects flip over the potentially verifying cards A and 4 (jointly) or A (alone), but less than 10% of the time flip over the potentially falsifying cards of A and 7.[83]

However, the fact that most subjects interpret conditional statements as claims that can be verified but not falsified itself speaks against the standard truth functional view. For example, in a closed scenario problem it is possible to examine all possible options and so the truth values that we assign to the antecedent and consequent can be certain. An example of a closed scenario problem would be if we have a box in front of us with 2 (some relatively small, countable number) black or white balls in the box

---

[82] Oaksford, M., and Chater, N. (2007).

[83] Evans, J. St., Newstead, S. E., and Byrne, R. M. J., (1993) p. 101.

with numbers written on them and are asked whether the claim “If a ball in the box is white, then it has a 4 written on it” is true. It is possible to examine all balls and so, on both the standard truth functional and probabilistic accounts, the process required to falsify this conditional is possible. In an open scenario problem the truth value of the conditional may be probabilistic and so a counter-example may not provide falsification. For example, if subjects have a standard truth functional conditional in mind in an open scenario problem, falsifying a conditional is possible (they simply have to find a case where the antecedent is true and the consequent false). However, if subjects have a probabilistic conditional in mind, the antecedent may not be strictly true and the consequent strictly false, making falsification impossible. In these cases, rather than trying to prove the consequent false, it may make sense instead to examine cases that make the consequent true. For example, if asked to determine whether the conditional, "if you eat raw chicken, then you are likely to get salmonella poisoning” is true we do not evaluate all possibilities since the group of people who do not suffer from salmonella poisoning is so large and open ended. It wouldn't make sense to examine those who do not suffer from salmonella poisoning, rather we would examine people who do have salmonella poisoning to see if they have eaten raw chicken. The results on the Wason selection task, if they can truly be explained by the assertion that subjects are interpreting the task as an open scenario one, suggests that people are more inclined to interpret even the abstract conditionals given as being non falsifiable. According to the standard truth functional account, the antecedent and consequent of conditionals are strictly true or strictly false and falsifying a conditional is as easy as looking for a case where the antecedent is true and the consequent false. Thus, the fact

that subjects interpret these abstract conditionals as being non-falsifiable suggests that they are operating with a probabilistic conditional where falsification is not as easy since the antecedent may not be strictly true or the consequent strictly false.

The construction or evaluation task most clearly indicates subject intuitions about the conditions leading to the truth or falsity of conditionals. This is because this type of task involves having subjects either fill in their own truth table to correspond to a conditional statement or to select one out of possible truth tables. An early study using the evaluation task done by Johnson-Laird and Tagart[84] allowed subjects to deem certain lines on the truth table 'irrelevant'. In this study subjects most often selected the TT case as true (50/96), the TF case as false (64/96) and the FF (37/96) and FT (40/96) cases as irrelevant.[85] Slightly more than half of subjects in this study appear to recognize that the antecedent and consequent both being true leads to a true conditional and they see that a true antecedent with a false consequent leads to a false conditional. Subjects tend not to recognize the cases where the antecedent is false as examples where the conditional is true. Most often subjects consider these cases "irrelevant" to the truth of the conditional, unless they are prompted by being given a conditional with a false antecedent. This has suggested to many researchers that subjects are operating with a faulty truth-table, i.e., a truth table that does not correspond to the material conditional.[86] A meta-analysis of different construction tasks examined by Evans, Newstead and Byrne shows that people almost always recognize

---

[84] Johnson-Laird, P. N., and Tagart, J. (1969).

[85] Ibid.

[86] See Evans, J. St. B. T., Newstead, S. E., and Byrne, R. M. J. (1993).

that a true antecedent and a false consequent leads to a false conditional; however, they only recognize that a true antecedent and true consequent leads to a true conditional approximately half of the time. When subjects were allowed to select their own truth value for the table which included either 'irrelevant' or 'neither' as a possible value, *subjects in none of the studies*[87] selected values for the truth table that corresponded to the standard truth functional view of conditionals. Across the different studies examined by Evans, Newstead and Byrne where subjects were not constrained by only two possible truth values, in *no* case did *any* subject select values that correspond to the truth table for the material conditional. In the studies where subjects were constrained by having only two possible truth values, they selected values that agreed with the truth table for the material conditional approximately half the time.[88]

Even when subjects are given the most simple formulations of conditionals, and even in cases where they are given the constraint of bi-valence, barely half of subjects select a truth table that is consistent with the standard truth functional view. If we wished to set up a situation where subjects would be *forced* to select an interpretation in accordance with the material conditional, regardless of their actual intuitions, we could hardly do better than this— and still half of these subjects do not think that conditional truth is determined by the standard truth functional view. This shows that Jackson's claim that the standard truth functional account of indicative conditionals is

---

[87]The studies examined are: Taplin, J. E. (1971),Evans, J. St. B.T. (1977), Wildman, T. M., and Fletcher, H. J. (1977), Marcus, S. L. and Rips, L. J., (1979), Kern, L. H., Mirels, H. L., Hinshaw, V. G. (1983), Rumain, B. Connell, J., and Braine, M. D. S. (1983), and Markovits, H. (1988).

[88] Evans, J. St. B. T., Newstead, S. E., and Byrne, R. M. J. (1993), p. 52.

the most representative of natural language use and natural language users intuitions about conditionals is clearly false.

To summarize the last section, the results across different studies done by experimental psychologists studying human reasoning show results that challenge the standard truth functional view of conditionals. This is the case even when the subjects are given abstract problems with no conversational implicature to explain away these conflicting intuitions. So, while Jackson's view gives a compelling description of assertibility conditions that solves many of the problems encountered by the view outlined by Quine, his account cannot explain away the wildly counter intuitive implications of the standard truth functional view. At least not if he remains committed to presenting a view that is sensitive to the linguistic/psychological data. This is because his reason for rejecting the intuitions that conflict with his account is based upon a fallacious conclusion drawn from Lewis’s triviality proof. In the following two chapters I will describe a probabilistic interpretation of conditionals and show how it is consistent with data from experimental psychologists studying human reasoning. If our goal in developing an account of conditionals is to model natural language use and remain faithful to the linguistic/psychological evidence, then a probabilistic account of conditionals is the best option.

Chapter Three: Stalnaker's Account of Conditionals

In this chapter I outline Stalnaker's account of conditionals which I argue in chapter four is the probabilistic theory of conditionals that most closely models ordinary language usage. First I give an historical description of the development of possible world semantics, the semantics that Stalnaker's account is based on. Next, I outline Stalnaker's account of conditionals, and describe Jackson and David Lewis's philosophical arguments against this account. I defend Stalnaker's account of conditionals against Lewis and Jackson claiming that their non-unified accounts of conditionals are not warranted or justified in light of the linguistic/psychological data. I further argue that Stalnaker's selection function, while occasionally leading to semantic underdetermination, is more consistent with the use of conditionals in natural language than the alternative offered by Lewis.

Possible Worlds Semantics

C. I. Lewis's attempt to explain away the paradoxes of the material conditional by distinguishing between logically true versus contingently true conditional-like formulations paved the way for Kripke's modal logic which forms the basis of more contemporary interpretations of conditionals. Before Kripke developed his more robust modal logic, Carnap introduced an important development in the form of possible worlds semantics.[89] Carnap's system corrects for the fact that C. I. Lewis's logic did not have an adequate semantic characterization by bringing in a notion of logical truth based on Leibniz's possible worlds. Thanks to Carnap we now have the ability to provide a

[89] Carnap, R., (1946).

semantics for modal logic in the form of possible worlds. Possible worlds are models of sentential logic that allow us to describe possibility and necessity. This gives us the definition: $\Box\phi$ is true iff ϕ is true in all possible worlds and $\Diamond\phi$ is true iff ϕ is true in at least one possible world.

The basic idea behind possible world semantics is that we can imagine a set of worlds where everything that is possible is true in one or more of the possible worlds. So, we have a world where Caesar did not cross the Rubicon, a world where Wittgenstein did not write the *Tractatus,* etc. A possible world interpretation begins with the nonempty set D of all possible objects as well as the set K which are all the possible worlds (one of which is designated the actual world). We assign to each world (w) of the set (K) a set of objects that exist in that world which gives us (V) a valuation which assigns T or F to each sentence letter relative to a particular possible world. With these valuations we can determine the truth value for each sentence in sentential logic. Here are some examples of how valuations work for each connective in sentential logic where ϕ and ψ are wff of sentential logic and α is an arbitrarily chosen world:

$\mathrm{V}_\alpha(\sim\phi)$=T iff $\mathrm{V}_\alpha(\phi)$=F

$\mathrm{V}_\alpha(\phi \rightarrow \psi)$=T iff either $\mathrm{V}_\alpha(\psi)$=T or $\mathrm{V}_\alpha(\phi)$=F

$\mathrm{V}_\alpha(\phi \wedge \psi$=T) iff both $\mathrm{V}_\alpha(\psi)$=T and $\mathrm{V}_\alpha(\phi)$=T

$\mathrm{V}_\alpha(\phi \vee \psi)$=T iff either $\mathrm{V}_\alpha(\psi)$=T or $\mathrm{V}_\alpha(\phi)$=T

$\mathrm{V}_\alpha(\phi \equiv \psi)$=T iff either $\mathrm{V}_\alpha(\psi)$=T and $\mathrm{V}_\alpha(\phi)$=T, or $\mathrm{V}_\alpha(\psi)$=F and $\mathrm{V}_\alpha(\phi)$=F

This allows us now to determine whether a sentence of modal logic is true or not. For example, □(P → Q) is true iff, in our set K of possible worlds, in every world either P is false or Q is true.

One important feature of Kripke's modal logic[90] is that we can define the necessity operator in terms of the possibility operator, for example, □ is equivalent to ~◊~. Another important development from Kripke's system of modal logic is that in "Semantical Analysis of Modal Logic"[91] Kripke introduced a way to describe the relation between possible worlds. This is R, which is a relation between elements of models. It's a binary relation between worlds so that w_1Rw_2 means that w^2 is accessible from w_1. In other worlds, ϕ is possible in w_1 if and only if it is true in some world accessible to w_1.

Kripke's modal logic is both logically important and ubiquitous but its relevance to the paradoxes of material implication is primarily due to the work of Robert Stalnaker and David Lewis. Stalnaker expands on Kripke's modal logic developing a probabilistic modal logic with a probabilistic conditional that best expresses uses of conditionals in natural language. This development was particularly exciting because it seemed to bridge a gap between algebra and probability theory. This made the account extremely attractive to philosophers of logic as the probability calculus is well defined and seemed like it could provide a more solid mathematical grounding for conditionals. Another reason for optimism about Stalnaker's early account of conditionals was that this

---

[90] Kripke, S. A., (1959a), and (1959b).

[91] Kripke, S.A., (1959b).

account is based upon Ramsey's test[92] which struck many as highly intuitive, i.e., it seems to come very close to the meaning of conditional statements in natural language. Specifically, Stalnaker's early account of conditionals appeared to give a solid mathematical account of conditionals and a semantics that accounted for common sense intuitions and avoided many of the paradoxes associated with the standard truth-functional account of conditionals. In this and the following chapter I aim to examine one of the main arguments in support of Stalnaker's hypothesis–- that this account of conditionals does justice to common sense intuitions and can accurately model conditional use in natural language.

Stalnaker's Theory of Conditionals

In "A Theory of Conditionals"[93] Robert Stalnaker develops a theory of conditionals that is unified (both indicative and subjunctive conditionals are analyzed in the same way) and that avoids the paradoxes of the material conditional by analyzing conditional statements using a method motivated by the Ramsey test. The Ramsey test is a test of conditional truth wherein you hypothetically add the antecedent to your stock of beliefs and then see whether the consequent follows. The Ramsey test requires modification since Ramsey's initial remarks only discuss conditionals where the antecedent is false or has an unknown truth value. Stalnaker's version of the Ramsey test takes the basic form:

---

[92] Where we hypothetically add the antecedent to our set of beliefs and see whether the consequent is true.

[93] Stalnaker, R., (1968).

> "First, add the antecedent (hypothetically) to your stock of beliefs; second, make whatever adjustments are required to maintain consistency (without modifying the hypothetical belief in the antecedent); finally, consider whether or not the consequent is true."[94]

Stalnaker argues that both indicative and subjunctive conditionals should be analyzed in the same basic way, so an addition to this test is needed to allow for cases where the antecedent is believed to be true. Stalnaker argues that, in cases where we already believe the antecedent to be true, we can use the same basic analysis where the addition of the antecedent simply requires no revision of our other beliefs. I refer to the result of these adjustments on ones' stock of beliefs, or stock of beliefs in cases where no adjustment is needed as an antecedent-including-scenario.[95] This modified version of the Ramsey test is what describes belief conditions for conditional statements, i.e., this is the method we use to determine whether or not to believe a given conditional.In chapter four and five I will argue that these belief conditions closely resemble the psychological account of mental models.

Stalnaker's modified version of Ramsey's test gives the epistemic conditions for belief in a conditional and must now be connected to possible world semantics in order to give us truth conditions for conditional statements. The introduction of possible worlds semantics results in the following account of conditional truth: "consider a possible world

---

[94] Ibid., p. 44.

[95] In the pages that follow I use the term "antecedent-including-scenario" to mean either: the subject's current set of beliefs if they already believe the antecedent to be true, or a possible set of beliefs that a subject has after adjusting their current beliefs to make the antecedent true. In some cases subjects will construct multiple possible antecedent-including-scenarios if the addition of the antecedent to their stock of beliefs can lead to different, conflicting scenarios.

in which A is true, and which otherwise differs minimally from the actual world. 'If A then B' is true (false) just in case B is true (false) in that possible world".[96] In Stalnaker's description of belief conditions for conditionals, i.e., how we decide whether to believe a given conditional, we are asked how probable the consequent would be given that the antecedent is true. This leads to Stalnaker's hypothesis, p(if A then B) = p(B/A), i.e., the probability of a conditional is equivalent to the probability of the consequent given the antecedent. For example, if we are analyzing the conditional "if I roll an even number on this die, then I will roll a six" there is a 1/2 chance of rolling an even number and a 1/6 chance of rolling a six. In one of the three cases where an even number is rolled, that number will be a six. Thus, this conditional is true in 1/3 of cases and the probability of the conditional is roughly .33. Stalnaker's account of conditionals is presented as an extension of Kripke's modal logic.

For the semantic characterization of his account, Stalnaker uses primarily the modal logic developed by Kripke where every possible world is associated with a probability space.[97] Stalnaker's account relies upon two further additions to Kripke's modal logic. The first is his use of λ as "a member of K [the set of all possible worlds] which is to be understood as the absurd world."[98] This is the world containing all the contradictions and their consequences. By introducing this element, Stalnaker gives his account the ability to interpret a conditional where the antecedent is metaphysically impossible or a contradiction. The other addition to Kripke's modal logic is the selection

---

[96] Ibid., p. 45.

[97] We have in Stalnaker's system σ, a subset of Ω, which selects a single outcome from a set of possible outcomes.

[98] Ibid., p. 45.

function (f).[99] The selection function selects a possible world. This leads to conditional logic, an extension of modal logic, where we analyze the truth of conditionals in the following way: A conditional is true in the actual world when its consequent is true in the selected world where the antecedent is true. Modal logic provides the apparatus for speaking about possible worlds and the selection function provides the means to speak about particular possible worlds.

Stalnaker's selection function adds two constraints to Kripke's logic that lead to an ordering of possible worlds based on their similarity to the actual world and two more conditions that place further restrictions on this ordering. I am using Stalnaker's own formulations here as provided in "A Semantic Analysis of Conditional Logic"[100] where:

σ is an assignment of values to individual variables,

A and B are wffs,

α and β are individual possible worlds,

f is a function which assigns to each wff, each sequence σ, and each world that is a member of K, a member f(A, α, σ), and

I is an interpretation which gives us a valuation and an f function.

The restrictions are as follows:

$$I_{f(A, \alpha, \sigma)}(A, \sigma) = T$$

The first restriction ensures that the possible world selected is the world where the antecedent is true.

If $I_{\alpha}(A, \sigma) = T$, then $f(A, \alpha, \sigma) = \alpha$

---

[99] Ibid., p. 45.

[100] Stalnaker, R., and Thomason, R. (1970).

The second restriction ensures that if the antecedent is true in the actual world, then the actual world must be selected. This restriction is further elaborated by Stalnaker's claim that the selected world must differ minimally from the actual world. This means that any differences between the selected world and the actual world must be restricted to those differences that are required in order to adopt the antecedent as true. Stalnaker claims that, "one must choose [the possible world] that does the least violence to the correct description and explanation of the actual world".[101]

$f(A, \alpha, \sigma) = \lambda$ only if there is no $\beta \in K$ such that $\alpha R \beta$ and $I_{\alpha}(A, \sigma) = T$

This restriction placed on the selection of worlds ensures that the absurd world (λ) be selected only when there is no other possible world in which the antecedent is true. This condition is necessary because if the absurd world could be selected at any time, then any conditional statement could be made true.

If $I_{f(A, \alpha, \sigma)}(B,\tau)=I_{f(B, \alpha, \tau)}(A,\sigma) = T$, then $f(A, \alpha, \sigma) = f(B, \alpha, \tau)$

If w_1 is chosen over w_2 in a context where both are eligible, then w_1 must always be chosen over w_2. This last condition ensures that the ordering of possible worlds is consistent in that, if a selection makes w_1 prior to w_2 where both are eligible, then every other selection must also make w_1 prior to w_2.

Stalnaker introduces the corner (>) to represent his account of conditionals. The corner can be understood in the following way: $(P \supset Q)$ is entailed by $(P > Q)$ which is entailed by $\Box(P \supset Q)$. Stalnaker and Thomason define the truth conditions for the corner

---

[101] Stalnaker, R., (1968), p. 46.

as, $I_{\alpha}(A > B, \sigma)=T$ iff $I_{f(A, \alpha, \sigma)}(B, \sigma)=T$.[102] In an earlier paper, Stalnaker defines the corner thusly:

> "A > B is true in α if B is true in f(A, α)
>
> A > B is false in α if B is false in f(A, α)"[103]

In other words, P > Q is true in a given world iff Q is true in the selected world where P is true.

There are a number of logical implications of this account that Stalnaker argues more closely align with our intuitions about conditionals. Firstly, Stalnaker's account of conditionals avoids the paradoxes of material implication as outlined in the first chapter so that neither $(\sim\phi > (\phi > \psi))$ nor $(\phi > (\psi > \phi))$ are logically true. Here is a counter-example for $(\sim\phi > (\phi > \psi))$ where w_1 is the actual world:

W: w_1, w_2

$I_{w2}(\phi > \psi, \sigma)=F$ since $I_{f(\phi, w2, \sigma)}(\psi, \sigma)=F$

$I_{w1}[\sim\phi > (\phi > \psi), \sigma]=F$ since $I_{f(\sim\phi, w1, \sigma)}(\phi > \psi, \sigma)=F$

$V_{w1}(\phi)=F$, $(\psi)=T$

$V_{w2}(\phi)=T$, $(\psi)=F$

Here is a counter-example for $(\phi > (\psi > \phi))$ where w_1 is the actual world:

W: w_1, w_2

$I_{w2}(\psi > \phi, \sigma)=F$ since $I_{f(\psi, w2, \sigma)}(\phi, \sigma)=F$

$I_{w1}[\phi > (\psi > \phi), \sigma]=F$ since $I_{f(\phi, w1, \sigma)}(\psi > \phi, \sigma)=F$

---

[102] Stalnaker, R., and Thomason, R. (1970), p. 28.

[103] Ibid., p. 46.

$V_{w1}(\phi)$=T, (ψ)=F

$V_{w2}(\phi)$=F, (ψ)=T

The corner is also non-transitive, meaning that it avoids many of the other counter-intuitive implications of the standard truth functional conditional. For example, $(\phi > \psi)$ and $(\psi > \zeta)$ do not imply $(\phi > \zeta)$. Here is a counter-example for $((\phi > \psi)$ ^ $(\psi > \zeta)) > (\phi > \zeta)$ where w_1 is the actual world:

W: w_1, w_2, w_3

$I_{w1}(\phi > \zeta, \sigma)$=F since $I_{f(\phi, w1, \sigma)}(\zeta, \sigma)$=F

$I_{w3}(\phi > \psi, \sigma)$=T since $I_{f(\phi, w3, \sigma)}(\psi, \sigma)$=T

$I_{w2}(\psi > \zeta, \sigma)$=T since $I_{f(\psi, w2, \sigma)}(\zeta, \sigma)$=T

$I_{w1}[((\phi > \psi)$ ^ $(\psi > \zeta)) > (\phi > \zeta), \sigma]$=F since $I_{f[(\phi > \psi) \wedge (\psi > \zeta), w1, \sigma]}(\phi > \zeta, \sigma)$=F

$V_{w1}(\phi)$=F, (ψ)= F, (ζ)=F

$V_{w2}(\phi)$=F, (ψ)=T, (ζ)=T

$V_{w3}(\phi)$=T, (ψ)= T, (ζ)=F

Stalnaker relies on an example about Hoover to argue that conditionals in natural language are not necessarily transitive. In this example we have the conditional premises: “If Hoover had been born a Russian, then he would be a communist” and “If Hoover were a communist, then he would be a traitor.” Stalnaker argues that these premises ordinarily do not lead one to conclude “If Hoover had been born a Russian, then he would be a traitor.”

Another common inference blocked by Stalnaker's account of conditionals is strengthening of the antecedent. On this account $(\phi > \psi)$ does not imply $((\phi \wedge \zeta) > \psi)$. Here is a counter-example for $(\phi > \psi) > ((\phi \wedge \zeta) > \psi)$ where w_1 is the actual world:

W: w_1, w_2

$I_{w1}(\phi > \psi, \sigma)$=T since $I_{f(\phi, w1, \sigma)}(\psi, \sigma)$=T

$I_{w2}[(\phi \wedge \zeta) > \psi, \sigma]$=F since $I_{f(\phi \wedge \zeta, w2, \sigma)}(\psi, \sigma)$=F

$I_{w1}[(\phi > \psi) > ((\phi \wedge \zeta) > \psi), \sigma]$=F since $I_{f(\phi > \psi, w1, \sigma)}[(\phi \wedge \zeta) > \psi, \sigma]$=F

$V_{w1}(\phi)$=T, (ψ)=T, (ζ)=F

$V_{w2}(\phi)$=T, (ψ)=F, (ζ)=T

To defend this result Stalnaker relies on the oft-used example of the wet match. It may be true that, "If I strike this match, then it will light" without it being true that, "If I soak this match in water and strike the match, then it will light."

Contraposition is also invalid; so $(\phi > \psi)$ does not entail $(\sim\psi > \sim\varphi)$. Here is a counter-example for $(\phi > \psi) > (\sim\psi > \sim\phi)$ where w_1 is the actual world:

W: w_1, w_2

$I_{w1}(\phi > \psi, \sigma)$=T since $I_{f(\phi, w1, \sigma)}(\psi, \sigma)$=T

$I_{w2}(\sim\psi > \sim\phi, \sigma)$=F since $I_{f(\sim\psi, w2, \sigma)}(\sim\phi, \sigma)$=F

$I_{w1}[(\phi > \psi) > (\sim\psi > \sim\phi), \sigma]$=F since $I_{f(\phi > \psi, w1, \sigma)}(\sim\psi > \sim\phi, \sigma)$=F

$V_{w1}(\phi)$=T, (ψ)=T

$V_{w2}(\phi)$=T, (ψ)=F

Stalnaker defends this by arguing that a person may believe that the following statement is true— "If the United States halts the bombing, then North Vietnam will not

negotiate”— if she believes that North Vietnam wants a complete withdrawal of American forces. She would in this case not believe that the contrapositive— “If North Vietnam negotiates, then the United States will not halt the bombing”— is true.

Stalnaker argues that his account of conditionals models natural language use more reliably than the standard truth-functional account primarily because it accommodates a broader range of situations and contexts. He claims that the most beneficial aspect of this type of analysis is that, when a connection between the antecedent and consequent is relevant, this account includes it, but when a connection is not relevant, it is not a necessary condition. This is because if the person analyzing the conditional believes that there is a connection between the antecedent and the consequent, when they add the antecedent into their belief set they will hold the consequent to be true. For example, if I believe that there is some relevant connection between the propositions “the Chinese enter the Vietnam war” and “the American military resort to nuclear weapons” such that the first guarantees the second, then a modified Ramsey’s test will lead to a true value for the conditional, “If the Chinese enter the Vietnam conflict, the United States will use nuclear weapons.” Whereas if I don’t believe that there is any connection between the Chinese entering the war and the American military resorting to nuclear weapons, then I will believe there is no change in the state of affairs relevant to the consequent and say that the conditional is false.

The common criticisms of Stalnaker do not tend to focus on these inference forms that are no longer valid in his account. Typically, those who take issue with Stalnaker’s account do so because they feel that a probabilistic conditional is too imprecise to be of use in scientific contexts or because they think that Lewis's triviality

results show the impossibility of a probabilistic conditional.[104] Specifically, Stalnaker's account leaves the epistemic conditions for some conditionals as vague or undefined. I will examine and respond to these criticisms below.

Philosophical Criticisms of Stalnaker

Most modern accounts of conditionals embrace possible world semantics to one degree or another as it is generally agreed that people use conditionals often to talk about possibilities. As Jackson says,

> "conditionals are hypothetical not categorical. They are, in some sense, about possibilia which include their antecedents. It is thus a mistake in principle to hold that the meaning of hook exhausts the meaning of any genuinely conditional construction."[105]

Stemming from this agreement, we get a number of different views about how conditionals should be analyzed with some authors trying to retain truth functionality for indicative conditionals and some embracing a more unified way of analyzing conditionals. There are differences also between the ways in which David Lewis and Stalnaker analyze subjunctive conditionals with Lewis claiming that a comparative similarity relation is how we select the closest possible world and Stalnaker claiming that that a selection function should be used. As seen in the previous chapter, Jackson

---

[104] These criticisms of Stalnaker's account have been made by Frank Jackson ((1980) Conditionals and Possibilia. *Proceedings of the Aristotelian Society*, New Series, Vol. 81, p.132.) and David Lewis((1973), Counterfactuals and Comparative Possibility, *Journal of Philosophical Logic*, Vol. 2, No. 2. Reprinted in Harper et al. (eds.) *Ifs,* (1981) D. Riedel.).

[105] Jackson, F., (1980), p. 125.

agrees with Lewis that indicative conditionals should be analyzed in a standard truth functional way and gives an explanation of conditionals that relies upon the notion that probability conditions affect the assertibility of a conditional but not its truth conditions. In this section I outline and respond to the main criticisms that David Lewis and Jackson level against Stalnaker's account of conditionals.

The dispute between Stalnaker and Jackson/Lewis is primarily over whether there is a significant difference between indicatives and subjunctives, and whether we should analyze them in roughly the same or in different ways. Grammatically, the indicative conditional is used to state matters of fact and the subjunctive conditional is used when the antecedent is known to be false or uncertain. It is not always clear in discussions by philosophers whether conditionals with antecedent truth values that are unknown, e.g., "If it starts to rain, then the ballgame will be cancelled" which can be grammatically expressed in the indicative mood are nonetheless logically subjunctives or not. Jackson holds that only counterfactuals are subjunctives and argues that there are significant differences in how natural language users use indicatives and subjunctives, which justifies a different approach for each. The primary motivation behind both Jackson's and Lewis's acceptance of a non-unified account of conditionals however, is that they want to retain the simplicity of the standard truth functional account of conditionals for indicatives.

Jackson's Criticisms of Stalnaker

In Stalnaker's theory of conditionals both indicative and subjunctive conditionals are analyzed in much the same way (the difference being simply that there is a further

constraint on the selection function in the case of indicatives). The fact that Stalnaker has a unified account is the primary criticism levelled against this account by Jackson who, as seen in the previous chapter, argues that only counterfactuals should be analyzed in a probabilistic way. According to Jackson, the indicative conditional construction in Stalnaker's account "acts as a signal" that we are taking for granted a certain context. Jackson argues that,

> "[Stalnaker's] suggestion is that the subjunctive-counterfactual construction acts as a signal to range further afield; the indicative construction acts as a signal to keep within the bounds of what is presupposed in the context."[106]

This characterization of Stalnaker's account is appropriate since in Stalnaker's account the difference between the selection function for indicative and subjunctive conditionals is that, for the former, we outline a context set of possible worlds that take into account the background information and the selected world must be selected from this context set. One example that Jackson believes shows the wrong-headedness of Stalnaker's unified account is the Oswald-Kennedy example given in "Conditionals and Possibilia."[107] In this example we have two conditionals–- "If Oswald didn't kill Kennedy, then someone else did" and "If Oswald hadn't killed Kennedy, then someone else would have"–- the first of which is in the indicative form and the second in the subjunctive form. According to Jackson's view, the first conditional is false since the antecedent is false and the second conditional is potentially true depending on the closest possible world we examine. Jackson argues that on Stalnaker's account we are to take for

[106] Ibid., p. 127.

[107] Ibid.

granted that someone killed Kennedy (because this is part of the context and Stalnaker's theory requires that we choose possible worlds that do the least possible violence to the description of the actual world) and so we should only look at possible worlds that maintain this. In other worlds, Jackson argues that Stalnaker's account of conditionals requires that, when evaluating the indicative conditional, we only look at possible worlds where Kennedy is killed. However, whether Kennedy is killed is the point of contention for these conditionals. Jackson argues that evaluating this indicative conditional using Stalnaker's selection function requires taking for granted the very fact that the conditional is meant to evaluate. According to Jackson, Stalnaker is correct in assuming that the indicative construction requires further restrictions; however, he argues that, "in Stalnaker's case the additional restriction is agreement with what is being taken for granted in the context; ... I think that [he] should have been more restrictive still."[108] What Jackson means here by 'more restrictive' is that indicatives should not be subject to a possible worlds analysis at all.

This counter-example provided by Jackson is only effective if we assume that the fact that someone killed Kennedy is the most relevant feature of the context of the actual world. There are other relevant contextual considerations such as Kennedy's policies, the content of Kennedy's speeches, Kennedy's social and professional relationships, etc., that are taken into consideration when we choose the world most like the actual world excepting that Oswald did not kill Kennedy. These seem at least as relevant to a consideration of, "If Oswald had not killed Kennedy, someone else would have," than the murder of Kennedy itself. In other words, there are many relevant

---

[108] Ibid., p. 128.

contextual considerations that must be taken for granted in this context, but it is not clear that the murder of Kennedy is one of them.

Ellis criticizes Jackson's account of subjunctive and indicative conditionals, arguing that the account offered by himself and Stalnaker offers, “a unified and comprehensive account, where such an account seems to be required; and no drastic reconstrual of the empirical data (of logical intuition) is required to make them fit the theory.”[109] I agree that there is no reason to assume that the analyses for indicatives and subjunctives should be radically different and that a vastly different analysis is not justified. Essentially, we agree that the possible worlds account is appropriate for counterfactual conditionals, we have no reason to assume and no ability to justify the claim that people use indicatives in relevantly different ways, therefore, we should analyze indicatives using the possible worlds approach as well.

Lewis’s Criticism of Stalnaker

David Lewis agrees with Jackson that indicative conditionals and subjunctive conditionals should have different analyses and he also takes issue with Stalnaker's choice of a selection function, i.e., how possible worlds are selected in Stalnaker’s account. Lewis claims that the problem with Stalnaker's selection function is that it entails what he dubs the uniqueness assumption and the limit assumption, both of which he argues are unrealistic or impractical. Lewis's first problem with Stalnaker's analysis is that it “depends upon a thoroughly implausible assumption: that there will never be more than one closest A-world”.[110] This he calls the uniqueness assumption.

[109] Ellis, B. (1984), p. 50.

[110] Lewis, D., (1973).

According to Stalnaker, the uniqueness assumption is the assumption that, “no distinct possible worlds are ever equally similar to any given possible world.”[111] In order to demonstrate the unrealistic nature of this assumption, Lewis uses an example with Bizet and Verdi, i.e., "If Bizet and Verdi are compatriots, then Bizet is Italian; or If Bizet and Verdi are compatriots, then Verdi is French”. He claims that an A-world where both Bizet and Verdi are French is equally as close as one where Bizet and Verdi are both Italian and that there is no plausible way to distinguish between both of these worlds. Lewis uses this example to claim that sometimes there is no single closest A-world.

Another problem with Stalnaker's account, according to Lewis, is that it leads to what Lewis calls the limit assumption. This is the assumption that, “given that some A-world is accessible from *i*, we no longer assume that there must be exactly one closest A-world to *i*; but we still assume that there must be at least one.” Lewis argues that there could be possible worlds that become closer and closer to the actual world without end. As an example Lewis uses the case where we are looking for a world where someone is greater than seven feet tall which differs minimally from the actual world where this person is seven feet tall. Lewis claims that this is not as easy as being able to claim that the closest possible world is where the person is 7+X feet tall. He claims that there is nothing preventing there from being a 7+X/2 or a 7+X/4, or a 7+X/n.

The main motivation Lewis gives for the rejection of Stalnaker's analysis is that it “is founded on comparative similarity... 'closeness' of worlds” and that “that comparative similarity is hopelessly imprecise unless some definite respect of comparison has been specified.”[112] Lewis claims that it is possible to try to reconcile this problem by claiming

---

[111] Stalnaker, R., (1981b).

[112] Lewis, D. (1973), p. 59.

that counterfactuals themselves are also imprecise and therefore, two imprecise concepts can somehow be “rigidly fastened to one another” (which is similar to how Stalnaker actually attempts to solve this problem) to leave us with something that is precise (namely, the connection between them). In defence of Stalnaker's position stands the fact that we frequently judge similarity without a specific criterion (or at least not an explicit one). Lewis claims, however, “such imprecision we can live with. Still, I grant that a counterfactual based on comparative similarity has no place in the exact language of science.”[113]

Stalnaker's Defence

Stalnaker defends his selection function by arguing that some vagueness in the selection of possible worlds is actually one of the benefits of his theory of conditionals. He claims that a general theory of vagueness can account for the fact that the abstract semantic theory leads to determinate results while the use of the selection function in particular cases sometimes results in indeterminacy. The idea is, when there are several possible worlds and it isn’t clearly the case that one is the closest possible world (as with Lewis’s “person greater than seven feet tall” example) we simply decide which worlds fall into the clearly closest, clearly not-closest, and penumbral categories (with super-valuations the categories are of course true, false, and penumbral). If all of the worlds in the clearly closest worlds yield a true value for the conditional, then the conditional is true. For example, if we have the conditional, “If John is taller than seven feet, he can reach the eight foot high shelf” we can put in the “clearly closest” all of the

[113] Ibid., p. 60.

worlds where John is taller than seven feet but shorter than eight feet for example. It doesn't matter whether John is 7 feet 1 inches or seven feet 2 inches, etc., because in all of these cases the counterfactual conditional is true. In cases where it matters whether John is 7'1" or 7'2" the conditional will be false unless it specifies this. Stalnaker argues that the selection function will ignore irrelevant similarities, solving Lewis's problem in some cases, and further, that it solves the problem in other cases by making the selection function undefined in cases where "every millimetre matters." This is because, Stalnaker claims,

> "if every millimetre matters then it is just inappropriate to use the antecedent 'if the line were more than an inch long'. This would, in such a context, be like using the definite description 'the shortest line longer than one inch'. The selection function would be undefined for antecedents in such contexts."[114]

Stalnaker claims that the addition of super-valuations to his selection function is not an *ad hoc* move because it is necessary to give an account of semantic indeterminacy in order "to account for pervasive semantic underdetermination in natural language."[115] We frequently use terms that fail to meet Lewis's criteria for precision, we nonetheless interpret and use these sentences.

Stalnaker claims that his uniqueness assumption is not a real problem because the value for the Bizet and Verdi conditionals remains the same in both evaluations. So, in a situation where we are unsure whether Bizet is more likely to be Italian or Verdi more likely to be French, the values for "If, Bizet and Verdi are compatriots, then Bizet and Verdi are French" and "If Bizet and Verdi are compatriots, then Bizet and Verdi are

---

[114] Stalnaker, R., (1981b), p. 97.

[115] Ibid., p. 90.

Italian" conditionals are the same. Both of these conditionals are indeterminate on Stalnaker's analysis, as there is no clear closest possible world and an attempt to super-valuate the worlds leads to both true and false values. Stalnaker argues that this is a more intuitive interpretation of these types of conditionals than an account that applies a precise value to counterfactuals that are clearly indeterminate. Lewis removes this indeterminacy since he is a realist about possible worlds. So, for example, there is some fact of the matter about which Bizet and Verdi world is the closest, we just don't have epistemic access to other worlds so we don't know which conditional is true. I take this to be an unsatisfactory solution in the sense that it leaves us with epistemic indeterminacy in these cases and creates metaphysical problems surrounding the existence of possible worlds.

In Lewis's analysis a similarity relation is used instead of a selection function. His final analysis is: "'A → C' is true at *i* iff some (accessible) AC-world is closer to *i* than any ~AC-world, if there are any (accessible) A-worlds".[116] The problem with Lewis's analysis is that his rejection of the limit assumption, combined with his account of might and would conditionals, leads to the result that "there is no real number x such that my height might be x, if I were over seven feet tall." This is because "for every positive *x* the counterfactual 'if I were over seven feet tall, then I would be under 7 + x feet tall' comes out true."[117] Therefore, "for every real number x, my height would not equal x feet if I were over seven feet tall." Due to these obviously absurd results, Stalnaker claims that

---

[116] Lewis, D., (1973), p. 63.

[117] Harper, W., (1981), pp. 3 - 38

his account based on comparative similarity makes just as much sense as the idea that the basis for selection is a similarity relation.

It would be difficult for Lewis to argue against this point since his own claim that possible worlds must have a certain minimal difference from one another seems to imply that there can be a limit to how close to the actual world possible worlds can be. Lewis argues, “differences never come singly, but in infinite multitudes”[118] which he attempts to justify by using the example of Caesar crossing the Rubicon. Lewis contends that if Caesar had not crossed the Rubicon, then the rest of the world cannot possibly be exactly the same as it is in this world where Caesar did cross the Rubicon—for there must be some differences in Caesar's character or the external world to account for these changes. That is, we cannot have a possible world that is just like the actual world except for one feature, or one that, “does not differ gratuitously from ours”. This implies that every world that differs from the actual world must differ more than some minimal limit, or that a possible world cannot differ from the actual world by only one characteristic. If every world “differs gratuitously” from the actual world then it seems impossible to have worlds such that we cannot distinguish which is the closest possible world. Lewis claims that in examples of the form “if Bobby is taller than 1 m, then he can ride on the roller coaster,” we cannot distinguish the closest possible world because if Bobby is 1 m + x there is no reason to assume that there will not be a world where Bobby is 1 m + x/2 or 1 m + x/4, and so on. With the requirement that possible worlds must differ minimally from one another, it is incoherent to say that we could not distinguish between these worlds because the difference in the size of Bobby must not

---

[118] Lewis, D., (1973), p. 59.

be the only difference and therefore we would still be able to conceive of a closest possible world. Therefore, Stalnaker's limit assumption is intuitively warranted (even if we use Lewis's intuitions). Further Stalnaker's selection function avoids the absurd results of the similarity relation for conditionals like the "If John's height is over seven feet tall…" variety.

Jackson's first criticism of Stalnaker's account of conditionals is unsuccessful as he is unable to give compelling reasons to support different analyses for indicative and subjunctive conditionals. For no other logical terms do we take a shift in grammatical mood to indicate that we should have two separate analyses. Jackson's second argument— that Stalnaker's judgement of comparative similarity conflicts with how we interpret the, "If JFK had not been shot, …" type of conditionals willfully misinterprets and oversimplifies the considerations of what is taken for granted in a particular context. Lewis's arguments against the limit assumption and the uniqueness assumption also fail to show that Stalnaker's account of conditionals is flawed. The use of super-valuations corrects for a situations where there is no unique closest possible world and ignoring irrelevant similarities or making the similarity function undefined in cases where every millimetre matters solves the problem Lewis has with the limit assumption.

The real issue in the disputes between Lewis, Jackson and Stalnaker is whether it is more important to maintain absolute precision in our theory of conditionals or if it is more important to correctly account for the intuitions of natural language users. Jackson's primary reason for rejecting Stalnaker's probabilistic account of conditionals is that he wishes to retain the simplicity of the material conditional. Obviously developing a logical account of conditional use involves a degree of simplification.

However, if our account of a logical term conflicts with the overwhelming majority of usages of that logical term (as the previous chapter showed is the case for the material conditional), then we have oversimplified— at the cost of descriptive accuracy. Lewis's primary reason for rejecting Stalnaker's account is that Stalnaker's account entails a certain degree of semantic indeterminacy. Stalnaker correctly asserts, as chapter 4 will show, that his analysis is in greater accord with the intuitions of natural language speakers than Lewis's view. Lewis agrees that Stalnaker's view captures ordinary language intuitions better than his view does and even claims that his own view "does sound like a contradiction."[119] He nonetheless rejects Stalnaker's account claiming that the cost of this view, in terms of loss of precision, is too great. Lewis seems to think that accepting even this small degree of indeterminacy will render conditionals useless for the precise applications of science.

There are two related problems with Lewis's claim that his precise conditional account is better for scientific applications. Firstly, this claim is a diversion since Lewis gives no specific examples of contexts where such precision is required. If Lewis believes that conditionals as used in scientific contexts require a different analysis than ordinary language conditionals he needs to provide evidence of this divergent usage. As it is, his argument amounts to nothing more than an unjustified association between science and logical precision. Secondly, as Alan Hájek argues in "Triviality Pursuit,"[120] the conditional as used in certain scientific contexts (he specifically mentions quantum mechanics) often involves antecedent values that are vague or absent. The important

---

[119] Stalnaker, R., (1981b), p. 92.

[120] Hájek, A., (2011).

question is whether Stalnaker's conditional truly does conform to ordinary language intuitions. If it can be shown that Stalnaker's account of conditionals is consistent with broader use of the conditional in natural language than the alternative views, such an increase in descriptive accuracy is worth accepting a degree of semantic indeterminacy. Particularly if, as Stalnaker claims, the descriptive accuracy of his account is partially due to this semantic indeterminacy. In the following chapter I aim to show that data from experimental psychologists studying conditional reasoning supports the claim that the conditional used in ordinary language is a unified probabilistic one much like the account offered by Stalnaker.

Chapter Four: Outline of the Experimental Support for a Probabilistic Account

In this chapter I outline the experimental results that support the claim that the most descriptively accurate psychological account of conditionals is based on a probabilistic conditional. My purpose in the following two chapters is to get a better picture of the phenomena that the logical theory is attempting to model— natural language use of conditionals— by including assessments of conditional reasoning by naive reasoners as shown through studies done by experimental psychologists. This experimental data shows that most uses of conditionals in natural language are interpreted probabilistically which supports the claim that the logical account that is best able to describe conditional use in natural language will also be probabilistic. I outline the results on psychological tasks that an account based on the material conditional cannot accurately describe and argue that a probabilistic theory of conditionals such as those provided by Oaksford and Chater or Schroyens and Schaeken are able to describe these uses. I argue that the most descriptively accurate psychological theory (the alternative mental models theory) shares many essential features and a basic structure with the belief conditions for conditionals outlined by Stalnaker. This provides support for the claim that the logical theory of conditionals provided by Stalnaker, or an alternate probabilistic account that shares many of the same features, is the best able to capture conditional reasoning by naive reasoners.

These experimental results are taken from a number of different tasks used in the study of human reasoning by experimental psychologists. In the first section of this chapter I examine the results that are called ‘biases’ in the psychological literature.

These biases are results that deviate from the standard truth functional interpretation of conditionals. Evans, Newstead and Byrne state, “deviations from logical principles are however often *systematic*.”[121] By “logical principles” they mean an interpretation of conditionals consistent with the material conditional and by “systematic” they mean responses that are consistent across different studies (and occasionally different types of task). So, biases are the systematic ways that subject responses fail to align with the material conditional. The fact that these biases are widespread has led many psychologists to reexamine the earlier assumption that these responses are normative errors. Most influentially, Oaksford and Chater[122] argue that these biases can be explained as rational strategies if we take subjects to be operating with a probabilistic rather than a material conditional.

The first section of this chapter is a study of some of the recognized biases in the psychological literature and an explanation of how a probabilistic psychological account of conditionals can describe these biases as rational reasoning strategies. I present this as evidence that a probabilistic psychological account of conditionals is the most descriptively accurate account since the many convoluted reasons given for subjects' performance errors that were needed to account for subject responses under previous psychological theories of conditionals are unnecessary if we take subjects instead to be using a probabilistic conditional. While the claim that subjects are relying upon a material conditional requires the majority of subject responses to be considered performance errors or heuristics, the claim that subjects are relying upon a probabilistic conditional has the ability to describe almost all responses as rational belief testing

[121]Evans, J. St. B. T., Newstead, S. E., Byrne, R. M. J., (1993), p. 270.

[122] Oaksford, M., and Chater, N., (2007).

methods. This provides support for the claim that the conditional in typical natural language use is probabilistic. Firstly, as this description of subject responses is simpler and more consistent than the alternative. Secondly, this theory is able to posit possible reasons for particular biases rather than simply claiming that, for example, the inclusion/ exclusion of a negation renders people incapable of reasoning.

The second section of this chapter will be a discussion of probabilistic accounts of conditional reasoning proposed by experimental psychologists. I will show that these probabilistic accounts predict results on tasks with conditionals with far greater accuracy than the non-probabilistic accounts. Even allowing for the ad hoc assumptions required to explain the majority of subject responses as performance errors, the assumption that subjects in reasoning tasks are operating with a material conditional in mind still accounts for less of the data. The fact that a probabilistic account of conditionals is both more simple and better at predicting subject responses I put forward as evidence for the conclusion that natural language users are primarily operating with a probabilistic conditional. Finally, I argue that Stalnaker's account of conditionals, or a probabilistic account that shares many of the same features, are the only logical accounts of conditionals with the essential features required to be compatible with such a psychological theory.

Effects and Biases

In the psychological literature, the responses that deviate from the standard truth functional account of conditionals are referred to either as effects or biases. I will be examining many of these recognized biases in order to argue that the deviation from the

standard truth functional account is explained by the hypothesis that these subjects are interpreting the conditional according to a different— probabilistic— conditional. First, a note is needed on what psychologists mean when they call a set of responses a bias. Evans et al. define a bias by stating that a bias is “a systematic error, relative to the normative model provided by formal logic.”[123] They also define bias as, “systematic attention to a logically irrelevant feature of the task”[124] by which they mean that subjects are attending to characteristics that are not relevant from a standard truth functional view. As will be seen, these features are logically relevant if we examine the responses with a different account of conditionals in mind. Specifically, a probabilistic account of conditionals can convincingly account for: the variation seen in subject responses, the suppression effect, confirmation bias, matching bias and negative conclusion bias.

Variation in Subject Responses

One of the effects that challenges the assumption that subjects are reasoning with a standard truth-functional account of conditionals is the variation among subject responses. The variation effect shows that the background knowledge that subjects have about particular topics will impact their responses on tasks. Subject responses are also effected by things such as their ability to imagine alternative possibilities. On a standard truth-functional view, this extra knowledge and ability to construct alternatives should be irrelevant to subject responses. For example, subjects are evaluating conditional statements such as, “If there is a dog in the house, then the dog belongs to Mary” which they are told to take as true, and told that the unconditional statement

[123]Evans, J. St. B. T., Newstead, S. E., Byrne, R. M. J. (1993), p. 44.

[124] Ibid., p. 49.

“there is a dog in the house” is also to be taken as true. Subject knowledge about Mary’s dog or their ability to envision situations where the dog in the house does not belong to Mary should be irrelevant if subjects are evaluating the conditional according to the standard truth functional account because the material conditional makes the truth of the conditional certain. Only if subjects take the truth of the conditional to be probabilistic would such considerations be relevant to what conclusions can be drawn from the two statements above.

The mental models theory, developed initially by Johnson-Laird,[125] proposes that subjects reason by constructing models that represent a world in which the antecedent is true and then use this model to generate their conclusions. Mental models theorists claim that people do not reason using formal rules but rather construct a mental model of a world in which the premises are true and then “look” to see which conclusions follow. According to mental models theory, variation among subject responses is explained by the fact that some people are better able to construct and flesh out models, either because of a natural ability to keep more models in mind, or because of more personal experience with the subject matter. A probabilistic analysis of conditionals can explain this deviation in a similar way. The antecedent-including-scenario that the subject would construct will contain more detail if the subject is familiar with the subject matter and certain subjects may have a natural ability to conceive of more of the antecedent-including-scenarios than others. Both of these hypotheses are testable— if these theories are correct, then a subject would reason better when they are able to imagine more possibilities (more models or more examples of antecedent-

---

[125]Johnson-Laird, P. N. (1983).

including-scenarios). Two studies appear to support this hypothesis; In 1984 Markovits argued that a subject's ability to imagine more alternate possibilities would affect her performance.[126] In this study subjects were given three conditional reasoning tasks and an additional question with the following story designed to test their ability to come up with alternate possibilities:

> "When David has homework to do, he gets into a bad mood. I saw David after school today and he was in a bad mood. Can you imagine what could have put David in a bad mood?"

Markovits found that subjects who were able to imagine more possibilities (they listed more examples of things that could have put David in a bad mood) committed fewer fallacies on the tests of conditional reasoning and tended to interpret the conditional as material rather than as a biconditional, subjects who could imagine fewer possibilities for the cause of David's mood made more fallacy errors and they tended to interpret the sentence as biconditional rather than material conditional.[127]

A study done by Cummins in 1991 also lends support to a probabilistic account of conditionals. In this study subjects are asked for "disabling conditions" which are additional requirements for the consequent that are not contained in the conditional itself. An example of the disabling conditions generation task given in this experiment is:

> Rule: If Joyce eats candy often, then she will have cavities.
>
> Fact: Joyce eats candy often, but she does not have cavities.

---

[126] Markovits, H. (1984).

[127] Evans, J. St. B. T., Newstead, S. E., and Byrne, R. M. J., (1993), p. 57.

> Please write down as many circumstances as you can that could make this situation possible.

The conditionals used in the study were then ranked according to how many disabling conditions subjects could come up with. Then these conditionals were embedded in valid and invalid argument forms of *modus ponens*, *modus tollens*, affirmation of the consequent, and denial of the antecedent. Cummins found that subjects make fewer valid inferences on conditionals for which they can think of many of these "disabling conditions."[128] This suggests that subjects do not take the conditional premise as given, but rather analyze its degree of truth by testing through models (or possible scenarios) whether the truth of the antecedent guarantees the truth of the consequent.

Confirmation Bias

One of the effects seen in the Wason selection task[129] (and also in the truth table evaluation task[130]) is called confirmation bias. In the Wason selection task, less than ten percent of people[131] opt to turn over the *p* and *not-q* cards (when the conditional they are given is "if p, then q") which have the potential to falsify the conditional (given the standard truth-functional account of conditionals). Instead subjects more frequently opt

---

[128] Cummins, D. D., Lubart, T., Alksnis, O., and Rist, R. (1991).

[129] In the Wason selection task subjects are given three cards, one with a P, one with a Q, one with a 3 and one with a 7 on them, and asked which cards they need to flip over to evaluate the conditional "If there is a P on one side of the card, then there is a 3 on the other side." See page 58-59 for a more detailed description.

[130] In this task subjects are given a conditional statement and asked to fill in the truth values for the unconditional statements making it up or are given the values for the antecedent and consequent and are asked to fill in the values for the conditional.

[131] Wason, P. C., and Johnson-Laird, P. N. (1972).

to turn over the *q* card and the *p* card. Wason and Johnson-Laird conclude, “typical results were: p and q cards, 46 percent; p card only, 33 percent; p,q, and not-q cards, 7 percent; and p and not-q cards, 4 percent.”[132] These results led to a description of confirmation bias (in Wason's early papers it is called verification bias).[133] In an early study done by Wason, subjects were found to select most often the true consequent and true antecedent and rarely to select the falsifying case of false consequent. This led Wason to propose the theory that subjects are operating with a “defective” truth table, i.e., a truth table that conflicts with the standard truth functional account.[134] Another common interpretation of this systematic bias is that people tend to select options which will confirm, rather than falsify, the conditional they are given.

According to Oaksford and Chater, people interpret conditionals given in the Wason selection task probabilistically which explains why subjects select the true antecedent and true consequent cards. Their argument is based upon the assumption that, in real world reasoning, opting to turn over the true antecedent and true consequent cards is associated with the highest information gain. In other words, subjects are predisposed to analyze the conditional given in the Wason selection task probabilistically and thus the rational selections would be the cards that give the subject the most information. As discussed in chapter two, a common interpretation of the results of this particular task is that subjects are misinterpreting what was intended to be a closed-scenario problem (where it is appropriate to rely on deductive inference) as an

---

[132] Elio, R., (2003), p. 197.

[133] Evans, J. St. B. T., Newstead, S. E., and Byrne, R. M. J., (1993), p. 101.

[134]Wason, P. C., (1966).

open-scenario problem (where it is appropriate to use inductive inference). Oaksford and Chater[135] argue that real-world situations involve verification and not falsification and that subjects may be flipping over the cards that will verify rather than falsify the conditional given. The fact that subjects are interpreting conditional statements as claims that can be more easily verified than falsified supports the argument that subjects are employing a probabilistic interpretation of these conditional statements.

Subjects analyzing abstract data may employ strategies that would be rational when examining data with content. For example, subjects provided with the conditional "If you turn your key in the ignition, then the car will start" may imagine situations where they turn their key in the ignition. The question is whether it is rational for subjects to imagine possible situations where the car does start or those where the car does not start. If we are operating with a standard truth-functional interpretation of conditionals, then subjects should choose to examine only situations where the car does not start–- which would mean that the majority of subjects' responses on the Wason selection task appear to be normative errors. If however, as a probabilistic account would suggest, subjects are looking for a general connection between the antecedent and the consequent, then it is rational for subjects to select the true consequent card. For example, showing that there is a case where you turn your key in the ignition and the car does not start, does not disprove the conditional if we consider the conditional probabilistically, i.e., as a non-exceptionless generalization. Confirmation bias may actually show that subjects do not interpret conditionals standard truth-functionally as evidenced by the fact that their selections indicate that they are not attempting to falsify

---

[135] Oaksford, M,, and Chater, N. (2007).

the rule by searching for counter-examples–- they are looking to see whether the rule holds generally.

The claim that subjects are simply operating with a probabilistic interpretation of conditionals in mind is further supported by experiments that involve asking the subject which case would falsify the rule given. Subjects appear to know which case will falsify the rule but opt instead to select the confirming cards— this indicates that their selections are not merely based on poor reasoning. As Evans et al. argue,

> "another interesting aspect of confirmation bias to come from these studies seems to suggest that this bias cannot be explained by arguing that people are merely bad reasoners or do not understand the task at hand. In some studies the subjects are asked to explicitly state which conditionals would falsify the rule expressed by the conditional. Subjects appear to easily see that the *p* card and the *not-q* card are the cards they need to flip over in order to falsify the rule they are given— yet this does not cause them to change their initial decision to select the non-falsifying cards."[136]

These intuitions lead subjects to make selections on the Wason task that are thoroughly incompatible with a standard-truth-functional analysis of conditionals but appear to show that conditionals are interpreted most often as a general rule.

---

[136]Evans, J. St. B. T., Newstead, S. E., and Byrne, R. M. J., (1993), p. 102

Matching Bias

Another bias seen in the Wason selection task and the truth table task is called matching bias. Matching bias refers to the tendency to select values for cards that match the rule. For example, when subjects are given the conditional “If P, then not Q” they tend to choose to examine the potentially falsifying cases of the true antecedent and false consequent more frequently than they do when they are given a conditional with a positive antecedent and consequent. In the Wason selection task, subjects choose to turn over the card that will verify the antecedent more often on rules with an affirmative antecedent and cards that would falsify the antecedent more often when the antecedent is negated in the rule. Likewise, subjects choose to turn over the card that will verify the consequent more often when the consequent is affirmed and attempt to falsify the consequent more often when the consequent is negated.[137] In 1972 Evans performed a study where subjects were given a conditional rule and a possible truth table and asked which line on the truth table falsifies the conditional rule given. He noted that subjects choose the line where the antecedent is true and the consequent false as the correct falsifier more often when the rule given is “if P, then not Q” and least often when the rule given is “if not P, then Q”.[138]

The historical explanation of matching bias was that subjects are employing an heuristic, i.e., subjects turn over the P and Q cards simply because they have just seen a P and a Q. This assumption does little to explain subjects' selections–- it merely discounts them as extra-logical errors. Both probabilistic accounts and standard truth functional accounts of conditionals would recognize the P and Q as the falsifying cards

---

[137]Evans, J. St. B. T., Newstead, S. E., and Byrne, R. M. J., (1993), p. 110.

[138]Evans, J. St. B. T (1972).

in a task where the conditional given is, “If P, then not Q” and thus subjects' responses in this case are in line with these normative theories. The reason why matching bias is challenging is that it appears as though the introduction of the negation has suddenly made subjects better reasoners. Subjects appear to realize which cards will falsify the conditional when the negation is present–- but on a standard truth functional view this negation should be irrelevant to which cards the subject decides to flip over. This led early researchers to the conclusion that subjects are employing heuristics.

Oaksford and Chater's description of this bias is similar to their description of confirmation bias where they rely on the expected information gain of particular cards based on a probabilistic interpretation of the conditional. They argue that,

> “the ‘contrast set’ account of negation shows that because of the rarity assumption— that most categories apply to a minority of items— *negated* categories are high probability categories. Having a high probability antecedent or consequent alters the expected information gains associated with the cards … Consequently, matching bias is a rational hypothesis testing strategy after all.”[139]

The rarity assumption is explained by Oaksford and Chater’s contrast class hypothesis. The assumption behind this hypothesis is that in most categories of objects, the negation of a category is far more common than the non-negated category. For example, given a standard deck of cards the chance of selecting an Ace is 4/52 whereas the chance of selecting a non-Ace is 48/52. The fact that subjects are more inclined to select the falsifying cards when the consequent is negated is rational if

---

[139] Oaksford, M., and Chater, N., (2009), p. 79. Clarification: the ODS model here refers to the Optimal Data Selection model proposed by Oaksford and Chater.

subjects are interpreting the conditional probabilistically. Subjects may (reasonably) assume that for most categories the probability of ‘not Q’ is much higher than the probability of ‘Q’ and, when the negation is present, they should be expected to turn over the falsifying cards.

The fact that Oaksford and Chater are able to describe matching bias as a rational belief testing mechanism makes their account far more compelling than accounts that rely on the hypothesis that subjects are simply employing an heuristic. The claim that subjects are reasoning with a probabilistic conditional has the ability to provide a unified account for why the introduction of a negation makes subjects better reasoners in the case of matching bias but worse reasoners in the case of the conditional inference task (where they tend to accept *modus ponens* but reject *modus tollens*). The alternative description that suggests that subjects reason more poorly with negations except when they are using the heuristic “select terms that match those in the question conditional” is needlessly complicated.

Negative Conclusion Bias

Negative conclusion bias is seen in the conditional inference tasks— more specifically the acceptance task where subjects are asked whether a certain conclusion is valid given a set of premises. In the conditional inference task (sometimes referred to as the conditional syllogism task)[140] the subject is given a conditional statement along with another premise and is asked either what follows from the two statements, whether a particular conclusion is valid, or to choose from a list of possible conclusions. For

[140] For example, Barrouillet, P, and Gauffroy, C. (2013).

example, the subject will be given the statements “if I am a bear then I have claws” and the statement “I do not have claws” and is asked what, if anything, follows from the two statements. In another common form of this task the subject is also given the statements “I am not a bear,” “I do have claws,” ”I am a crocodile,” and “I am a bear” as a list of possible conclusions and is asked to select all or any of the conclusions that can be validly drawn. Occasionally tasks of this form are also referred to as acceptance tasks. In this task the subject is given a set of premises and a set of conclusions representing the valid inferences of *modus ponens* and *modus tollens* and the common invalid inferences of affirmation of the consequent and denial of the antecedent and is asked which, if any, of the conclusions follow. Negative conclusion bias is the tendency in these tasks for subjects to more readily accept an inference (valid or invalid) when the conclusion is negated.

There are some experiments that seem to counter the claim that the negative conclusion bias truly is systematic. In 1981 Evans and Brooks found that when subjects were asked to perform concurrent articulation with a working memory load[141] the bias disappeared.[142] In another study there appears to be evidence that negative conclusion bias disappears when working with less abstract problems. In 1975 Fillenbaum studied the acceptance of denial of the antecedent with conditionals that can be interpreted as threats or demands. The introduction of negations did not impact the results in this

---

[141]Evans, J. St. B. T., and Brooks, P. G. (1981). In this experiment subjects were asked to articulate a series of digits while performing conditional reasoning tasks.

[142]Ibid.

study.[143] In light of experimental results surrounding permission schemas[144] it seems as though the results from Fillenbaum's study should be interpreted as bearing on the particular nature of threats and demands themselves rather than on the systematicity of negative conclusion bias itself. Given that there are only these few studies disconfirming negative conclusion bias (and the types of conditionals used in these studies are elsewhere shown to be idiosyncratic), I propose that negative conclusion bias is in fact systematic.

The experimental evidence suggests that, with the exception of *modus ponens* (which shows no negative conclusion bias), subjects tend to accept inferences more frequently (either valid or invalid) when the conclusion is negated.[145] As seen in the previous section, Oaksford and Chater have supplied a framework such that biases involving negation can be explained by assuming that subjects are relying upon a probabilistic interpretation of conditionals, rather than a standard truth-functional one. They claim that they "appeal to the idea that most categories apply only to a minority of objects. Hence the probability of an object being, say, red is lower than the probability of it not being red."[146] According to Oaksford and Chater, "therefore, an apparently

---

[143]Fillenbaum, S. (1975).

[144]A permission schema is a particular way of wording a conditional such that it can be interpreted as a social rule. Subjects who are given conditionals in this form overwhelmingly choose standard truth-functional answers particularly in the Wason selection task but also in other tasks.

[145] Pollard, P, and Evans, J, St. B. T. (1980).

[146] Oaksford, M., and Chater, N., (2009), p 76.

irrational negative conclusion bias can be seen as a rational “high probability conclusion” effect.”[147]

In this section I have argued that a probabilistic account of conditionals, such as that proposed by Oaksford and Chater, has the ability to render many systematic effects and biases seen in the psychological literature on human reasoning rational belief testing mechanisms. If we assume that subjects are reasoning with a probabilistic account of conditionals, negative conclusion bias and matching bias both become rational belief testing strategies. Overall, the assumption that subjects in these reasoning tasks are interpreting conditionals probabilistically provides a far better description of the results than the alternatives. It should be noted that the probabilistic account is not universally accepted among experimental psychologists.[148]

Probabilistic Accounts Offered by Experimental Psychologists

There are two promising psychological theories about reasoning with conditionals that rely upon a probabilistic interpretation. The first is a straightforward probabilistic account based on Bayesian probability theory put forward by Oaksford and Chater and the second is an amendment to the traditional mental models theory put forward by Schroyens and Schaeken. These psychological accounts of conditionals are able to predict nearly all of the subject responses on the conditional inference tasks. I argue that the alternative mental models theory proposed by Schroyens and Schaeken is similar in important ways to Stalnaker’s account of conditionals, supporting the idea that the logical account of conditionals most compatible with this psychological theory is

[147] Ibid., p 77.

[148] For example see: Frosch, C. (2011).

Stalnaker's account of conditionals. The success of Schroyens and Schaeken's account of conditionals at predicting results on the conditional inference task lends further support to the claim that natural language users primarily interpret conditionals probabilistically and specifically in accordance with an account that has the essential features of Stalnaker's account.

Oaksford and Chater attempt to describe experimental data on reasoning tasks by assuming a probabilistic view of human reasoning that they call the Optimal Data Selection Model.The probabilistic model offered by Oaksford and Chater takes many of the effects and biases discovered in studies on human reasoning to be rational belief testing strategies. Most notably they convincingly conceive of negative conclusion bias, matching bias, and confirmation bias as rational strategies for testing beliefs. They can also accommodate the data seen in the conditional inference task, claiming that

> "the conditional probability model provided a better fit for 54 out of 65 studies ..., and accounted for 84.5% ... of the variance ...That is, even when error is allowed for in the logical model, the conditional probability model fits the data much better."[149]

Oaksford and Chater argue that the results of some recent studies are only explicable if we assume a probabilistic theory of human reasoning. As Oaksford and Chater argue,

> "there is also data that seem to be consistent only with a conditional probability interpretation. For example, purely probabilistic manipulations in both the selection task and the conditional inference

[149]Oaksford, M., and Chater, N., (2003), p. 369.

task have been shown to have marked effects on participants inferential behavior."[150]

A significant problem with the optimal data selection model is that the calculations that explain many of the effects and biases involving negation rely upon the contrast class hypothesis discussed on page 101. Oaksford and Chater claim that this hypothesis is entirely rational if the subjects take themselves to be engaged in an inductive rather than a deductive task, in other words, if they take themselves to be testing a general rule. However, it is not clear how Oaksford and Chater would explain the operation of the rarity hypothesis without relying upon something like mental models. In order for this hypothesis to truly be rational, subjects would need to have some way of modelling the information (if the information given is not abstract in nature) in order to determine whether the rarity hypothesis is correct for this specific type of information–- otherwise the contrast class hypothesis seems merely to be another heuristic. For example, if subjects are given the premise "the ball will be white" in a case where there are an even number of white and black balls, then the assumption that the negation of this premise is more likely would not be rational. Thus, Oaksford and Chater's theory may require some acceptance of mental models. Also interesting to note is that the alternative mental models theory proposed by Schroyens and Schaeken[151] is consistent with more data on the conditional inference task than Oaksford and Chater's theory–- implying that mental models are better at accounting for

---

[150]Ibid., p. 376.

[151] Schroyens, W. J., and Schaeken, W. (2003), and (2008), and Schroyens, W., Schaeken, and W., Dieussaert (2008).

and predicting subject behavior on conditional reasoning tasks than the computational model proposed by Oaksford and Chater.

Mental Models Theory

Mental models theory was proposed most notably by Johnson-Laird, and also by Byrne and Schaeken. According to this theory, subjects reason by constructing models that represent a world in which the antecedent is true and then use this model to generate their conclusions. Johnson-Laird argues that there are three computational constraints on human reasoning: “to maintain semantic information, to simplify, and to reach a new conclusion.”[152] There are also three extra-logical constraints; “people do not usually throw away semantic information”, “conclusions should be simple or parsimonious”, and “conclusions should be informative.”[153] So reasoners occasionally avoid deductive inferences that decrease semantic information, that reassert something that has already been stated, and subjects tend to search for conclusions that contain new information. Johnson-Laird and other mental models theorists explicitly reject the notion that there are formal rules underlying subjects' selections on reasoning tasks. According to mental models theory, people do not reason using formal rules but rather construct a mental model of a world in which the premises are true and then “look” to see which conclusions follow.

According to Johnson-Laird, human reasoning entails three stages or steps. The first step is the model-constructing step where subjects construct a model of a world in which the given premise is true; in this step background knowledge relating to the content of the premise is used by the subject to flesh out the model which could account

---

[152]Johnson-Laird, P. N., and Byrne, R. M. J. (1991), p. 22.

[153]Johnson-Laird, P. N. (1983).

for the apparent content effects on human reasoning. The second step is the model-combining step where the subject combines the different models they have constructed for the initial premises (when there is more than one premise given) and combine them to form another parsimonious model that is a combined set. For example, they take the model they have constructed for “if I turn my key in the ignition, then my car will start” and combine this with the model they have constructed for “I turn my key in the ignition,” they then discover, and eliminate, any contradictions or inconsistencies in the final set. The third stage of human reasoning requires revising the models that have been created to either produce possible conclusions or to falsify a given conclusion.

Schroyens and Schaeken accept most features of traditional mental models theory but change the theory in some experimentally and philosophically significant ways. Firstly, traditional mental models theorists, such as Johnson-Laird, do not represent the possibility that subjects will construct a model where the antecedent is true and the consequent is false. This model is empirically impossible when the conditional is taken to be true, but is hypothetically possible and Schroyens and Schaeken believe that many subjects hypothetically adopt this model in order to search for possible affirmations of it. Secondly, Schroyens and Schaeken propose that a non-mandatory search for counterexamples is a large part of subjects reasoning procedures. For example, subjects will hypothetically adopt a model that falsifies the conclusion and then examine the acceptability or probability of this model. Schroyens and Shaeken “propose that after people have constructed a mental model of the conditional rule and perhaps fleshed it out, they then perform a validating search of long term memory for potential counter-examples.”[154] Thirdly, Schroyens and Schaeken argue that validation by falsification is not the only strategy used by subjects in conditional reasoning tasks, occasionally subjects seek to confirm the rule given.

---

[154]Oaksford, M., and Chater, N., (2003), p. 373.

The computational model constructed by Schroyens and Schaeken is called SSCEPPTRE (a syntactic-semantic counter-example prompted probabilistic thinking and reasoning engine). The processing tree for SSCEPPTRE can be explained as follows: people initially construct a model where the antecedent and consequent are both true and thus accept *modus ponens* except when they can construct a model where the antecedent is true and the consequent is false and deem this model likely. Subjects also accept affirmation of the consequent except when they construct and deem likely a model where the antecedent is false and the consequent true. Subjects do not initially construct models that are consistent with *modus tollens* or denial of the antecedent so their likelihood of accepting these inference forms is based on their ability to construct a hypothetical model that confirms the inference form against their ability to construct a model that falsifies the form.

The theory proposed by Schroyens and Shaeken predicts results on the conditional inference task better than the historical theories of formal rules and mental models. This theory also accommodates the experimental data better than the straightforward probabilistic analysis offered by Oaksford and Chater. According to Oaksford and Chater, “their model provides a small (8.2%), but significant, improvement in the proportion of variance accounted for over the conditional probability model.”[155] Note that this increase means that SSCEPPTRE accounts for 92.7% of variance in the conditional inference task. The alternative mental models account also can explain why *modus ponens* and affirmation of the consequent are accepted more often than *modus tollens* and denial of the antecedent as subjects' initial models are consistent with the

[155]Ibid., p. 374.

positive inference forms but not the negative inference forms. This model also explains why subjects are more likely to accept *modus ponens and modus tollens* than they are to accept denial of the antecedent and affirmation of the consequent, since the cases that potentially falsify the valid inference forms require the antecedent to be true and the consequent false. This is not empirically possible and most subjects reject this as a likely model.

Stalnaker's Account of Conditionals and Mental Models

Stalnaker's account of conditionals shares both a philosophical foundation and many essential features with the alternative mental models theory of conditionals. Both have some basis in Ramsey's test, both have a probabilistic interpretation of conditionals, both offer a unified account for both indicatives and subjunctives, and both have the ability to represent complex suppositions. The belief conditions for conditionals provided by Stalnaker's theory and the process of human reasoning that is outlined by Johnson-Laird et al. are also similar in significant ways. The former requires looking to a scenario where the antecedent is true in order to determine the truth value of a conditional and the latter argues that people reason by constructing models where the premises given are true. To be compatible with the alternative mental models account of reasoning a logical account of conditionals would need to share these characteristics. By showing that Stalnaker's account (or alternate account with these features) is compatible with the most descriptively accurate psychological account I intend to show that Stalnaker's, or similar probabilistic account, is able to describe the majority of competent uses of conditionals as seen in reasoning tasks.

The characteristic which distinguishes mental models theory from formal rules theory is the same as that which distinguishes Stalnaker's account from the standard truth functional account. Both of these theories share the assumption that conditional statements are essentially about possibility. In their explanation of traditional mental models theory Johnson-Laird and Byrne argue that, “the antecedent of a conditional establishes two possibilities, either two factual possibilities or a fact and a counterfactual possibility. The antecedent refers to a possibility, and the consequent is interpreted in that context.”[156] This is one of the few times that Johnson-Laird and Byrne write about the analysis of conditional statements singularly (not as part of a premise set). With this statement we can see that mental models theorists share with Stalnaker the underlying assumption that conditionals are about possibilities. Specifically, both mental models theorists and Stalnaker accept Ramsey’s test where we add the antecedent hypothetically to our stock of beliefs and then check to see whether the consequent is true. In spite of this acceptance of Ramsey's test, Johnson-Laird et al. do not believe that conditionals are interpreted probabilistically and argue, as will be seen in the next chapter, that experimental evidence is not in support of this view. However, Schroyens and Schaeken propose the alternative mental models theory that does embrace a probabilistic analysis of conditionals.

It is also notable that the proponents of mental models theory claim that subjunctive and indicative conditionals are amenable to roughly the same analysis. Other empirical psychologists tend not to write about subjunctives since there is relatively little experimental data on this subject. This poverty of experimental data is

---

[156]Johnson-Laird, P. N., and Byrne, R., (2002), p. 649.

unfortunate for logicians since the question of whether indicatives and subjunctives admit of the same analysis or are different in significant enough ways to warrant separate analyses is still a point of contention for those working on the logic of conditionals. This issue is part of what separates the analyses of Stalnaker and Lewis/ Jackson for example. Stalnaker argues that the only difference between an indicative and a subjunctive conditional is that the antecedent-including world that is considered in the former happens to be the actual world. For logicians and philosophers who accept a unified probabilistic account,[157] the claim that subjunctive conditionals are about possibility is the more obvious claim and generally the assumption that comes first. By which I mean, it is first assumed that subjunctives are in some sense about possibility and then indicatives are explained in terms of this. Johnson-Laird et al. assume that indicatives are based on possibility and then explain subjunctives in terms of this assumption. They argue that, “basic indicative conditionals have core meanings that refer to sets of possibilities and... basic subjunctive conditionals also have such core meanings.”[158] So, for these proponents of mental models theory a subjunctive conditional refers to sets of possibilities just as do indicative conditionals.

These similarities evince their significance primarily because they lead to similar ideas about how a conditional is analyzed in practice or how reasoning with conditionals proceeds. Stalnaker's belief conditions for conditionals where a possible antecedent-including-scenario is imagined bears many similarities to the mental models theory where a model is constructed that corresponds to the initial statement given and is then adjusted based on further information. According to mental models theory, subjects

---

[157] For example, Davis, Ellis, and Stalnaker.

[158]Johnson-Laird, P. N., and Byrne, R. (2002), p. 653.

reason by constructing models that represent a situation in which the initial premises are true and then use this model to generate their conclusions. Johnson-Laird et al. argue that subjects construct a mental model of a situation in which the premises are true and then “look” to see which conclusions follow. If we take Stalnaker's belief conditions for conditionals, whereby we imagine a possible scenario (or scenarios) in which the antecedent is true and then look to see whether the consequent is true in that antecedent-including-scenario, then we are left with essentially the same theory (adjusting for the fact that Stalnaker is speaking of conditional statements where Johnson-Laird et al. are speaking of arguments).

Mental models theory attempts to explain the significant difference between experimental results involving *modus ponens* and those involving *modus tollens* (the former is accepted far more frequently and content effects appear to be slightly less relevant). The proposed explanation is that the first model constructed by the subject is the one where the antecedent and the consequent are both true.[159] In *modus ponens* this model does not need to be revised, whereas for *modus tollens* it does. For *modus tollens* the model constructed is one where both the antecedent and the consequent are true but the second premise rules out this model, so multiple models need to be constructed increasing the difficulty of this inference.[160] This is why *modus tollens* is a more difficult inference— subjects make the inference less often and it takes subjects on average more time to complete *modus tollens* inferences— than *modus ponens*.

---

[159] Johnson-Laird, P. N., and Byrne, R. M. J. (1991).

[160] Ibid., p. 22.

For Stalnaker, the antecedent-including-scenarios that are imagined may contain both consequent including and consequent not including scenarios. Given that the studies where this competence is exhibited all involve sets of statements, there is a possible way for Stalnaker to explain this increased competence with *modus ponens*. It may be that when reasoning with sets of statements the premises take the place of the antecedent and a possible scenario is imagined that includes them all. For example, the subject could take the combined premises of the argument as the antecedent of a conditional— "((If P, then Q) and ~Q), then ~P". The subject would take the first premise "if P, then Q" as describing the antecedent-including-scenario they are meant to imagine. This could lead to the same initial scenario as proposed by mental models theory since subjects would most likely be imagining possible scenarios where both P and Q are true. When they are then given "~Q" as a second premise, this would force them to alter their initial imagined scenario and imagine possible situations where "if P, then Q" and "~Q" are both true. The practical application of Stalnaker's theory (the steps that people go through when reasoning) would be nearly identical to those proposed by the mental models theory. Stalnaker could adopt the explanation offered by mental models theorists and claim that the initial set of antecedent-including-scenarios that is examined by a subject must be revised when they are given the negated consequent as a premise, i.e., the negated consequent as a premise forces subjects to examine the set of possible scenarios that they had previously constructed and eliminate those scenarios where the consequent is true. This would account for why subjects are less proficient with *modus tollens* inferences and could also account for why subjects take more time on tasks involving *modus tollens* inferences.

Stalnaker's Account and the Suppression Effect

In addition to being compatible with the psychological theory that I have argued is the most descriptively accurate, Stalnaker's account of conditionals can also account for a phenomenon seen in the psychological experiments that has proven difficult for psychological theories to convincingly reconcile. Specifically, Stalnaker's theory can resolve the tension in what experimental psychologists refer to as the suppression effect.

The suppression effect is the terminology used to refer to the different factors that can suppress the acceptance of valid or invalid inferences. For example, subjects will be given a conditional statement, “If there is a ball in the box, then the ball is red” and an unconditional statement, “there is a ball in the box” along with additional statements that suppress particular inference forms such as, “none of the balls is red.” The results of Byrne's experiments on the suppression effect where subjects fail to make valid inferences due to the introduction of additional conditional premises, can be accounted for by Stalnaker’s account of conditionals. In Byrne's study the conditional “if the library stays open late, then she will study” is given to subjects. When subjects are given the additional statement “the library stays open late” they accept the conclusion “she studies” with regularity and ease; showing a basic, and widely accepted, competence with *modus ponens*.

In Byrne's study some of the subjects are given the additional conditional “If she has a paper to write, then she will study” which suppresses the subjects' *modus ponens* inference. Byrne found that the group with simple arguments (a statement containing

the conditional and a statement containing the antecedent) made the *modus ponens* inference 96 percent of the time; the group with additional arguments (two conditionals are given each with the same consequent but different antecedents and a statement containing the antecedent of the first conditional is given) made the *modus ponens* inference only 33 percent of the time. The additional conditional statement gives a second condition for the consequent. Stalnaker's account of conditionals can accommodate the suppression of valid inferences since the addition of further statements will affect the antecedent-including-scenarios that are imagined.

More interesting is that Stalnaker's account is not only able to reconcile the suppression of valid and invalid inferences through additional conditional premises but can also account for why we see the suppression effect when there is an implied or actual time lag and when modal words are included in the premises, conclusion, or question given to subjects. Specifically, for Stalnaker's account of conditionals the additional statements would limit the possible antecedent-including-scenarios that are selected. On Stalnaker's account the antecedent-including-scenario that is selected must do the least violence to the current world as possible which means that we must select an antecedent-including-scenario that takes as true all the things that we take for granted in the context. Adding additional conditions in the tasks gives the subjects additional information that must be taken for granted in the context, i.e., that should be included in the truths of their selected antecedent-including-scenario. For example, in Byrne's study described above subjects would be selecting possible scenarios where the antecedent "she has a paper to write" and "the library stays open late" both obtain. This would lead subjects to mistakenly infer that the antecedent is a conjunction. If this

were the case, then the subject's responses, given that we are told only that she has a paper to write, would be that she may or may not go to the library— which is the conclusion that subjects overwhelmingly chose. Note that the form of this particular task can be seen as giving insight into natural language users acceptance of the inference form strengthening the antecedent. Subjects on these tasks reject antecedent strengthening which is an inference form that Stalnaker's account also rejects.

The results of the second part of the study also provide support for Stalnaker. Based on the account put forward by Stalnaker we would predict that, if subjects are given the premise, "she has a paper to write and the library stays open late," they will make the *modus ponens* inference since the antecedent-including-scenario they are selecting is one where "the library stays open late" and "she has a paper to write" both obtain. All of the subjects in the second part of Byrne's study chose this conclusion.

Byrne also showed that the duration cited to take place between the conditional premise and the second premise has an impact on subjects' valid and invalid inferences. In problems where a short duration was said to have taken place between the first and second premise valid inferences were suppressed (*modus ponens* dropped from 72 percent to 41 percent), and invalid inferences were suppressed in the context of a long duration (denial of the antecedent dropped from 44 percent to 19 percent). Byrne proposed that, "the short duration reminds subjects that there may be background conditions that might not be met."[161] For example, subjects are given the paragraph:

---

[161] Evans, J. St. B. T., Newstead, S. E., and Byrne, R. M. J., (1993), p. 59.

> "a policeman came up to James at the protest and told him that if he went into the library, then he would be arrested. The next day James went into the library to study for an exam."

The time lapse said to have taken place in this paragraph reminds subjects that there are qualifying conditions on the conditional which also limits the antecedent-including-scenarios that can be selected. For example, they are no longer examining possible scenarios that include only the antecedent "James went into the library" but realize that the possible scenario examined also must contain "the protest is still going on." In this case, the subject's rejection of the *modus ponens* inference is warranted since the time lapse may signal that the conditional premise is no longer true.

Another of the factors leading to the variability in the suppression of valid and invalid inferences appears to be the use of modal operators in either the conditional the subjects are asked to evaluate or in their evaluation itself. In 1973 Taplin and Staudenmayer[162] performed a study designed to determine the effect of procedural changes on deductive reasoning. Their primary goal was to determine how people interpret "if…, then…" statements— whether as a conditional, biconditional, or inconsistently. Taplin and Staudenmayer got different acceptance results depending upon whether they asked the subjects if the inference was always true or false or sometimes true or false. They found that, when subjects had "sometimes true" or "sometimes false" among the possible conclusions, subjects interpreted the "if…, then…" statement given more often as a truth-functional conditional rather than as a biconditional or inconsistently (36 percent versus 2.8 percent). The conclusion that

---

[162] Taplin, J. E., and Staudenmayer, H. (1973).

Taplin and Staudenmayer draw from these results is that they “provide further evidence to indicate that the meaning of a sentence is not solely a function of the intrinsic properties of the sentence itself but is also dependent upon context as well.”[163] Alternately a conclusion that could be drawn is that subjects analyze conditionals probabilistically and therefore they are more inclined to claim that a conditional is “sometimes true” or “sometimes false” since this is in agreement with the subject's potential evaluations. If subjects receive the conditional, “if you turn the key in the ignition, your car will start” they most likely imagine a variety of possible scenarios where it is true that they turn the key in the ignition. In this case the conditional would be true in only the possible scenarios where the car has a full gas tank, where there is nothing wrong with the engine, where the battery is not dead, etc. In some of the antecedent-including-scenarios the conditional is true and in some the conditional is false, so, given a probabilistic account of conditionals, subjects would be expected to select “possibly true” or “possibly false” as their analysis. This is the response overwhelmingly seen in these tasks.

A probabilistic theory of conditionals that is consistent with maintaining that subjects answer conditional reasoning problems by constructing mental models— such as Stalnaker’s account— has the ability to account for most of the effects and biases seen in the literature on conditional reasoning. It can accommodate the invalid inferences accepted in the cases of suppression effect studies, it can interpret both matching bias and negative conclusion bias as rational belief testing strategies, and it is also able to account for the variation among subject responses and why content effects

---

[163] Ibid., p. 530.

matter since these will affect the antecedent-including-scenarios selected. Also, Scroyens and Schaekens' model, which combines mental modelling and probabilistic analysis can predict the outcome of particular experiments better than the other proposed models. The fact that this psychological theory of conditionals shares essential features with Stalnaker's account of conditionals supports the claim that the most descriptively accurate logical account of conditionals will be one with these features. The necessary features being: some basis in the Ramsey test, a probabilistic conditional, a unified account for both indicatives and subjunctives, and the ability to represent complex suppositions. Therefore, the predictive accuracy of SSCEPPTRE supports the conclusion that Stalnaker’s hypothesis aligns with competent natural language use of conditionals. Further, this data suggests that a theory of conditionals that does justice to natural language use must be both probabilistic and must be consistent with the evidence that subjects construct mental models when faced with conditional reasoning problems. Stalnaker’s account is the only account with these features. There are some experiments by psychologists studying human reasoning that are designed to show that a probabilistic theory of conditionals is not descriptively accurate. In the next chapter I will be examining experimental evidence that is presented to challenge the view that the most descriptively accurate account of conditionals must be probabilistic.

Chapter Five: Experimental Evidence that Challenges the Probabilistic Account

In this chapter I examine the empirical data that challenges a probabilistic interpretation of conditionals. First, I examine arguments from Braine and O'Brien who address Stalnaker's early interpretation of conditionals[164] and conclude that it could not be considered a general account of conditional reasoning. I argue that the characteristics Braine and O'Brien consider detrimental to Stalnaker's account are actually advantages and that their claim that Stalnaker's hypothesis fails to adequately describe a psychological mechanism for conditional use is not a problem for a logical theory. Next I address arguments from Johnson-Laird and Byrne whose main consideration against a probabilistic account of conditionals is their argument that such accounts are not descriptively accurate. I examine the experiment cited in support of this view and conclude that the design of the experiment does not support their conclusion. In the last section of this chapter I examine a study done by Douven and Verbrugge that purports to challenge the claim that people interpret conditionals using something like Adams's thesis. Adams's thesis is roughly that the assertibility of a conditional correlates or moderately correlates with the probability of the consequent given the antecedent. As(A→C) ≈ P(C/A). They conclude that most forms of conditionals are only interpreted using a very weak version of Adams's thesis. I examine their results in detail and argue that the conclusion of their paper— that Adams's thesis is not descriptively accurate— is too strong. The experiment cited by Douven and

---

[164] As described in chapter three pp. 65 - 72 and outlined in: Stalnaker, R., (1968); and Stalnaker, R., and Thomason, R. (1970).

Verbrugge shows merely that conditional probability is not the only factor subjects use in assessing the truth of a conditional. Their results show that subjects rely heavily upon a weak version of Adams's thesis when evaluating all types of conditional, lending support to the notion that a general account of conditionals will be probabilistic. Further, experimental evidence against the descriptive accuracy of Adams's thesis would only show that any probabilistic account of conditionals must be based on truth rather than acceptability or assertibility— it would not show that all probabilistic accounts are descriptively inaccurate.

Psychologists on the Possible Worlds Approach

Though an examination of empirical results from psychologists on human reasoning is rarely found among the philosophical literature on logic, psychologists fairly often incorporate theories in philosophy or logic into their own work or examine logical theories in a psychological way. Braine and O'Brien examine Stalnaker's possible worlds approach to conditionals and examine its potential usefulness as a psychological theory and Johnson-Laird and Byrne discuss whether a probabilistic account of conditionals is descriptively accurate. Braine and O'Brien examine possible worlds semantics to determine if it better defines people's knowledge of 'if' than mental models or suppositions. They conclude that possible worlds semantics is not sufficient to describe natural language use of conditionals since it is more complex than suppositions. Meanwhile, Johnson-Laird and Byrne report the results of an experiment that they take to show that subject assessments of the probability of conditionals do not match their assessments of conditional probabilities.

Braine and O'Brien on Stalnaker

Braine and O'Brien, in their article "A theory of If: Lexical Entry, Reasoning Program and Pragmatic Principles," evaluate whether Stalnaker's logical account of conditionals is suitable as a psychological theory of conditional reasoning. As I do not argue that Stalnaker's account of conditionals should take the place of a psychological theory, part the argument made by Braine and O'Brien in this article is not entirely pertinent. However, their arguments can be used to evaluate both the descriptive accuracy of the type of probabilistic account that Stalnaker offers, and the similarities between the mental models account and Stalnaker's account of conditionals that I argued for in the previous chapter.

Braine and O'Brien begin their examination of Stalnaker's account of conditionals by dismissing the idea that possible worlds can function like mental models. I will be examining this argument, not to show that possible worlds should take the place of mental models, but to argue that Stalnaker's account of conditionals and mental models theory are not as dissimilar as Braine and O'Brien argue. They argue that Stalnaker's account of conditionals fails to function well as a psychological theory because possible worlds carry too much information and are too complex to play the psychological role that mental models or suppositions do. Their argument is inspired in great part by the fact that the truth table for a mental model is incomplete and because probabilities (for most versions of mental models) are not represented. This is not an argument against Stalnaker's account of conditionals as a logical account unless Braine and O'Brien can show that Stalnaker's account is inherently incompatible with the psychological mechanisms underlying conditional reasoning. Braine and O'Brien fail to show this

incompatibility and they make too much of the difference between the Ramsey test used for Stalnaker's belief conditions and mental models. Primarily, they fail to consider the aspect of Stalnaker's account that is intended to describe belief conditions— his epistemic analysis. Possible worlds, as part of Stalnaker's epistemic analysis of conditional reasoning, are really ways to model complex suppositions since natural language users clearly do not imagine actual possible worlds when reasoning. The use of the term 'possible world' in the logical analysis is a far more complex notion and it is this logical usage, and not the possible worlds described in Stalnaker's description of belief conditions that Braine and O'Brien take issue with. The main issue with the criticism outlined by Braine and O'Brien is that they attempt to evaluate Stalnaker's account as a general account of conditional reasoning, but rather than examining the aspect of Stalnaker's account that attempts to describe this natural language reasoning — the belief conditions— they examine the logical analysis instead. Possible worlds, in Stalnaker's epistemic analysis, operate much like mental models in that our primary concern is whether the antecedent and consequent are true or false. More importantly, the epistemic analysis of conditionals offered by Stalnaker relies on a modified version of the Ramsey test where we add the antecedent to our set of beliefs to see whether the consequent follows— a process very much like that involved in mental modelling inasmuch as it is a way to model complex suppositions. Therefore, the difference between mental models and possible worlds is not nearly as wide as Braine and O'Brien imply— the epistemic analysis of conditionals offered by Stalnaker, since it is also a method of modelling complex suppositions, is in fact compatible with mental models.

Further, Stalnaker offers a description of the truth conditions for conditional statements in addition to his description of belief conditions. This connection between an epistemological and a logical account of conditionals leads to a number of benefits over the alternative mental models account. Most importantly, Stalnaker's account has the benefit of offering a logical account that eliminates the underdetermination of the epistemic account. A degree of uncertainty is expected in an epistemic account of conditionals but our logical account of conditionals (since it outlines the truth conditions for conditional statements) should provide determinate results for all conditional statements. It is Stalnaker's connection between the belief conditions for a conditional and a possible worlds analysis that allows for a determinate semantics for conditional statements. Stalnaker's account also has the benefit of connecting this account of conditionals to probability theory— the calculus of which is relatively well-defined. This gives us an account of conditionals which is far more mathematically rigorous than the alternative mental models theory.

The main criticism that Braine and O'Brien level against Stalnaker's theory of conditionals is that, according to them, the way of working out which possible world one is examining and what that possible world contains is too vague. This objection refers to the practical implications of possible worlds as a psychological theory. For example, they claim that, taken as a psychological theory, Stalnaker's account of conditionals doesn't explain how people determine what is in a given possible world and which possible world to select in cases where the conditional is subjunctive. In Stalnaker's theory, the selection function determines that the possible world selected be one where the antecedent is true and that it be one as much like the current world as possible.

Presumably, Braine and O'Brien, much like David Lewis, do not consider the selection function to be determinate enough. In other words, Stalnaker's theory doesn't provide enough detail as to how people compute possible worlds and how they decide what is in those worlds. Braine and O'Brien argue that Stalnaker's theory is incomplete because it does not leave us with the ability to determinately compute which possible world we should select and how that possible world differs from the actual world. Braine and O'Brien argue that,

> "the reasoner must have a cognitive representation of the semantics ... However, the semantics itself says nothing about how to compute what the world in which Napoleon attacked earlier would be like, nor to what kinds of world similarities the reasoner should attend."[165]

As argued in Chapter Three, this characteristic can be interpreted as one of the main benefits of Stalnaker's semantics. This is because there does not appear to be any one singularly relevant link for the analysis of all conditionals. Sometimes a conditional is used to signify a causal connection, sometimes a logical one, sometimes a metaphysical one, etc. For example, we may think of the connection between the antecedent and consequent in a causal conditional such as "if I turn my key in the ignition, the car will start" in a very different way from how we think of the connection between the antecedent and the consequent in the counterfactual conditional, "If JFK had not been shot, people would be living on Mars," and this very differently from how we think of the connection between the antecedent and the consequent in a conditional that is made true or false by the meanings of the terms such as, "If Jeremy is 2 meters

---

[165] Braine, M., and O'Brien, D., (1991), p. 199.

tall, then he is an odd number of meters tall." However, our different conceptions about the relevant similarities in each of these cases changes nothing about how each of the conditionals as a whole is interpreted. So the fact that, in Stalnaker's account, the reasoner is free to posit a causal, logical, metaphysical or other link depending upon what is relevant in the context is a benefit of his theory and not a detriment in the sense that it makes no sense to include a description of a difference that does not make a difference. Stalnaker's hypothesis describes conditional use in all cases regardless of whether the particular context warrants a causal, logical or metaphysical link between the antecedent and consequent. Providing an account that distinguishes these different interpretations would only be beneficial if it were shown that they systematically lead to different interpretations of conditionals— which is not the case. Braine and O'Brien argue that an account of reasoning is not complete without a clear explanation of how the antecedent world is selected by natural language users and how people determine which other propositions are true in that world. They claim that, given a more detailed and specific account "all the vague talk of world similarities could be dropped."[166]

I agree with Braine and O'Brien that some account of how people determine the similarities between models is necessary for a complete explanatory psychological account of human reasoning. However, the absence of a clear psychological mechanism for selecting models doesn't make possible worlds semantics descriptively inaccurate— unless Braine and O'Brien could argue that there is some feature of possible worlds semantics that would make it *impossible* to outline a psychological mechanism for the selection of models. For Stalnaker, there is a certain vagueness

---

[166] Ibid., p. 199.

imbued in our selection of possible worlds— he believes that this corresponds to and incorporates the fact that people frequently use conditionals in vague ways. If this is true, then a solidification of the possible world selection process (like the one that we get with Lewis) would be more parsimonious and more pleasing from a computational point of view but would be a horribly inaccurate representation of human reasoning. More importantly, positing a psychological mechanism for conditional use is not the purpose of a logical theory. Stalnaker has no position on the psychological mechanism involved in the selection of models— as this is the purview of a psychological theory. The lack of a specified psychological mechanism here obviously cannot be used as a measure of an adequate logical theory as the purpose of a logical theory is not to provide such an explanation. The psychological theory presented in chapter four (SSCEPPTRE) also does not posit a psychological mechanism for selecting models. This is because the theory is a descriptive rather than an explanatory psychological account as described in chapter one.[167] This failure does not speak against the theory and in favour of an alternate theory since, as Braine and O'Brien themselves point out, no alternative theories are able to posit such a mechanism either. In other words, while these theories provide an accurate descriptive account of conditional use in natural language, there is further work to be done in the area with regard to outlining a mechanism for how people select different models.

[167] Specifically, on page 8 and 9.

Johnson-Laird and Byrne on Probabilistic Conditionals

Johnson-Laird and Byrne outline what they take to be empirical evidence against Adams's thesis (they also mention Stalnaker's account of conditionals) and conclude that their theory based on mental models is a more accurate representation of how people use conditionals. They describe a probabilistic view of conditionals as the claim that "the probability of a conditional, *p(If A then C)*, is close to the conditional probability *p(C/A)*"[168] and argue that "the seductive nature of the claim depends in part on the syntax of English."[169] This description of a probabilistic view is very close to Stalnaker's hypothesis (we simply need to change the wording from "is close to" to "is equal to." However, the definition provided by Johnson-Laird and Byrne is not an accurate description of Adams's thesis as this thesis is usually interpreted as a claim about the assertibility or acceptability of a conditional rather than as a claim about the truth of a conditional.

Johnson-Laird and Byrne begin their discussion of conditional probabilities by admitting that there is quite a lot of empirical support for the conditional-probability hypothesis[170] some of which I myself covered in the previous chapter. They argue that the empirical support for Adams's thesis is caused by subjects slightly misunderstanding the question being asked such that they supply the probability for the wrong clause. Specifically they argue that subjects may take a question of the form "Is it true that if

---

[168] Johnson-Laird, P. N., and Byrne, R., (2002), p. 650.

[169] Ibid., p. 650.

[170] Specifically they cite the following studies: Stevenson, R. J., and Over, D. E., (1995), Evans, J. St. B. T., Handley, S. H., and Over, D. E., (2003), Hadjichristidis, C., Stephenson, R. J., Over, D. E., Sloma, S. A., Evans, J. St. B. T., and Feeney, A., (2001), and Oberaber, K., and Wilhelm, W., (2003).

Paolo has the Ace, then Maria has the King?" to mean "If Paolo has the ace, then is it true that Maria has the King?" Meaning that rather than supplying the probability of the conditional, Johnson-Laird and Byrne believe that subjects are instead giving the probability of the consequent. This would lead subjects to give the same probability for questions about the probability of a conditional Pr(If P, then Q) and the conditional probability Pr(Q/P) because they translate the former conditional into the latter. According to Johnson-Laird and Byrne "Investigators must take extra pains to ask for the absolute probability of a conditional instead of this conditional probability, or else participants are in danger of confusing the two."[171]

They attempt to block this misunderstanding by slightly altering the sentence representing conditional probability in this experiment. Typically, when evaluating the similarities between interpretations of the probability of a conditional and conditional probability subjects would be given the following narrative and sentences:

> "There are three cards on the table, a 3, 6 and an 8. Paolo selects one at random and then selects another card at random"
>
> "What is the probability that 'if Paolo has the 3, then he also has the 8' is true?" (testing the probability of the conditional)
>
> and
>
> "If Paolo has the 3, what is the probability that he also has the 8?" (testing the conditional probability)

[171] Johnson-Laird, P. N., and Byrne, R., (2002), p. 650.

Johnson-Laird believes that the latter sentence may be misleadingly interpreted by subjects and instead proposes sentences like the following to represent conditional probability:

“Paolo shows his card: It is the 8. What is the probability that he also has the 3?”

The motivation for this sentence is a particular interpretation of conditional probability that comes from Kemeny, Snell, and Thompson who describe conditional probability thusly:

> “The probability of p after the receipt of the information q is called its *conditional probability*, and it is denoted by Pr(P/Q) which is read ‘the probability of p given q”[172]

Based upon this understanding, Johnson-Laird asserts that the problem with using a sentence like “If Paolo has the 3, what is the probability that he also has the 8” to determine conditional probability is that the antecedent is not obviously true. Thus, Johnson-Laird and Girotto have attempted to come up with a sentence expressing conditional probability that clearly indicates that the antecedent is to be taken as true. This sentence is “Vittorio says ‘Paolo has the 3’. Given that, indeed, Paolo has the 8, what is the probability that this Vittorio’s assertion is true?” One issue with this particular understanding is Johnson-Laird and Girotto have provided an interpretation that does not consider conditionals where the antecedent is false or unknown to be evaluable according to conditional probability. Both Stalnaker and Adams have accounts of conditional probability that apply to the evaluation of conditionals in cases where the antecedent value is unknown or false. I believe that Johnson-Laird and Girotto’s

---

[172] Kemeny, J., Snell, L., and Thompson, G. (1957), p. 97.

oversimplified understanding of conditional probability and the way it has influenced their formulation of the problem presented to the subjects in this experiment has led to results that differ drastically from other experiments testing naive reasoning about conditionals. I will explain this concern further after I describe the experiment and the results.

Johnson-Laird and Byrne cite a study by Johnson-Laird and Girotto[173] which they take to show that subjects do not analyze conditionals probabilistically but rather according to mental models of them. In this study 20 subjects are asked to evaluate sentences uttered by the experimenter. The first question, meant to test conditional probability is:

> There are three cards face down on a table: a 3, a 6, and an 8. Paolo takes one card at random, and then he takes another card at random. Paolo shows one of his cards: it is the 8. Vittorio says "Paolo also has the 3". Given that, indeed, Paolo has the 8, what is the probability that this Vittorio's assertion is true?[174]

The second question, meant to test the probability of a conditional, is:

> There are three cards face down on a table: a 3, a 6, and an 8. Paolo takes one card at random, and then he takes another card at random. Vittorio says "If Paolo has the 8, then he also has the 3."What is the probability that this Vittorio's assertion is true?[175]

---

[173] Girotto, V., and Johnson-Laird, P. N., (2002).

[174] Girotto, V., and Johnson-Laird, P. N., (2002), pp. 218-219.

[175] Girotto, V., and Johnson-Laird, P. N., (2002), p 212.

According to mental models theory, subjects reason about conditionals by constructing models that represent a world in which the antecedent is true and then use this model to generate their conclusions. Johnson-Laird and Girotto propose in this paper an understanding of conditionals that differs slightly from the Ramsey test (and Johnson-Laird's earlier understanding of conditional reasoning) by proposing that subjects may also represent models where the antecedent value for the conditional is false.[176] The mental models theory, as outlined in this particular paper, predicts that subjects reason about probabilities of conditionals by representing the possibilities compatible with the premises and assuming that each of these possibilities is equiprobable. Johnson-Laird and Girotto predict that subjects will give answers to the second sentence in line with three potential strategies. In the first strategy, called the "equiprobable" strategy, subjects look only at mental models where the antecedent is true (so they discount every possibility where Paolo does not have the 3) and thus give a probability of 1/2 to the sentence "what is the probability that this Vittorio's assertion is true?" In the second strategy subjects interpret the conditional as a conjunction since they take the only confirming instance of the rule to be the case, i.e., where both Paolo has the 3 and also has the 8. They thus give a probabilistic estimate of 1/3 to the sentence representing conditional probability. Lastly, subjects who use a "complete" strategy have made a complete model of all of the possibilities for the conditional "Given that, indeed, Paolo has the 8, what is the probability that this Vittorio's assertion is true?" and thus give a probability that takes into account every possibility consistent with that conditional— 2/3. The description of the complete strategy would be an interpretation in

---

[176] For a longer description of mental models theory see my introduction to the theory on page 108.

line with the material conditional and the equiprobable strategy would be one consistent with Adams's thesis.

The results showed that 85% of subjects gave 1/2 as the probability of the conditional "If Paolo has the 8, then he also has the 3." The results for the conditional probability— the probability that Vittorio's assertion is true— were split with 40% of subjects giving the answer 1/2, 20% giving the answer 1/3, and 10% giving the answer 2/3. We can see that only approximately 40% of the subjects gave the same answer to both questions[177] and that Johnson-Laird and Girotto's predictions about the sentence they took to represent conditional probability were correct in that subjects interpreted this sentence in many different ways. According to Johnson-Laird and Byrne,

> "the results showed that naive individuals do not tend to give the same estimates for the two probabilities. The majority give the correct answer for the conditional probability, but the modal response for the probability of the conditional was the one based on mental models, and a handful of participants even responded 2/3"[178]

My primary concern with this particular experiment is that the second question meant to test conditional probability is unlikely to actually test conditional probability and is formulated in such a way that I think subjects are less likely to interpret it as a hypothetical. Reformulating the conditional probability question as— "Vittorio says 'Paolo has the 3'. Given that, indeed, Paolo has the 8, what is the probability that this Vittorio's assertion is true?"— is likely to mislead subjects. The introduction of an aside

---

[177] Since Johnson-Laird and Girotto do not show the percentage of subjects who gave the same answers for both questions this is an estimate based upon which subjects gave 1/2 as a response to both questions.

[178] Johnson-Laird, P. N., and Byrne, R., (2002), p. 651.

and additional sentences may further confuse some subjects as to which clause is the most pertinent. The fact that the conditional phrase ‘given that’ is separated from the antecedent by an aside, the fact that this sentence is phrased in such a way as to lead subjects to assume that the antecedent is true, and the fact that the utterance they are asked to identify the probability of is not formulated as a hypothetical— likely increased the chances that subjects would give the probability for the consequent (which 20% of subjects did). The results on this study differ drastically from other studies— for example, Stevenson, R. J., and Over, D. E., (1995), Evans, J. St. B. T., Handley, S. H., and Over, D. E., (2003), Hadjichristidis, C., Stephenson, R. J., Over, D. E., Sloma, S. A., Evans, J. St. B. T., and Feeney, A., (2001), and Oberaber, K., and Wilhelm, W., (2003)— that have shown great support for the conditional-probability hypothesis (Adams’s thesis). It is far more likely that the deviant results in this study are caused by the strange choice made in the phrasing of the statement about conditional probability rather than that subjects in every other experiment testing conditional probability and probabilities of conditionals were misunderstanding the task asked of them. Further, the results in a previous iteration of this experiment where subjects were asked to reason aloud actually supports the claim that subjects interpret the probabilities of conditionals as conditional probabilities. Johnson-Laird and Girotto describe subject interpretations of the sentences on probabilities of conditionals, “despite our attempts to elicit judgements of the probability of the conditional *as a whole,* four of the participants assumed that the antecedent of the conditional was true, and accordingly made an estimate of the conditional probability of the consequent in this circumstance.”[179]

---

[179] Girotto, V., and Johnson-Laird, P. N., (2002), p. 218.

Thus, without further experimental support, this particular task does not produce a significant challenge to the claim that subjects tend to give the same results for conditional probabilities and the probabilities of conditionals. In other words, neither Stalnaker's hypothesis nor Adams's thesis have been proven descriptively inaccurate by this particular study.

Douven and Verbrugge

A more significant challenge to a probabilistic theory of conditionals comes from Douven and Verbrugge. Douven and Verbrugge publish the results of studies designed to show whether Adams's thesis is descriptively accurate for a range of different types of indicative conditional. If Adams's thesis is shown not to be descriptively accurate, then one of the motivations for retaining a probabilistic account of conditionals would be eliminated. However, since Adams's thesis is a claim about the assertibility of a conditional and not the truth of a conditional, evidence against Adams's thesis does not count as evidence against probabilistic truth values for conditionals. For example, Stalnaker's hypothesis, as it is based on truth and not assertibility, would still be descriptively accurate. Douven and Verbrugge's conclusion is that only a very loose form of Adams's thesis (that the acceptability[180] of a conditional correlates or moderately correlates with the probability of the consequent given the antecedent) holds for the three types of indicative conditional that they analysed. Only the deductive form of conditional, and not the abductive or inductive forms, showed a correspondence to a

[180] In previous discussions of Adams's thesis I have defined it in terms of assertibility. Assertibility is appropriateness of utterance and acceptability is reasonableness of belief (Douven and Verbrugge define it as justified believability).

strong form of Adams's thesis. Douven and Verbrugge claim that their experiment "lend[s] further support to the thesis that acceptability is not (fully) determined by probability"[181] This contrasts the results of some previous studies[182] that claim to have found a high correlation between the probabilities assigned to conditionals and conditional probability.

Douven and Verbrugge's experiments appear to show a disconnect between acceptability and conditional probability, I will explain the type of challenge these results are taken to place on Adams's thesis. I conclude by explaining why a probabilistically based theory of human reasoning is promising, even in light of the results from Douven and Verbrugge's studies. Even if a strong version of Adams's thesis, that conditional probability is equal to the acceptability of a conditional, does not seem to hold for many types of indicative conditional, the claim that conditional probabilities are equal to the probabilities of conditionals is still possible. It should also be noted that Douven and Verbrugge do not examine subject responses to subjunctive conditionals in their study and this type of conditional is, according to the intuitions of most logicians, the most likely to be evaluated according to Adams's thesis.

Douven and Verbrugge's study examines four different versions of Adams's thesis which they abbreviate and define in the following ways:

WAT1: The acceptability of 'P → Q' is the same as the probability of (Q/P)

---

[181] Douven, I., and Verbrugge S., (2010), p. 6.

[182]For example, Evans and Over (2003), Evans, Handley and Over (2003), Over and Evans (2010).

WAT2: If the acceptability of ‘P → Q’ is high/middling/low, the probability of (Q/P) will also be high/middling/low.

WAT3: The acceptability of ‘P → Q’ highly correlates with the probability of (Q/P)

WAT4: The acceptability of ‘P → Q’ moderately correlates with the probability of (Q/P)

They also divide conditionals into those that are inferential “If you live in London, you like the rain,” and those that are contextual “If you turn your key in the ignition, then your car will start” using only conditionals of the inferential type for their study. This choice is also notable as it is contextual conditionals that seem most likely to be interpreted according to a strong version of Adams’s thesis.[183] They further divide conditionals into three types— deductive, inductive, and abductive. The deductive conditionals are constructed so that the conclusion is made certain by the addition of the premises. The abductive conditionals are constructed so that the conclusion is made more likely by the addition of explanatory considerations and the inductive conditional is constructed so that the conclusion is made more likely on the bases of statistical considerations.

Subjects were given a booklet with thirty different conditionals in a mix of the above defined types. Half of the subjects were asked to give the probability of the conditional and half were asked to give the acceptability of the conditional. Douven and Verbrugge mention that they verified that subjects understood acceptability in the

---

[183] I take this to be intuitively true but also the work of Cheng and Holyoak (Cheng, P. W., and Holyoak, K. J. (1985)) shows that subjects are more likely to interpret inferential conditionals in a straightforwardly truth-functional way.

epistemic sense (rather than as social acceptability) but they fail to give a detailed account of how this understanding was verified.

An example of the format of the questions given is:

> Context: According to a recent report written on the authority of the Dutch government, many primary school students in the province of Friesland (where many people still mainly speak Frisian) have difficulty with spelling. Jitske is a student of a primary school somewhere in the Netherlands.
>
> Conditional: If Jitske goes to a Frisian primary school, then she has difficulty with spelling.

Douven and Verbrugge's study showed that WAT1 holds for deductive conditionals; WAT3 holds for abductive conditionals; and WAT4 holds for inductive conditionals. According to Douven and Verbrugge, these results show that Adams's Thesis is only descriptively accurate in the weakest sense. They claim that, conditional probability cannot equal assertibility as the two are only moderately correlated for the three types of indicative conditional studied (and strongly correlated only in the case of deductive conditionals). The studies outlined by Douven and Verbrugge in *The Adams Family* provide a much more detailed examination of subjects' analyses of conditional probability than has been seen before. They claim that Adams's thesis, "yields the wrong predictions of people's judgements of the acceptability of important subclasses of the class of so-called inferential conditionals."[184] In other words, certain subclasses of

---

[184]Douven, I., and Verbrugge S., (2010), p. 302.

inferential conditionals (abductive and inductive) do not show a strong correlation between the conditional probability and subjects' assessments of acceptability. According to Verbrugge and Douven, “the results suggest a family of theses, each pertaining to a different type of conditional, about how conditionals relate to the relevant conditional probabilities.”[185] For example, only in the case of deductive indicative conditionals were the assessments of conditional probability close enough to the assessments of acceptability that they can reasonably be said to be the same.

The upshot of this study for Douven and Verbrugge is that, though the experiments showed “that the acceptability of inferential conditionals highly correlates with their corresponding conditional probabilities,”[186] different versions of Adams's Thesis hold for different types of conditionals. A straightforward version of Adams's Thesis, i.e., the “strong version” P(if P, then Q) = P(Q/P), does not offer a descriptively accurate account of conditional reasoning except in limited cases.

The results of Douven and Verbrugge's studies show that Adams's thesis is not nearly as descriptively accurate as was once believed. This does not however, discount a probabilistic theory of conditional reasoning. Firstly, because there is still a correlation between the conditional probability and acceptability of a conditional, even if the acceptability and conditional probability are not equivalent. This indicates that conditional probability has some effect on the acceptability of a conditional even if it does not seem to be the sole determiner of it. These results can actually support the claim that the truth of a conditional is determined probabilistically since we expect some connection between the truth and acceptability of a conditional for subjects. Secondly,

---

[185]Ibid., p. 302.

[186]Ibid., p. 312.

Douven and Verbrugge's studies do not analyze subjunctive conditionals for which the strong version of Adams's thesis intuitively seems more likely to hold than for the indicative forms. Thirdly, this study used only inferential conditionals and not contextual conditionals that are far more likely to be interpreted probabilistically. Finally, many more studies need to be done in order to flesh out this complicated picture of conditional reasoning. Not only are there relevant alternative types of conditional that Douven and Verbrugge's studies do not cover, there is also the need to see whether the results obtained by Douven and Verbrugge can be replicated across studies.[187] Also, there is good reason to suppose that, for the purposes of conditional reasoning, additional classes of conditionals may need to be recognized–- for example, social contract or permission schema conditionals (see Cheng and Holyoak).[188] Therefore, Douven and Verbrugge's study provides some reason to question Adams's thesis though this evidence is not as strong as it could be (for example, if the results were replicated and if the conditionals most likely to be interpreted according to Adams's thesis weren't removed from the study). If Adams's thesis is shown to be descriptively inaccurate, then this would have implications for the type of probabilistic account of conditionals we should adopt. It would mean that any probabilistic theory of conditionals should be based on truth rather than assertibility (such as Stalnaker's account for example).

Johnson-Laird and Byrne, and Braine and O'Brien do not offer a convincing case against a probabilistic interpretation of conditionals. Braine and O'Brien rely upon the claim that possible worlds cannot provide a useful psychological function, rejecting the

---

[187]Schroyens indicates that these results are replicated in a study done by Schroyens et al. but the results have not yet been published.

[188] Cheng, P. W., and Holyoak, K. J. (1985).

notion that Stalnaker's account is consistent with the mental models theory. Johnson-Laird and Byrne rely on the results of a study designed to prove that subjects do not reason probabilistically when given conditional statements. This study fails to convincingly test subject assessments of conditional probability due to the strange formulation in the question given to the subjects. Douven and Verbrugge's studies provide some relatively weak evidence that Adams's thesis is descriptively inaccurate, but they do not show that a probabilistic theory based on truth is likewise descriptively inaccurate. The broader conclusion of these studies for philosophers is that a descriptively accurate account of conditionals must be, in some sense, probabilistic. This is because, though Douven and Verbrugge have shown that a strong version of Adams's thesis is not consistent with the data from their studies, they have still shown a high correlation between subjects' judgements of conditional probability and assertibility for a range of different types of conditional statements.

The previous two chapters have shown that there is overwhelming experimental evidence in support of the claim that naive reasoners primarily use probabilistic conditionals. Also, as has been seen in this chapter, the evidence challenging the descriptive accuracy of a probabilistic account is limited and relatively weak. I conclude that the logical account most likely underpinning the conditional use of naive reasoners is probabilistic. This evidence, combined with evidence in the form of expert intuitions provided in chapter one, provides compelling support for the claim that Stalnaker's account, or similar probabilistic account, of conditionals is the most descriptively accurate logical account.

Conclusion

I have argued that a probabilistic account of conditionals (specifically Stalnaker's account or alternate probabilistic account with the essential features of Stalnaker's account) is able to describe the aggregate of competent uses of the conditional in natural language better than any other logical account. I conclude that Stalnaker's account of conditionals is the most descriptively accurate logical account. In chapter one I outlined expert intuitions that conflict with the standard truth functional account of conditionals in the form of the paradoxes of the material conditional. Some of the counter-intuitive results of the truth table for the material conditional are that every conditional with a false antecedent is true and every conditional with a true consequent is true. I showed that C. I. Lewis's attempt to resolve the paradoxes of the material conditional by developing an account of strict implication fails as this account simply results in strengthened paradoxes that arise from contradictions or logical truths. For example, in Lewis's account every conditional with an antecedent that is a contradiction is true and every conditional with a consequent that is a tautology is true. I also argued that connection accounts fail to satisfy expert intuitions about conditional use as they reject some inferences that are intuitively valid and accept some inferences that are intuitively invalid.

In chapter two I outlined an attempt to save the standard truth functional account of conditionals by explaining the paradoxes of the material conditional in terms of assertibility, i.e., the claim that the conditionals outlined in the paradoxes of the material conditional are true but not assertible. I argued that Quine's account of assertibility fails as it cannot accommodate the use of counterfactuals in natural language. I then

described Jacksons's account of conditionals whereby conditionals have both truth conditions and conditions of assertibility. Indicative conditionals in this account are able to retain the truth table for the material conditional and the counter-intuitive aspects of this truth table are explained away in terms of assertibility. Jackson's account based on assertibility is much more successful than Quine's but ultimately also fails to capture natural language use of conditionals as the intuitions that conflict with the material conditional for naive reasoners are far more widespread than Jackson claims. Jackson attempts to argue that these intuitions must be mistaken since David Lewis's triviality proof showed that probabilistic conditionals result in triviality or contradiction. I showed that the triviality proof does not lead to such a strong conclusion and that Jackson's account of conditionals fails to adequately describe natural language use of conditionals.

In chapter three I outlined the probabilistic, possible worlds account of conditionals that Stalnaker developed in 1968. Stalnaker's account is loosely based on Kripke's modal logic and relies upon possible world semantics. An absurd world that allows for contradictions, and a selection function that determines the ordering of possible worlds, are added to Kripke's modal logic. The selection function restricts the ordering of possible worlds in the following ways: the antecedent must be true in the selected world; if the antecedent is true in the actual world, then the actual world must be selected (the world selected must differ as minimally as possible from the actual world); the absurd world must only be selected if it is required in order to make the antecedent true; and ordering of possible worlds must remain consistent. Jackson argues that the unified account offered by Stalnaker is flawed as only subjunctive

conditionals should be analyzed probabilistically. This argument fails to successfully challenge Stalnaker's account as we don't typically consider differences in grammatical tense to require different logical analyses and Jackson provides no non ad hoc reason why conditionals should be different. David Lewis argues that Stalnaker's selection function leads to semantic indeterminacy. I defend Stalnaker's account against criticisms from David Lewis by arguing that the small degree of semantic indeterminacy in Stalnaker's account is worth the greater degree of descriptive accuracy.

In chapter four I outline some empirical support for a probabilistic psychological account of conditionals by arguing that the use of conditionals by naive reasoners is primarily probabilistic. My purpose in this chapter is to get a better picture of the phenomena— natural language use of conditionals— by including assessments of conditional reasoning by naive reasoners. The first part of this chapter is an examination of how a probabilistic psychological account is able to explain the systematic responses on reasoning tasks with conditionals. A probabilistic psychological account can explain variation among subject responses and the reason for performance errors in the suppression effect experiments. Specifically, I argued that the theory of conditionals provided by Schroyens and Schaeken is the most descriptively accurate psychological account. I then argued that Stalnaker's account of conditionals is the logical theory most compatible with these probabilistic theories of conditional use from psychologists (for example, SSCEPPTRE). Stalnaker's account shares essential characteristics with the psychological account of conditionals presented by Schroyens and Schaeken. The ability of the alternative mental models account to reliably predict responses on the conditional reasoning task and define what were previously considered performance

errors as rational belief testing mechanisms, supports the descriptive accuracy claims of a logical account of conditionals that shares a basic structure with this psychological account.

In chapter five I outline some experimental challenges to a probabilistic account. I argue that Brain and O'Brien's challenge to Stalnaker's account fails as it merely shows that Stalnaker's logical account would make an incomplete psychological account of conditionals, not that Stalnaker's hypothesis is descriptively inaccurate. I outline an experiment that Johnson-Laird and Byrne take as evidence against Adams's thesis and Stalnaker's hypothesis but conclude that, due to the formulation of the question in this task, the experiment does not in fact test assessments of conditional probability. Finally, I examined experiments that Douven and Verbrugge present against Adams's thesis. These arguments present a moderate challenge to the claim that Adams's thesis is descriptively accurate but do not present any challenge to the claim that Stalnaker's hypothesis is descriptively accurate.

I have presented evidence from both expert reasoners in the form of philosophical arguments and naive reasoners in the form of psychological experiments in human reasoning that support a probabilistic interpretation of conditionals. The experimental evidence that naive subjects are operating with a probabilistic conditional in mind is overwhelming and there is almost no evidence to be found that counters this view. A probabilistic account of conditionals (such as that provided by Stalnaker in 1968) is able to model the vast majority of competent uses of the conditional. Therefore, such an account is the most descriptively accurate logical account of conditionals, i.e., is the

account that most aptly models conditional use in natural language and provides the best representational model of the data of conditional use in natural language.

Sources Cited

Adams, E. (1975). *The Logic of Conditionals*. Dordrecht: Reidel.

___. (1964). On the Reasonableness of Inferences Involving Conditionals, *Proceedings of the 13th International Congress of Philosophy*, Vol. 5, Mexico City, pp, 1 - 9.

___. (1965). A Logic of Conditionals, *Inquiry*, Vol. 8, pp. 166 -197.

___. (1966). Probability and the Logic of Conditionals, in Hintikka, J., and Suppes. P., (eds.), *Aspects of Inductive Logic*, Amsterdam: North Holland, pp. 256–316.

___. (1968). Probability and the Logic of Conditionals, in *Aspects of Inductive Logic*, J. Hintikka, J., and Suppes, P., (eds.), *Aspects of Inductive Logic*, Amsterdam: North Holland, pp. 265 - 316.

___. (1970). Subjunctive and Indicative Conditionals, *Foundations of Language,* Vol. 6, pp. 89 - 94.

___. (1976). Prior Probabilities and Counterfactual Conditionals, in *Foundations and Philosophy of Epistemic Applications of Probability Theory,* Harper, W., and Hooker, C. (eds), The University of Western Ontario Series in Philosophy of Science, Volume 6a, pp 1-21.

___. (1987). On the Meaning of the Conditional, *Philosophical Topics*, Vol. 15, pp. 5 – 22.

Anderson, A.R. (1951). A Note on Subjunctive and Counterfactual Conditionals, *Analysis*, Vol. 12, pp. 35 - 38.

Anderson, A. R. and N. D. Belnap, Jr. (1975). *Entailment: The Logic of Relevance and Necessity, Vol. 1*, Princeton, Princeton University Press.

Anderson, A. R., Belnap N. D. Jr., and Dunn, J. M. (1992). *Entailment: The Logic of Relevance and Necessity (Vol. 2)*, Princeton, Princeton University Press.

Austin, J. L. (1962). *How to Do Things With Words*. Cambridge: Harvard University Press.

Baddeley, H. (1990). *Human memory: Theory and practice*. Hove: Lawrence Erlbaum and Associates.

Bazerman, M., and Neale. M. (1986). Heuristics in Negotiation, in *Judgement and Decision Making: An Interdisciplinary Reader,* Arkes and Hammond (eds), Cambridge: Cambridge University Press, pp. 311 – 321.

Beall, and van Fraasen. (2003). *Possibilities and Paradox*. Oxford: Oxford University Press.

Bennett, J. (2003). *A Philosophical Guide to Conditionals*, Oxford University Press.

Bermudez J. L. (1999). Rationality and the Backwards Induction Argument, *Analysis,* Vol. 59: pp. 243 - 248.

___. (2000). Rationality, Logic and Fast and Frugal Heuristics: Review of Gigerenzer et al, *Behavioral and Brain Sciences,* Vol. 23, pp. 744 - 745.

___. (2006). Animal Reasoning and Proto-logic, in *Rational Animals*, Hurley, S. L., and Nudds, M. (eds), Oxford: Oxford University Press.

___. (2007). Language and Thought, in *Handbook to the Philosophy of Psychology*, Prinz. J. (ed), Oxford: Oxford University Press.

Bermudez J. L., and Millar. A. (eds.) (2002). *Reason and Nature: Essays in the Theory of Rationality*, Oxford University Press.

Birnbaum, A. (1962). On the Foundations of Statistical Inference. *Journal of American*

Statistical Association, Vol. 57, No. 298, pp. 269 – 326.

Blackburn, S. (1986). How Can We Tell Whether a Commitment has a Truth Condition?, in *Meaning and Interpretation*, Travis C. B. (ed.), Blackwell, pp. 201 – 232.

Braine M.D.S. (1978). On the Relation Between the Natural Logic of Reasoning and Standard Logic, *Psychological Review* Vol. 85, pp. 1 - 21.

___. (1990). The Natural Logic Approach to Reasoning, in *Reasoning, Necessity, and Logic: Developmental Perspectives*, Overton. W. F. (ed.), Hillsdale: Erlbaum.

Braine M. D. S. and O'Brien, D. (1991). A theory of If: Lexical Entry, Reasoning Program and Pragmatic Principles. *Psychological Review*, Vol. 98, pp. 182 - 203.

Braine M. D. S., Reiser, B. J., and Rumain, B., (1984). Some Empirical Justification for a Theory of Natural Propositional Logic, in *The Psychology of Learning and Motivation: Advances in Research and Theory*, Vol. 18, Bower, B. (ed.), New York: Academic Press.

Braine, M. D. S., and Politzer, G. (1991). Responses to Inconsistent Premises Cannot Count as Suppression of Valid Inferences, *Cognition*, Vol. 38, pp. 103 - 108.

Brandom, R. (1981). Semantic Paradox of Material Implication. *Notre Dame Journal of Formal Logic.* Vol. 22, No. 2, pp. 129 - 132

Bower, B. (1996). Rational Mind Design: Research into the Ecology of Thought Treads on Contested Terrain, *Science News*, Vol. 150, pp. 24 - 25.

Byrne, R. M. J. (1989). Suppressing Valid Inferences with Conditionals. *Cognition*, Vol. 31, pp. 61 - 83.

Cantlon, J., and Brannon, E. (2006). Adding Up the Effects of Cultural Experiences on the Brain, *Trends in Cognitive Sciences*, Vol. 11, No. 1, pp. 1 - 4.

Carnap, R. (1946). Modalities and Quantification. *The Journal of Symbolic Logic*, Vol. 11, pp. 33 – 64.

Carroll, L. (1895). What the Tortoise Said to Achilles. *Mind*, Vol. 4, pp. 278 – 80.

Carruthers, P. (1996). *Language, Thought, and Consciousness*. Cambridge: Cambridge University Press,

Carruthers, P, and Boucher, J. (eds.), (1998). *Language and Thought*. Cambridge: Cambridge University Press.

Cheng, P. W., Holyoak, K. J. (1985). Pragmatic Reasoning Schemas. *Cognitive Psychology*, Vol. 17, pp. 391 - 416.

Cheng, P. W., Holyoak, K. J., Nisbett, R. E., and Oliver, L. M. (1986). Pragmatic Versus Syntactic Approaches to Training Deductive Reasoning. *Cognitive Psychology*, Vol. 18, pp. 293 - 328

Chomsky, N. (1965). Aspects of the Theory of Syntax. Cambridge: MIT Press

Cohen, L. J. (1971). Some Remarks on Grice's Views about the Logic Particles of Natural Language, in *Pragmatics of Natural Languages*, Bar-Hillel, I. (ed.), Dordrecht: Reidel, pp. 50 - 68.

Cosmides, L. (1989). The Logic of Social Exchange: Has Natural Selection Shaped how Humans Reason? Studies with the Wason Selection Task, *Cognition*, Vol. 31, pp. 187 - 276.

Cosmides, L., and Tooby, J. (1996). Are Humans Good Intuitive Statisticians After all? Rethinking Some Conclusions From the Literature on Judgement Under Uncertainty, *Cognition,* Vol. 58, pp. 1 - 73.

Cummins, D. D., Lubart, T., Alksnis, O., and Rist, R. (1991). Conditional Reasoning and

Causation. *Memory and Cognition*, Vol. 19, pp. 274 - 282.

De Neys, W. (2006). Dual Processing in Reasoning, *Psychological Science* Vol. 17 No. 5, pp. 428 - 433.

Douven, I., and Verbrugge, S. (2010). The Adams Family. *Cognition*. Vol. 117.

Dugan, C. M., and Revlin, R. (1990). Response Options and Presentation Format as Contributors to Conditional Reasoning. *Quarterly Journal of Experimental Psychology*, Vol. 42A, pp. 829 - 848.

Dummett, M. (1973). *Frege*, Duckworth: London.

Dunn, J. M. (1986). Relevance Logic and Entailment, in *Handbook of Philosophical Logic*, Vol. 3, Guenthner, F., and Gabbay, D. (eds), Dordrecht: Reidel, pp. 117 – 24.

Edgington, D. (1995). On Conditionals. *Mind,* Vol. 104, No. 414, pp. 235 - 329.

___. (1997b) Commentary, in *Conditionals*, Edgington, D., Woods, M. and Wiggins, D, (eds). Clarendon Press.

Edgington, D., Woods, M. and Wiggins, D. (1997a). *Conditionals*, Clarendon Press.

Ericsson K., A., and Herbert A. S., (1980). Verbal Reports as Data, *Psychological Review,* Vol. 8, pp. 215 - 251.

Elio, R. (ed.) (2003). *Common Sense, Reasoning, and Rationality*. Oxford University Press.

Ellis, B. (1984). Two Theories of Indicative Conditionals, *Australasian Journal of Philosophy*, Vol. 62, No.1.

Engel, P. (1991). *The Norm of Truth*. University of Toronto Press.

Evans J. St. B. T.(2003). In Two Minds: Dual Process Accounts of Reasoning, *Trends in Cognitive Sciences* Vol 7, pp. 454 - 459.

___. (1993). Bias and Rationality, in *Rationality*, Manktelow, K. I., and Over, D. E. (eds), London: Routledge.

___. (1992). Reasoning and Relevance in the Selection Task. Unpublished Manuscript, University of Plymouth.

___. (1977). Linguistic Factors in Reasoning. *Quarterly Journal of Experimental Psychology*, Vol. 29, pp. 297 - 306.

___. (1972). Interpretation and Matching Bias in a Reasoning Task. *British Journal of Psychology*, Vol. 24, pp. 193 – 199.

Evans, J. St. B. T., Handley, S. H., and Over, D. E. (2003). Conditionals and Conditional Probability. *Journal of Experimental Psychology: Learning, Memory and Cognition*, Vol. 29, pp. 321-335.

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

Evans, J. St. B. T., Newstead, S. E. (1977). Language and Reasoning: A Study of Temporal Factors. *Cognition*, Vol. 8, pp. 265 - 283.

Evans, J. St. B. T., and Brooks, P. G. (1981). Competing with Reasoning: A Test of the Working Memory Hypothesis. *Current Psychological Research*, Vol. 1, pp. 139 - 147.

Evans, J. St. B. T., and Lynch J. S. (1973). Matching Bias in the Selection Task. *British Journal of Psychology*, Vol. 64, pp. 391 - 397.

Evans, J. St. B. T., and Pollard, P. (1990). Belief Bias and Problem Complexity in

Deductive Reasoning, in *Cognitive Biases*, Caverni, J. P., Fabre, J. M., and Gonzales, M. (eds), Amsterdam.

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

Fisher, J. (2007). *On the Philosophy of Logic*. Wadsworth.

Frege, G. (1884). The Foundations of Arithmetic, (Die Grundlagen der Arithmetik: eine logisch-mathematische Untersuchung über den Begriff der Zahl), Breslau: W. Koebner, reprinted in *The Frege Reader*, Beaney, M. (ed), Oxford: Blackwell, 1997. Page numbers are cited from the 1997 printing.

___. *(1879?). Logic (Logik)*, unpublished manuscript translated by Long and White in Posthumous Writings, 1979, Hermes, H., Kambartel, F., and Kaulbach, F. (eds), Chicago: University of Chicago Press, reprinted in *The Frege Reader*, Beaney, M. (ed), Oxford: Blackwell, 1997. Page numbers are cited from the 1997 reprinting.

___. *(*1879). Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens, Halle a. S.: Louis Nebert. Translated as Concept Script, a Formal Language of Pure Thought Modelled Upon That of Arithmetic, by Bauer-Mengelberg, S., in vanHeijenoort, J. (eds), From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931, Cambridge, MA: Harvard University Press, 1967.

___. (1894). Review of E. G. Husserl, Philosophie der Arithmetik I, reprinted in *Frege, Collected Papers on Mathematics, Logic, and Philosophy*, Wiley-Blackwell, 1991. pp. 195 – 209.

Frosch, C. (2011) Is Everyday Causation Deterministic or Probabilistic? *Acta*

Psychologica, Vol. 137, No. 3, pp. 280-291.

Gamut, L. T. F. (1991). *Logic, Language, and Meaning*. Vol. 1. Chicago: The University of Chicago Press.

Gavanski, I., and Hoffman, C. (1986). Assessing Influences on One's Own Judgements: Is There Greater Accuracy for Either Subjectively Important or Objectively Influential Variables?, *Social Psychology Quarterly,* Vol. 49 No. 1, pp. 33 - 44.

Gigerenzer, G. (1991). How to make Cognitive Illusions Disappear: Beyond Heuristics and Biases, *European Review of Social Psychology* Vol. 2, pp. 83 - 115.

___. (1991). On Cognitive Illusions and Rationality, *Poznan Studies in the Philosophy of the Sciences and the Humanities* Vol. 21, pp. 225 - 249.

___. (1993). The Bounded Rationality of Probabilistic Mental Models, in *Rationality: Psychological and Philosophical* Perspectives, Manktelow, K. I., and Over, D. E. (eds), London: Routledge, pp. 284 - 313

___. (1994). Why the Distinction between Single Event Probabilities and Frequencies is Important for Psychology, in *Subjective* Probability, Wright, G., and Ayton, P. (eds). New York: Wiley, pp. 129 - 161.

___. (1996). On Narrow Norms and Vague Heuristics: a Reply to Kahneman and Tversky, *Psychological Review*, Vol. 103, pp. 592 - 596.

___. (1997). The Modularity of Social Intelligence, in *Machiavellian Intelligence II,* Whiten, A., and Byrne, R. (eds), Cambridge: Cambridge University Press, pp. 264 - 280.

___. (1998). Ecological Intelligence: An Adaptation for Frequencies, in *The Evolution of Mind,* Cummins, D., and Allen, C. (eds), New York: Oxford University Press, pp. 9

- 29.

Gigerenzer G., and D. G. Goldstein (1996). Reasoning the Fast and Frugal way: Models of Bounded Rationality. *Psychological Review,* Vol. 103, pp. 650 - 669.

Gigerenzer G., and Hoffrage, U. (1995). How to Improve Bayesian Reasoning Without Instruction: Frequency Formats, *Psychological Review*, Vol. 102, pp. 684 - 704

Gigerenzer G, Hoffrage, U., and Kleinbolting, H.,(1991). Probabilistic Mental Models: A Brunswikian Theory of Confidence, *Psychological Review*, Vol. 98, pp. 506 - 528.

Girotto, V. and Johnson-Laird, P. (2004). The Probability of Conditionals, *Psychologia*, Vol. 47, pp. 207-225.

Goel, V. (2007). Anatomy of Deductive Reasoning, *Trends in Cognitive Science*, Vol. 11, No. 10, pp. 435 - 441.

Goodman, N. (1947). The Problem of Counterfactual Conditionals, *Journal of Philosophy*, Vol. 44, pp. 113 - 128.

___. (1955). *Fact, Fiction, and Forecast*, Cambridge, MA: Harvard University Press.

Grice, H. P. (1975). Logic and Conversation, *Syntax and Semantics, Vol. 3, Speech Acts*, Cole, P. and Morgan, J. L. (eds), reprinted in *Conditionals*, Jackson, F. (ed), 1991, Oxford University Press. Page numbers are cited from the 1991 reprinting.

Griggs, R. (1995). The Effects of Rule Clarification, Decision Justification, and Selection Instruction on Wason's Abstract Selection Task, in *Perspectives on Thinking and* Reasoning, Newstead, S., and Evans, J., (eds) pp. 17 - 39.

Griggs, R. and Cox, J. (1982). The Elusive Thematic Materials Effect in Wason's Selection Task. *British Journal of Psychology*, Vol. 73, pp. 407 - 420.

Haack, S. (1978). Philosophy of Logics. Cambridge: Cambridge University Press.

Haaparanta, L., and Koskinen, H. J., (eds.), (2012). Categorie*s of Being: Essays on Metaphysics and Logic*, Oxford: Oxford University Press.

Hacking, I. (1979). What is Logic?, *Journal of Philosophy*, Vol. 86, pp. 285 – 319.

Hadjichristidis, C.,, Stephenson, R. J., Over, D. E., Sloma, S. A., Evans, J. St. B. T., and Feeney, A. (2001) On the Evaluation of 'if p then q' Conditionals. *Proceedings of the 23rd Annual Meeting of the Cognitive Science Society*, Edinburgh. pp. 409 - 414 (Page numbers are taken from pdf as the file itself has no printed page numbers.)

Hájek, A. (2003). What Conditional Probability Could not Be. *Synthese,* Vol. 137, No. 3 pp. ,273 – 323.

___. (2011). Triviality Pursuit, *Topoi,* Vol. 30.

Hanna, R. (2006). *Rationality and Logic*. Cambridge, Mass.: MIT Press.

Harman, G. (1986). *Change in View: Principles of Reasoning*. Cambridge Mass.: MIT Press.

___. (1973). *Thought*. Princeton: Princeton University Press.

Harper, W. L. (1981). A Sketch of some recent Developments in the Theory of Conditionals, in *Ifs*, Harper, W. L., Stalnaker, R., and Pearce, G., (eds), pp. 1 - 143.

Harper, W. L., Stalnaker, R., and Pearce, G. (eds) (1981). Ifs. Dordrecht: Reidel.

Henle, M. (1962). On the Relation Between Logic and Thinking. *Psychological Review*, Vol. 69, pp. 366 - 378.

Hock, S. J., and Tschirgi, J. E. (1985). Logical Knowledge and Cue Redundancy in Deductive Reasoning. *Memory and Cognition*, Vol. 13, pp 453 - 462.

Jackson, F. (1987). *Conditionals*. Oxford: Blackwell.

___. (1991). (ed) *Conditionals*. Oxford: Oxford University Press.

___. (1980). Conditionals and Possibilia, *Proceedings of the Aristotelian Society*, New Series, Vol 81.

___. (1984). Two Theories of Indicative Conditionals: Reply to Brian Ellis, *Australasian Journal of Philosophy*, Vol. 62, No. 1.

___. (1979). On Assertion and Indicative Conditionals. *The Philosophic Review*, Vol. 88, No. 4.

Jager, R. (1972), *The Development of Bertrand Russell's Philosophy*, London: George Allen and Unwin.

Johnson, D. K. (1996). The Paradox of Indicative Conditionals. Philosophical Studies, Vol. 83, No. 1, pp. 93 - 112

Johnson-Laird, P. (1981). Comprehension as the Construction of Mental Models, *Philosophical Transactions of the Royal Society of London*. Series B, Biological Sciences, Vol. 295, No. 1077, pp. 353 - 374.

___. (1983). *Mental Models*. Cambridge, Mass.: Harvard University Press.

Johnson-Laird, P., and Byrne, R. (1991). *Deduction*. Hillsdale, N. J.: Lawrence Erlbaum.

___. (1992). Modal Reasoning, Models and Manktelow and Over. *Cognition***.** Vol. 43, No. 2, pp. 173-182

___. (2002). Conditionals: A Theory of Meaning, Pragmatics, and Inference. *Psychological Review* Vol. 109, No. 4.

Johnson-Laird, P. N., Byrne, R., and Shaeken, W. (1992). Propositional Reasoning by Model. *Psychological Review* Vol. 99, No. 3.

Johnson-Laird, P., and Tagart, J. (1969). How Implication is Understood, *The American Journal of Psychology*, Vol. 82, No. 3, pp. 367 - 373.

Johnson-Laird, P. N., Legrenzi, P., and Legrenzi, M. S. (1972). Reasoning and a Sense of Reality. *British Journal of Psychology*, Vol. 63, pp. 395 - 400.

Kalish, D., Montague, R., and Mar, G. (1980). *Logic*. Oxford University Press.

Kant, I., (1781). *Critique of Pure Reason*, P. Guyer, P., and Wood, A. (eds. and trans.), Cambridge: Cambridge University Press, 1997.

Kapsner, A. (2013). From Semantics to Logic: The Scenic Route, *Teorema: Revista Internacional de Filosofía*, Vol. 32, No. 1, El legado de Dummett/ Dummett's Legacy, pp. 77 - 94.

Kemeny, J., Snell, L., and Thompson, G. (1957). Introduction to Finite Mathematics (3rd ed.). Prentice Hall

Kern, L. H., Mirels, H. L., Hinshaw, V. G. (1983). Scientists' Understanding of Propositional Logic: An Experimental Investigation. *Social Studies of Science*, Vol. 13, pp. 131 - 146.

Kraut, R. E., and Lewis, S. H. (1982). Person Perception and Self-Awareness: Knowledge of Influences on One's Own Judgments, *Journal of Personality and Social Psychology* Vol. 42, pp. 448 - 460.

Kolmogorov, A. N. (1933). Grundbegriffe der Wahrscheinlichkeitrechnung, Ergebnisse Der Mathematik; translated as Foundations of Probability, New York: Chelsea Publishing Company, 1950.

Kripke, S. A. (1959a). A Completeness Theorem in Modal Logic, *Journal of Symbolic Logic*, Vol. 24, pp. 1 – 14.

___. (1959b). Semantical Analysis of Modal Logic, *Journal of Symbolic Logic*, Vol. 24, pp. 323 – 324.

Kusch, M. (2007). “Psychologism” *The Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/entries/psychologism. Subsection “Recent Re-evaluations”.

Lairde, D. A. (1932). How the Consumer Estimates Quality by Subconscious Sensory Impressions, *Journal of Applied Psychology* Vol. 16, pp. 241 - 246.

Lewis, C. I. (1912). Implication and the Algebra of Logic, *Mind*, Vol. 21, No. 84.

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

___. (1973). Counterfactuals and Comparative Possibility, *Journal of Philosophical Logic*, Vol. 2, No. 2. Reprinted in Harper et al. (eds.) *Ifs,* (1981) D. Riedel. Page numbers from the 1981 reprinting.

___. (1976). Probabilities of Conditionals and Conditional Probabilities, *Philosophical Review*, Vol. 85, No 3, pp 297 - 315.

Manktelow, K. I. and Evans, J. St. B. T. (1979). Facilitation of Reasoning by Realism: Effect or Non-Effect? *British Journal of Psychology*,, Vol. 70, pp. 477 - 488.

___. (1991). Social roles and Utilities in Reasoning with Deontic Conditionals. *Cognition*, Vol. 39, pp. 85 - 105.

Marcus, S. L. and Rips, L. J. (1979). Conditional Reasoning. *Journal of Verbal Learning and Verbal Behavior*, Vol. 18, pp. 199 - 233

Markovits, H. (1984). Awareness of The 'Possible' as a Mediator of Formal Thinking in Conditional Reasoning Problems. *British Journal of Psychology*, Vol. 75, pp. 367 - 376.

___. (1988). Conditional Reasoning, Representation, Empirical Evidence on a Concrete

Task. *Quarterly Journal of Experimental Psychology*, Vol. 40A, pp. 483 - 495.

___. (2013). *The Developmental Psychology of Reasoning and Decision-Making*. Psychology Press.

Nisbett, R., E., and Bellows, N. (1977). Verbal Reports about Causal Influences on Social Judgements: Private Access Versus Public Theories *Journal of Personality and Social Psychology*, Vol. 35: pp. 613 - 624.

Nisbett, R. E., and Wilson, T. D. (1977). Telling More than We Can Know: Verbal Reports on Mental Processes, *Psychological Review*, Vol. 84: pp. 231 - 259.

Oaksford, M, and Chater, N. (2007). *Bayesian Rationality: The Probabilistic Approach to Human Reasoning*. Oxford: Oxford University Press.

___. (2002). Commonsense Reasoning, Logic and Human Rationality, in *Common Sense, Reasoning, and Rationality*, Elio, R. (ed.), Oxford University Press.

___. (2009). Precis of Bayesian Rationality: The Probabilistic Approach to Human Reasoning. *Behavioral and Brain Sciences*, Vol. 32.

___. (2003). Conditional Probability and the Cognitive Science of Conditional Reasoning. *Mind and Language*, Vol 18, No 4.

Oaksford, M., Chater N., and Larkin J., (2000). Probabilities and Polarity Biases in Conditional Inference, *Journal of Experimental Psychology*, Vol. 26 No. 4, pp.883 - 99.

Oberauer, K. (2006). Reasoning with Conditionals: a Test of Formal Models of Four Theories, *Cognitive Psychology*, Vol. 53: pp. 238 - 283.

Oberaber, K., and Wilhelm, W., (2003) The Meaning(s) of Conditionals: Conditional Probability, Mental Models, and Personal Utilities. *Journal of Experimental*

Psychology: Learning, Memory and Cognition, Vol. 29 pp. 680-693

O' Brien, D., Braine, M., and Yang Y. (1994). Propositional Reasoning by Mental Models? Simple to Refute in Principle and Practice. *Psychological Review*, Vol 101, No 4.

Osherson, D. (1975). Logic and Models of Logical Thinking, in *Reasoning: Representation and Process in Children and Adults*, Falmagne, R. J. (ed), Hillsdale: Lawrence Erlbaum and Associates.

Pelletier, J., Elio, R., and Hanson, P. (2008). Is Logic all in Our Heads? *Studia Logica* ,Vol. 86, pp. 1 – 86.

Piaget, J. (1957). *Logic and Psychology*. New York: Basic Books.

Politzer, G., and Braine, M. (1991). Responses to Inconsistent Premisses Cannot Count as Suppression of Valid Inferences. *Cognition*, Vol. 38, pp 103 - 108.

Pollard, P., and Evans, J, St. B. T. (1980). The Influence of Logic on Conditional Reasoning Performance. *Quarterly Journal of Experimental Psychology*, Vol. 32, pp. 605 - 624.

___. (1981). The Effect of Prior Beliefs in Reasoning: An Associational Interpretation. *British Journal of Psychology*, Vol. 72, pp. 73 - 82.

Quine, W. O. (1962). *Methods of Logic*. London: Routledge & Kegan Paul.

___. (1953). *From a Logical Point of View*. London: Harper & Row.

___. (1960). *Word and Object*. Cambridge, Mass.: MIT Press, 2015.

___. (1986). *Philosophy of Logic*, 2nd. ed. Cambridge, Mass.: Harvard University Press.

Ramsey, F.P. (1929). General Propositions and Causality, in *F. Ramsey: Philosophical Papers*, ed. H. A. Mellor, Cambridge: Cambridge University Press, 1990. pp. 145 -

163.

___. (1930). On a Problem in Formal Logic, *Proceedings of the London Mathematical Society,* Vol. 30.

___. (1931). *The Foundations of Mathematics and Other Essays*. Braithwaite, R. B. (ed.) London: Routledge & Kegan Paul.

Rawls, J. (1971). *A Theory of Justice*, 2nd Edition 1999, Cambridge, MA: Harvard University Press.

Resnik, M. (1985). Logic: Normative or Descriptive? The Ethics of Belief or a Branch of Psychology? *Philosophy of Science*, Vol. 52, pp. 221 - 238.

Rips, L. J. (1989). The Psychology of Knights and Knaves, *Cognition*, Vol. 31, pp. 85 - 116.

___ (1990a). Reasoning, *The Annual Review of Psychology,* Vol. 41, pp. 321- 353.

___ (1990b). Paralogical Reasoning: Evans, Johnson-Laird and Byrne on Liar and Truth-teller Puzzles. *Cognition*, Vol. 36, pp. 291 - 314.

___ (1983). Cognitive Processes in Propositional Reasoning. *Psychological Review*, Vol. 90

___ (1986). Mental Muddles, in *The Representation of Knowledge and Belief*, Brand, M., and Harnish, R. M., (eds). Tucson: University of Arizona Press. pp. 258 - 286.

___ (1989). The Psychology of Knights and Knaves. *Cognition*, Vol. 31, pp. 85 - 116.

Rips, L. J., and Conrad, F. J. (1983). Individual Differences in Deduction. *Cognition and Brain Theory*, Vol. 6, pp. 259 - 289.

Rumain, B. Connell, J., and Braine, M. D. S. (1983). Conversational Comprehension Processes are Responsible for Reasoning Fallacies in Children as Well as

Adults. *Developmental Psychology*, Vol. 19, pp. 471 - 481.

Rumain et al. (1984). Some Empirical Justification for a Theory of Natural Propositional Logic, in *The Psychology of Learning and Motivation,* Bower (ed). New York: Academic Press.

Russell, B. (1956). *Logic and Knowledge*. R. C. Marsh (ed.). London: George Allen & Unwin.

Russell, B and Whitehead, A. N. (1910, 1912, 1913). *Principia Mathematica*, 3 vols., Cambridge: Cambridge University Press; 2nd ed., 1925 (Vol. 1), 1927 (Vols. 2, 3), Cambridge: Cambridge University Press, 1962.

Samuels R., Stich, S., and Tremolet, P. (1999). Rethinking Rationality: From Bleak Implications to Darwinian Modules, in *What is Cognitive Science?* Lepore, E. and Pylyshyn, Z (eds). Oxford: Blackwell.

Samuels R., and Stich, S. (2001). Irrationality: Philosophical Perspectives in *The International Encyclopedia of the Social and Behavioral Sciences,* Smelser, N. and Baltes, P. (eds). Science Direct

Samuels R., Stich S., and Bishop M. (2002). Ending the Rationality Wars: How to Make Disputes about Human Rationality Disappear, in *Common Sense Reasoning, and Rationality,* Elio, R. (ed.), New York: Oxford University Press. pp. 236 - 268.

Samuels R., and Stich S. (2002). Rationality, in *The Encyclopedia of Cognitive Science*, Vol. 3, pp. 830 - 837.

___. (2003). Rationality and Psychology, in *The Oxford Handbook of Rationality ,* Mele, A. (ed). New York: OUP. pp. 279 - 300.

Searle, J. (1969). *Speech Acts*, Cambridge University Press.

Sharpe, D., Cote, M. H., and Eaken, L. (1999). Reasoning about a Structured Object: Three- and- Four-Year-Olds' Grasp of a Borderline Case and an Unexcluded Middle, *Child Development*, Vol. 70, No. 4, pp. 866 - 871.

Schroyens, W. J., and Schaeken, W. (2003). A Critique of Oaksford, Chater, and Larkin's (2000) Conditional Probability Model of Conditional Reasoning. *Journal of Experimental Psychology: Learning Memory and Cognition*, Vol. 29, pp. 140-149

___. (2008). Deductive Rationality in Validating and Testing Conditional Inferences. *Canadian Journal of Experimental Psychology*. Vol. 62, No. 3, pp. 163 - 173.

Schroyens, W., Schaeken, W., and Dieussaert (2008). Issues in Reasoning about Iffy Propositions: "The" Interpretation(s) of Conditionals. *Experimental Psychology*. Vol. 55, No.2, pp. 113 - 120.

Smith, E.E., Langston, C. and Nisbett, R. (1992). The Case for Rules in Reasoning. *Cognitive Science*, Vol. 16, pp. 1 - 40.

Smith E. R., and Miller, F. (1978). Limits on Perception of Cognitive Processes: a Reply to Nisbett and Wilson, *Psychological Review*, Vol 85, pp. 355-362.

Stalnaker, R. (1968). A Theory of Conditionals in *Studies in Logical Theory, American Philosophical Quarterly* (Monograph Series, 2), Oxford: Blackwell, pp. 98 – 112. Reprinted in *Ifs* (1981), Harper, W.L., Stalnaker, R. and Pearce, G. (eds). Boston: D. Reidel Publishing Company. Page numbers from the 1981 reprinting.

___. (1981a). Probability and Conditionals, in *Ifs,* Harper, W.L., Stalnaker, R. and Pearce, G. (eds). Boston: D. Riedel Publishing Company

___. (1981b). A Defense of Conditional Excluded Middle, in *Ifs,* Harper, W.L., Stalnaker,

R. and Pearce, G. (eds). Boston: D. Riedel Publishing Company. pp. 87 - 104.

Stalnaker, R., and Thomason, R. (1970). A Semantic Analysis of Conditional Logic, *Theoria*. Vol. 36, pp. 23 - 42.

Staudenmayer, H. (1975). Understanding Conditional Reasoning with Meaningful Propositions, in *Human Reasoning* edited by R. J. Falmagne. New York: John Wiley.

Stevenson, R. J., and Over, D. E., (1995). Deduction from Uncertain Premises, *The Quarterly Journal of Experimental Psychology*, Vol. 48A, pp. 613-643

Stich, S. (1990). *The Fragmentation of Reason: Preface to a Pragmatic Theory of Cognitive Evaluation*, MIT Press.

Sugiyama, L., Tooby, J., and Cosmides, L. (2002). Cross-Cultural Evidence of Cognitive Adaptations for Social Exchange among the Shiwiar of Ecuadorian Amazonia, *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 99, No. 17, pp. 11537 - 11542.

Taplin, J. E. (1971). Reasoning with Conditional Sentences, *Journal of Verbal Learning and Verbal Behavior*, Vol. 10, pp. 219 - 225

Taplin, J. E., Staudenmayer, H. (1973). Interpretation of Abstract Conditional Sentences in Deductive Reasoning, *Journal of Verbal Learning and Verbal Behavior*, Vol. 12, pp. 530 - 542.

Thagard, P. (1982). From the Descriptive to the Normative in Psychology and Logic, *Philosophy of Science*, Vol. 49, pp. 24 - 42.

Thompson V. (1994). Interpretational Factors in Conditional Reasoning, *Memory and Cognition* Vol, 22, pp. 742 - 758.

Tourangeau R., Rips, L. J., and Rasinski, K. (2000). *Psychology of Survey Responses*, Cambridge: Cambridge University Press.

Van Duyne, P. C. (2002). Necessity and Contingency in Reasoning, *Acta Psychologica*, Vol. 40, pp. 85 - 101

Von Wright, G. H. (1957). *Logical Studies*. Routledge: Psychology Press.

Wason, P. (1966). Reasoning, in *New Horizons in Psychology,* Foss, B.M. (ed), pp. 135 - 151.

___. (1968). Reasoning about a Rule. *Quarterly Journal of Experimental Psychology* Vol. 20, pp. 273 - 281.

___. (1969). Regression in Reasoning? *British Journal of Psychology*,Vol. 60, pp.471 - 480.

Wason, P., and Johnson-Laird, P. (1972) *The Psychology of Reasoning: Structure and Content*. Cambridge, Mass.: Harvard University Press.

___. (1970). A Conflict Between Selecting and Evaluating Information in an Inferential Task. *British Journal of Psychology*, Vol. 61, pp. 509 - 515.,

___. (1972). *Psychology of reasoning: Structure and Content*. London: Batsford

Wildman, T. M., and Fletcher, H. J. (1977). Developmental Decreases and Increases in Solutions of Conditional Syllogism Problems. *Developmental Psychology,* Vol. 13, pp. 630 - 636

Wilkins, M. C. (1928). The Effect of Changed Material on the Ability to do Formal Syllogistic Reasoning. *Archives of Psychology*, No. 102.

Wilson T. D., and Nisbett, R. (1978). The Accuracy of Verbal Reports about the Effects of Stimuli on Evaluations and Behavior, *Social Psychology,* Vol. 41, pp. 118 - 131.

Wilson T. D., and Stone, J. I. (1985). Limitations of Self-Knowledge: More on Telling More than We Can Know, in *Review of Personality and Social Psychology,* Vol. 6, Shaver, P. (ed), pp. 167 - 183.

Wittgenstein, L. (1953). Philosophical Investigations. London: Blackwell.

___. (1978). *Remarks on the Foundations of Mathematics*, Anscombe, G. E. M., (rev. ed. and trans.).

Wright P., and Rip, P. D. (1981). Retrospective Reports on the Causes of Decisions, *Journal of Personality and Social Psychology* Vol. 40, pp. 601 - 614.