Coherent Attributions with Co-occurring and Interacting Causes

by

Kyle Edward Jennings

A dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy in Psychology in the Graduate Division of the University of California, Berkeley

Committee in charge:
Professor Kaiping Peng, Chair
Professor Thomas Griffiths
Professor Robert MacCoun

Spring 2010
Coherent Attributions with Co-occurring and Interacting Causes

Copyright 2010
by
Kyle Edward Jennings
Abstract

Coherent Attributions with Co-occurring and Interacting Causes

by

Kyle Edward Jennings
Doctor of Philosophy in Psychology
University of California, Berkeley
Professor Kaiping Peng, Chair

Many processes within social and personality psychology require individuals to attribute the cause of an effect for which there are multiple potential causes. Whether people make these attributions correctly has been a topic of long-standing debate. This work introduces a Bayesian identity that can determine the coherence (internal consistency) of people’s attributions regarding the presence of a candidate causal factor, given the occurrence of an effect and the presence of another causal factor. Letting $U$ be the factor whose presence is uncertain, $E$ be the effect that occurred, and $C$ be the factor known to be present, then the attribution of interest, $P(U | E, C)$ is equal to $P(U) \cdot \frac{P(C | U) \cdot P(E | C, U)}{P(C)}$, which are termed the prior probability, cause-cause co-occurrence, and relative effect likelihood, respectively. Intuitively, they express how likely the uncertain factor’s presence is in general, whether the two factors tend to occur together, and how much more likely the effect is when it is known that the uncertain factor is present, as compared to when the uncertain factor’s presence is unknown. This expression can be used to intuitively determine the coherent attribution in a particular scenario, or it can be applied quantitatively. Studies are conducted that assess attribution coherence relative to people’s reported assumptions and perceptions. Two of the studies ask directly for the beliefs in the identity, one using a Likert scale to represent the log-ratio of terms in the model, the other asking for the individual probabilities. People are generally coherent, with little difference cross-culturally (East Asian vs. European-American). The approach can be made to directly match the inferences in trait and attitude attribution studies by using probability density functions over continuous variables, and letting the expected value of the posterior distribution be the normative attribution. When applied to the attitude attribution paradigm, it is shown both via postdiction and prospectively that people’s attitude attributions are potentially coherent according to the model. Implications for discounting and augmenting, interactions between causes, the fundamental attribution error and correspondence bias, and cross-cultural attribution research are discussed, and recommendations are given for improving theory and research.
# Contents

List of Figures iv  
List of Tables v  
Acknowledgments vi  

1 Introduction 1  

2 Determining the Coherence of Ascriptions with Two Causes 3  
2.1 Background ........................................ 5  
2.1.1 Applications ...................................... 6  
2.1.2 Normative Claims .................................. 7  
2.2 Conceptualizations and Operationalizations ................. 10  
2.2.1 Conceptual Distinctions ............................. 11  
2.2.2 Operationalizations ................................ 13  
2.3 Determining Correctness ................................ 16  
2.3.1 Correctness Criteria ................................. 16  
2.3.2 Product and Process ................................. 18  
2.3.3 Scoping ............................................ 19  
2.4 Standard for One Uncertain Cause ......................... 19  
2.4.1 Supporting Evidence ............................... 21  
2.4.2 Comparison with Morris and Larrick ................. 25  
2.5 Empirical Demonstration ................................ 26  
2.5.1 Study Design ....................................... 26  
2.5.2 Methods ............................................ 29  
2.5.3 Results and Analysis ............................... 33  
2.5.4 Discussion .......................................... 36  
2.6 Modeling Causal Mechanisms ............................ 39  
2.6.1 Detailed Causal Schemata ......................... 40  
2.6.2 Application to Attitude Attribution .................. 41  
2.6.3 Interactive Causation .............................. 43  
2.7 Implications for Attribution Biases ....................... 43
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.7.1 Heeding Scope</td>
<td>44</td>
</tr>
<tr>
<td>2.7.2 Meaningful Manipulations</td>
<td>45</td>
</tr>
<tr>
<td>2.7.3 Demonstrating Biases</td>
<td>46</td>
</tr>
<tr>
<td>2.8 Conclusions</td>
<td>46</td>
</tr>
<tr>
<td>2.8.1 Limitations</td>
<td>47</td>
</tr>
<tr>
<td>2.8.2 Future Directions</td>
<td>48</td>
</tr>
<tr>
<td>2.8.3 Prospects for Attribution Research</td>
<td>52</td>
</tr>
<tr>
<td>3 Application to Attitude Attributions</td>
<td>53</td>
</tr>
<tr>
<td>3.1 Normative Model</td>
<td>54</td>
</tr>
<tr>
<td>3.2 Illustrations</td>
<td>57</td>
</tr>
<tr>
<td>3.2.1 Choice and Prior Probabilities</td>
<td>57</td>
</tr>
<tr>
<td>3.2.2 Prior Probabilities Across Populations</td>
<td>59</td>
</tr>
<tr>
<td>3.2.3 Degree of Compliance</td>
<td>61</td>
</tr>
<tr>
<td>3.3 Empirical Results</td>
<td>63</td>
</tr>
<tr>
<td>3.4 Conclusions</td>
<td>64</td>
</tr>
<tr>
<td>4 Cross-Cultural Application</td>
<td>65</td>
</tr>
<tr>
<td>4.1 Background</td>
<td>66</td>
</tr>
<tr>
<td>4.1.1 Explanations: Socially-Oriented Attributions</td>
<td>67</td>
</tr>
<tr>
<td>4.1.2 Inferential Goals and Two-Stage Attribution</td>
<td>69</td>
</tr>
<tr>
<td>4.1.3 Attention to Situational Information and Subtle Behaviors</td>
<td>71</td>
</tr>
<tr>
<td>4.1.4 Toward Determining Coherence</td>
<td>73</td>
</tr>
<tr>
<td>4.2 Determining Coherence</td>
<td>74</td>
</tr>
<tr>
<td>4.2.1 Core Judgment: Ascriptive Discounting and Augmenting</td>
<td>75</td>
</tr>
<tr>
<td>4.2.2 Normative Standard</td>
<td>76</td>
</tr>
<tr>
<td>4.2.3 Culture and Background Assumptions</td>
<td>77</td>
</tr>
<tr>
<td>4.3 Study</td>
<td>79</td>
</tr>
<tr>
<td>4.3.1 Methods</td>
<td>80</td>
</tr>
<tr>
<td>4.3.2 Results</td>
<td>82</td>
</tr>
<tr>
<td>4.3.3 Discussion</td>
<td>90</td>
</tr>
<tr>
<td>4.4 General Discussion</td>
<td>91</td>
</tr>
<tr>
<td>4.4.1 Assessing Normativity</td>
<td>91</td>
</tr>
<tr>
<td>4.4.2 Cross-Cultural Methodology</td>
<td>92</td>
</tr>
<tr>
<td>4.4.3 Future Research Directions</td>
<td>93</td>
</tr>
<tr>
<td>4.5 Conclusion: So, who’s right and who’s wrong?</td>
<td>94</td>
</tr>
<tr>
<td>5 Conclusions</td>
<td>95</td>
</tr>
<tr>
<td>References</td>
<td>97</td>
</tr>
</tbody>
</table>
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Illustration of $p(e \mid a, p = 2, s = .5)$, attitude attribution</td>
<td>54</td>
</tr>
<tr>
<td>3.2</td>
<td>Model of Jones and Harris (1967), Study 1</td>
<td>57</td>
</tr>
<tr>
<td>3.3</td>
<td>Model of Miyamoto and Kitayama (2002), Study 2</td>
<td>61</td>
</tr>
<tr>
<td>3.4</td>
<td>Illustration of strong/weak essays under strong/weak constraint, attitude attribution</td>
<td>62</td>
</tr>
<tr>
<td>4.1</td>
<td>Comparison of the normative attributions, culture study</td>
<td>89</td>
</tr>
<tr>
<td>A.1</td>
<td>Venn diagram to illustrate probabilities</td>
<td>107</td>
</tr>
<tr>
<td>A.2</td>
<td>Contradiction for non-sufficient, independent causes</td>
<td>108</td>
</tr>
</tbody>
</table>
## List of Tables

2.1 Sample characteristics, empirical demonstration ·························· 29  
2.2 Scenarios, empirical demonstration ······································ 31  
2.3 Predictions and outcomes, empirical demonstration ······················· 32  
2.4 Aggregate model comparisons, empirical demonstration ····················· 34  
2.5 Individual-level coherence, empirical demonstration ······················· 35  
3.1 Model predictions based on essay and situation strength, attitude attribution 58  
4.1 Sample criteria and characteristics, culture study ·························· 81  
4.2 Discounting and augmenting by group, culture study ························ 85  
4.3 Residuals for normative model predictions, culture study ····················· 87
Acknowledgments

There are innumerable people whom I should thank—advisors and mentors, colleagues and research assistants, friends and family—all of whom were “causes” of my completing this degree. However, as the discounting principle suggests, the perceived role of any one cause will diminish when other causes are made salient. Lest this happen, my thanks go simply to Rachel Beth Egenhoefer, for being the best ever in the entire Universe and beyond, and for being the one cause whom I truly could not live without.
Chapter 1

Introduction

Many processes within social and personality psychology require individuals to attribute the cause of an effect for which there are multiple potential causes. Kelley (1972a) introduced the discounting and augmenting principles to describe the reduced or increased confidence in one cause, given the presence of another cause. Discounting and augmenting are at the center of many important research paradigms, and are interesting phenomena in their own right. Additionally, determining when and by how much people ought to discount or augment bears directly on the fundamental attribution error and correspondence bias. Therefore, it is of great importance to understand these phenomena further.

Whether people discount and augment correctly has been a topic of long-standing debate. Complicating this debate is the variety of conceptualizations, operationalizations, and normative criteria in use by different researchers. Building on earlier work (Morris & Larrick, 1995; McClure, 1998), Chapter 2 attempts to clarify this research area by (a) classifying the conceptual and operational differences in the literature, (b) clarifying the kinds of standards that can be applied to people’s judgments, and (c) presenting normative standards that can be used to assess the coherence (internal consistency) of people’s judgments about the cause for a particular event when doubt is cast on a candidate cause due to the presence of an alternate cause.

Useful though it is, the Bayesian identity only specifies the relationships between the probabilities of dichotomous events. However, most attribution studies ask participants to make judgments about continuous quantities, such as traits or attitudes. While discrete events can be made to approximate judgments about continuous variables, important details are lost regarding how the judgments are made, and differences between the judgment and the confidence in the judgment. Chapter 3 remedies this problem by generalizing the identity to probability density functions over continuous variables. The approach is demonstrated with the attitude attribution paradigm, where observers must estimate the true attitude of an author who was assigned to advocate a particular position. Though previous research has convincingly shown that people’s attributions in this paradigm diverge from the ground truth, it is less clear whether they are at least coherent with people’s assumptions and perceptions. The model is used to reproduce classical findings regarding chosen
vs. assigned positions, base rates, and degree of compliance, and also fits newly collected data. The results suggest that people’s attributions in this paradigm may be coherent.

One of the most exciting developments in the last two decades of social psychology has been the emergence of cross-cultural comparisons into the research mainstream. This emergence has been felt particularly strongly in attribution research, where it has been argued that East Asians generally consider situational factors more than Westerners—that is, that they may not make the fundamental attribution error. However, previous research has been unable to definitely answer the obvious question of whether this difference implies that one culture is right and the other is wrong, or whether instead people in both cultures reach conclusions that are consistent with other culture-specific assumptions and demands. Answering this question requires clarity about what part of the attribution process is under consideration, and precision about how perceptions and background assumptions should be consolidated into a normatively correct conclusion. Using the Bayesian model, Chapter 4 compares participants across cultures. The results suggest that people’s dispositional attributions are equally coherent in both cultures when a situational cause is also present, but that European-Americans are more normative when the situational cause’s presence is unknown. The work also shows that the normative attribution can vary according to cultural and individual beliefs, and that failing to account for these factors can lead to an inaccurate understanding of cultural differences.

Overall, the work presented here illustrates a promising new approach to rigorously conducting research on social judgment, and for making responsible cross-cultural comparisons. The Bayesian identity is itself a useful tool that can be applied to many more research questions than are presented here. Additionally, the approach itself models an orientation toward research that considers the formal specification of theories to be as important as the use of inferential statistics, a set of priorities with great potential for improving the quality of social judgment research.
Chapter 2

Determining the Coherence of Ascriptions with Two Causes

Though it is no longer controversial to say that behavior is jointly determined by personality and the situation, questions remain about how these causal forces interact, and how perceivers understand their joint influence. The longstanding belief that people systematically underestimate the power of the situation (the fundamental attribution error; Ross, 1977) has been challenged both in general (Harvey, Town, & Yarkin, 1981; Funder, 1987; Gawronski, 2004) and by findings that situational neglect occurs to a lesser extent in East Asian cultures (J. G. Miller, 1984; Morris & Peng, 1994; Choi, Nisbett, & Norenzayan, 1999). Regardless, evidence that people do not perceive their social worlds correctly abounds (Gilbert & Malone, 1995). These contradictory perspectives require resolution.

Two phenomena, discounting and augmenting (Kelley, 1972a), are at the core of attributions involving multiple causes (such as personality and situational influences). Discounting occurs when the perceived role of one cause is diminished by the actual or possible presence of an alternate cause (or causes). Conversely, augmenting occurs when the perceived role of one cause is increased by the actual or possible presence of an additional cause (or causes). Though not all research involving attribution with multiple causes invokes discounting and augmenting directly, discounting and augmenting can and have been applied to a diverse variety of phenomena (McClure, 1998), and bear directly on basic questions like the fundamental attribution error (Hilton, 2007). Therefore, clarifying the diverse findings on discounting and augmenting promises to contribute greatly to the clarification of how people make attributions in general.

After their original proposal, discounting and augmenting received wide and often uncritical acceptance (Hansen & Hall, 1985). When put to greater scrutiny, however, disagreement about when and to what degree people should discount or augment is readily apparent (Morris & Larrick, 1995; McClure, 1998). Great progress toward clarifying this disagreement was made by reviews by Morris and Larrick (1995) and McClure (1998), and a recent review by Hilton (2007) has added needed clarity to attribution research more generally. However, there are still three broad barriers to progress, which this review aims to
remedy.

The first barrier is that researchers have conceptualized and operationalized what are ostensibly the same concepts in contradictory ways (Morris & Larrick, 1995). This review takes a closer look at the experimental methodologies used in the discounting and augmenting literature than was afforded in previous reviews, and suggests that the different methodologies actually require different normative standards. In addition to more minor distinctions, the long-recognized difference between explanation and attribution (Hilton, 1990, 1991; Hilton & Erb, 1996) is amplified, and the distinction between causal induction and causal ascription is discussed.

The second barrier to progress has been that researchers in social judgment (as in judgment and decision making more generally) have conceived of what makes a judgment “correct” in different ways. For instance, while some theories focus on whether people’s judgments are internally consistent with their beliefs and perceptions (e.g., Jones & Davis, 1965; Kelley, 1967; Ajzen & Fishbein, 1975; Morris & Larrick, 1995), others have focused on compelling demonstrations that people’s judgments don’t match reality (e.g., A. G. Miller, Jones, & Hinkle, 1981; Reeder, Fletcher, & Furman, 1989). Still other researchers have argued that neither of the latter questions matters as much as whether people can make judgments that are adaptive in real life (e.g., Swann, 1984; Funder, 1987; Andrews, 2001). In between these extremes are researchers who have variously used all of these criteria when considering correctness. This review discusses the relationship between different standards of correctness, and contextualizes past research under this rubric.

The final barrier to progress in discounting and augmenting research is that researchers have often not been explicit about what constitutes a correct judgment (Kelley & Michela, 1980; McClure, 1998), or have not properly justified their prescriptions for correctness (Harvey et al., 1981; Morris & Larrick, 1995). As a result, the discounting and augmenting literature is filled with seemingly contradictory findings that belie more agreement and a more positive take on human reasoning than is generally acknowledged (cf. Hilton, 2007). Though Morris and Larrick (1995) present a useful normative standard for discounting and augmenting with multiple sufficient and multiple necessary causes, their analysis does not extend beyond these cases, and is difficult to apply. Like Morris and Larrick, this review develops a Bayesian standard, but unlike Morris and Larrick the standard is not tied to previous theoretical notions, making it more general, flexible, and intuitive. Additionally, the review shows how the normative standard can be adapted from discrete probability judgments to judgments of continuous quantities, allowing it to directly match the inference problems in common experimental paradigms.

This review proceeds as follows. First, highlights of the development of discounting and augmenting are presented, and it is shown how these phenomena pervade many attribution problems. The major normative standards are reviewed in the process, with particular attention paid to past Bayesian standards. Next, the various operationalizations used in this literature are examined and classified, the different normative questions are identified, and the different ways to judge correctness are clarified. From here the review focuses on how to judge the coherence (internal consistency) of causal ascriptions (recovering the causal
structure for one particular event) when there is one cause whose role is known, and one cause whose role is unknown. Previous research results are reinterpreted in light of this standard, and the standard is compared with prior theories of discounting and augmenting. A small empirical study is run to demonstrate how to measure beliefs and assess coherence. Finally, examples are given of how the standard can be generalized beyond the presence or absence of discrete events to questions about continuous variables. In closing, the implications for research in attribution biases are discussed.

2.1 Background

The discounting and augmenting principles were first introduced by Kelley (1972a), who defended them by reframing past research results in terms of how perceivers reconciled multiple possible causes for the same behavior. Kelley cites Thibaut and Riecken (1955) as the “progenitor” (p. 8) of this line of investigation, and as an example of discounting. Thibaut and Riecken studied how a participant’s liking for another individual changed when that individual (a confederate of the experimenter) complied with the participant’s request for help. Though compliance should increase liking in general, they manipulated the relative status of the compliant individual, reasoning that the participant’s liking for the confederate would depend on perceiving the compliance as self-chosen. Whereas there is only one strong cause for a high status individual’s compliance (choosing to), there are two strong causes for a low status individual’s compliance (choosing to and feeling compelled to). As a result of multiple plausible causes for compliance, the key attribution (choosing to comply) is discounted for the low status person but not for the high status person, leading to the obtained differences in liking for that person.1

Like the other studies that Kelley reviews, Thibaut and Riecken’s study doesn’t test discounting directly, but can be construed as reflecting a reduction in the role of one cause given the presence of other plausible causes. In this vein, Kelley also cites Jones, Davis, and Gergen (1961), a study where participants judged the personality of individuals via recorded interviews in which they were said to be applying for a job. Participants (and ostensibly the recorded individuals) were aware that the job required particular personality characteristics. When the interviewees described themselves in terms consistent with the job requirements, participants were less likely to believe that the statements reflected the interviewee’s personality. However, when the interviewees described themselves in ways inconsistent with the job requirements, participants viewed those statements as highly reflective of the individual’s personality. Presumably, there are two reasons for making statements that match the job requirements (making accurate self-representations, and qualifying one’s self for the job) while there is only one reason for making statements that do not match the job requirements (accurate self-representation), and indeed good reasons not to (potentially disquali-

---

1McClure (1998) challenges the relevance of this study to discounting since it used a bipolar person-situation scale, which as discussed later, can overstate the amount of discounting. However, the effects for liking still suggest that discounting may have occurred.
fying one’s self from the job). Kelley differentiated these two causes as being facilitative and inhibitory (respectively) in that they would tend to either encourage or discourage the effect. In this way, the Jones et al. study can be seen as demonstrating augmenting, or increased confidence in a facilitative cause given the presence of an inhibitory cause.

The Jones et al. study was designed as a test not of discounting or augmenting, but of Correspondent Inference Theory (CIT; Jones & Davis, 1965). As Kelley (1972a, 1973) recognizes, this illustrates that discounting and (especially) augmenting are present in Correspondent Inference Theory itself. Central to CIT is the assumption that an action inevitably has both desirable and undesirable effects (e.g., going in debt in order to get the car one wants), and that something can be learned about the importance of a desirable effect by looking at what undesirable effects were assumed as part of the bargain. Among the predictions that Jones and Davis make is that the more effects a choice has, the less can be inferred about the value of any one of them. When the effects are treated at “causes” for the choice (i.e., the action was taken in order to obtain that particular effect), this is a clear parallel to the discounting principle’s claim that confidence in one cause’s role should be reduced when there are other possible causes. Likewise, CIT predicts that negative effects (causes) are greatly informative about the value to the individual of the positive effects, similar to the augmenting principle (Kelley, 1973; Jones & McGillis, 1976).

2.1.1 Applications

Discounting and augmenting are of more than historical interest, and are applicable beyond their typical use with internal and external causes for interpersonal and intrapersonal situations (McClure, 1998). Indeed, they may be part of a larger sort of reasoning, which Hansen and Hall (1985) call “multiple force causation” and Spellman (1996) and Fiedler (2007) refer to “trivariate reasoning”.

Discounting and augmenting can also be seen as relevant to specific attribution problems that underlie broad areas of intrapersonal behavior. Several authors have discussed Bem’s (1967) attributional theory of cognitive dissonance as applied to intrinsic motivation in discounting terms (Kelley, 1972b, 1973; Ajzen & Fishbein, 1975; Morris & Larrick, 1995; McClure, 1998). Attributions of ability have also been discussed in terms of discounting and augmented among the candidate causes of ability, effort, and task difficulty (Kun & Weiner, 1973; Kelley, 1972b; Ajzen & Fishbein, 1975).

Despite their generality, discounting and augmenting are most often discussed in terms of understanding person and situation causes of behavior. Kelley (1972a) talked about discounting as a tacit step in trait inference. Discounting can also be seen, directly or indirectly, in some of the most common paradigms in the attribution literature, and particularly in paradigms that have been used as evidence for attribution biases. Most famously, Jones and Harris (1967) found that when people attributed an author’s attitude on the basis of an

---

2The theory qualifies this to say effects not shared in common with other choices, a distinction that is not relevant here.
essay that the author ostensibly wrote, they apparently did not take into sufficient account that the author had been assigned which position to express. This can be interpreted as failing to sufficiently discount the author’s attitude as a cause of the essay in light of the alternative cause, the situational inducement. A similar pitting of competing causes for behavior can be seen in the silent interview paradigm (Snyder & Frankel, 1976), wherein an interviewee’s nervous behavior can be attributed either to the content of the interview questions, or to the interviewee’s personality. These two paradigms are often paired with manipulations of cognitive load or motivation (e.g., Gilbert, Pelham, & Krull, 1988), with the result often being that people’s attributions under cognitive load are more dispositional than more effortfully-made attributions (Gilbert & Malone, 1995), which can be interpreted as an interruption of the discounting process (McClure, 1998).

2.1.2 Normative Claims

Given discounting’s and augmenting’s roles in the study of attribution biases, it is of obvious interest to know whether and by how much discounting and augmenting are normatively justified. Kelley’s (1972a) original definitions for discounting and augmenting included both a description of the phenomenon (e.g., “the role of a given cause in producing a given effect is discounted”) and a claim about when the phenomenon should occur (“if other plausible causes are also present”, both p. 8). His justification for this claim is that by thinning the number of causes thought involved, “the attributor does pretty much what a good scientist would do” (p. 8), i.e., be parsimonious (Morris & Larrick, 1995). While Kelley does review a good deal of evidence showing that people do discount under conditions matching his statements, such evidence can support his claims as descriptive, but not as normative. What’s more, his statements lack the precision necessary to determine not just whether, but by how much people should discount. (Similar criticisms apply to other standards, e.g., Kruglanski, 1980.)

Kelley began to work toward a solution by advancing his Causal Schemata Model (CSM; Kelley, 1972b), which proposes that people have “causal schemata” or “general conception[s] ... about how certain kinds of causes interact to produce a specific kind of effect” (p. 151). Kelley describes causal schemata using a matrix representation with the presence or level of one cause (or kind of cause) in rows, another (kind of) cause in columns, and the presence (or level) of the effect in each cell. He suggests that causal schemata have “reversibility”, meaning that one can start with any sufficient subset of the information in the schema and from there determine the missing information. Thus, one could start knowing what effect occurred and what the level of one cause was, and from their determine the likely level of the other cause. He proposes several different kinds of causal schemata (e.g., multiple sufficient causes, multiple necessary causes, as well as more complex configurations), and speculates about what sorts of inferences would be logical for the various causal schemata (particularly with the multiple sufficient and multiple necessary causes schemata). Ultimately, however, he offers these questions as topics for future research, suggesting that the primary variable will be the “degree of unequivocality of inference” that would be jus-
tified in light of “different sets of information” (p. 169).

In possible contrast to Kelley’s discounting writings, Correspondent Inference Theory was intended to be a normative standard (Jones & McGillis, 1976). Its justifications were originally rooted in the logic of comparing what could be assumed about a person before an action was taken with what could be inferred about a person after the action was taken (Jones & Davis, 1965). In particular, effects that anyone would desire to achieve or choices that anyone would make in the same circumstances were thought not to add information relative to what would already have been assumed (Jones and Harris [1967] added the language of information theory to their exposition). Additionally, only conclusions about the person that differed from default assumptions about that person were considered to be “correspondent”.

In its original form, Correspondent Inference Theory allowed only dichotomous predictions of when information would or would not be gained from observing a behavior. Apropos discounting, Jones and Harris (1967) was designed to show that behavioral freedom would matter only for acts with a low prior probability of occurrence, with a correspondent inference possible only with high behavioral freedom. Thus, Jones and Harris completely expected to find attributed attitudes in line with either the assigned or chosen position for the anti-Castro essay, since the behavior of opposing Castro had a high prior probability. They likewise expected attributions in line with behavior for the freely chosen, anti-Castro essays. The surprise that motivated them and many researchers afterwards came at the finding that people did not revert to the default, anti-Castro attitude assumption when essay authors were assigned to write pro-Castro essays, and instead offered neutral attributions.

Like Kelley’s original statements of discounting and augmenting, CIT lacked the precision to account for anything more than directional outcomes. However, many authors (e.g., Ajzen, 1971; Trope, 1974, 1978; Ajzen & Fishbein, 1975) recognized that with its focus on belief revision upon encountering new information, CIT could be understood in Bayesian terms. Broadly, a Bayesian theory is one that has at its root some form of Bayes’ theorem, which prescribes how previous knowledge should be revised in light of new information. If $D$ is an hypothesised disposition that can be present or not, and $B$ is some behavior that relates to this disposition, then Bayes’ theorem says that:

$$P(D | B) = P(D) \cdot P(B | D)/P(B)$$

In words, the probability of the disposition being present after having observed the behavior is equal to the probability that the disposition was present prior to observing the behavior, raised in proportion to how likely it is that the behavior would be observed if the disposition were present, and lowered in proportion to how common the behavior is overall. This matches Jones and Davis’ statement that unusual choices (i.e., $P(B)$ low) are more informative than “normative” choices, a result amplified by Ajzen (1971) and Trope (1974).

Ajzen and Fishbein (1975) rewrite Bayes’ theorem as a statement regarding the prior odds of the disposition (or cause) and the posterior (after observing the behavior) odds of
the disposition (or cause):

\[
\frac{P(D | B)}{P(\overline{D} | B)} = \frac{P(B | D)}{P(B | \overline{D})} \cdot \frac{P(D)}{P(\overline{D})}
\]

In this form, the key determinant of whether there is any belief revision is what is called
the likelihood ratio, \(P(B | D)/P(B | \overline{D})\). If the data are no more or less likely to occur
with or without the hypothesis being true, then nothing new is learned when the data are
observed. For this reason, both Ajzen and Fishbein and Jones and McGillis (1976) define
a dispositional attribution as being a change in knowledge about a person relative to what
had previously been assumed.

Ajzen and Fishbein’s take on the discounting principle is that the presence of alter-
native causes would tend to increase \(P(B | D)\), and hence make an attribution less jus-
tified. However, they don’t consider whether that increase would be offset by changes in
\(P(B | D)\), nor do they handle causes that are not independent. What’s more, their dis-
cussion couches difficult attribution judgments in probabilistic terms without necessarily
elucidating how those judgments are or ought to be made (Fischhoff & Lichtenstein, 1978).
In this way, Kelley’s Causal Schemata Model is more specific and testable, as is Reeder
and Brewer’s (1979) Implicational Schemata Model (ISM), which like the CSM specifies
different forms of trait-behavior relationships in light of varying situational demands, and
asserts that the form of these relations dictates what kinds of inferences people can make
on the basis of observed behavior. In presenting their model, Reeder and Brewer refer
schematically to \(P(B_i | D_j)\), the probability of a behavior of level \(B_i\) given a disposition
at level \(D_j\). While they disavow claims to be a Bayesian model, this shorthand illustrates
the point that schematic models specify a relationship where the effect is the dependent
variable and the levels of various causes are the independent variables. In contrast, in attri-
bution the dependent variable is the level of one of the causes (e.g., level of a trait) while the
independent variables are the observed effect, and possibly knowledge about one or more
additional causes. As Bayes’ theorem shows, it is not in general true that \(P(B | D) = P(D | B)\). Mistaking one conditional probability for the other is a common mistake in
everday statistical reasoning (Koehler, 1996; Villejoubert & Mandel, 2002), and it appears to
be present in these formal theoretical treatments of attribution, as well, via the assumption
of reversibility (see also Einhorn & Hogarth, 1986).

An exception to the above statement are the Bayesian adaptations of Kelley’s CSM by
Morris and Larrick (1995). Morris and Larrick sought to generalize Kelley’s model by re-
 laxing his assumptions about the sufficiency of the causes and their statistical independence,
and the number of causes besides the two under consideration. Their analysis compares the
probability of a target cause being present before and after the presence of an alternative
cause is known, where in both cases an effect was observed that might relate to either cause.
Letting \(U\) be the cause whose presence is uncertain, \(C\) be the cause whose presence is later
learned to be certain, and \(E\) be the observed effect, they operationalize the two probabilities

\[\text{Medcof (1990) makes similar use of conditional probabilities without intending a Bayesian framework.}\]
of interest as \( P(U \mid E) \) and \( P(U \mid E, C) \), respectively.\(^4\) They then analyze different combinations of the sufficiency, independence, and number of causes in order to determine when discounting is warranted, which in their treatment means that \( P(U \mid E) > P(U \mid E, C).\(^5\)

As these examples show, Bayes’ theorem is quite applicable to the phenomena of discounting and augmenting, and can offer more detailed predictions (or prescriptions) than other kinds of theories. However, there are multiple ways that Bayes’ theorem can be applied, and there are multiple ways that discounting and augmenting can be operationalized. These conflicting views require resolution before moving forward.

### 2.2 Conceptualizations and Operationalizations

In a review of the actor-observer effect, Kelley and Michela (1980) note that the variety of interpretations of the theory made summarizing the research results difficult as “problems of operationalization become entangled with problems of conceptualization” (p. 477). A similar state of affairs affects discounting research, which has been interpreted and operationalized in many different ways (Morris & Larrick, 1995). The result is a confusing variety of experimental results purporting to be about the same phenomena, but that actually pertain to different questions. This section aims to clarify matters, first by adding some clarification to the concepts surrounding discounting, and then by paying particular attention to operationalizations.

Some of the confusion in the discounting literature stems from the vagueness of the original concept. Kelley’s first statement of the discounting principle is that “the role of a given cause in producing a given effect is discounted if other plausible causes are also present” (1972a, p. 8). While this definition suggests that one particular cause is discounted when other plausible causes are present, later Kelley states that when a person “is aware of several plausible causes, he \[sic\] attributes the effect less to any one of them than if he is aware of only one as a plausible cause” (p. 8 Kelley, 1972a). In the latter statement, the contrast is made between the role of a single cause and the roles of each of several causes, alternately suggesting one cause as the target or several causes at the target.

An additional ambiguity in both of the above forms of discounting is in what is meant by the cause’s “role” being reduced. Kelley suggests two possible ways to detect this reduction. The less problematic of the two occurs via a reduced willingness to infer that the magnitude of a cause is as great as would otherwise seem likely. Though this case involves the assumption that some other attribution (e.g., of a trait or attitude along a bipolar continuum) is mediated by a casual judgment,\(^6\) the outcome is straightforward to operationalize.

\(^4\)Morris and Larrick use \( A \) and \( B \) to refer to the two causes, which matches Kelley’s usage. However, since the two causes have a different status, the present treatment gives them more mnemonically useful labels.

\(^5\)Kun, Murray, and Sredl (1980) use \( P(U) > P(U \mid E, C) \), which is non-standard, but not necessarily over-discounting as Morris and Larrick (1995) claim.

\(^6\)As discussed later, this paper’s model does not make causal assumptions, but since Kelley did, it is relevant to report this concern.
(though presumably both the attribution and the confidence in the attribution may be affected).

More problematic is Kelley’s suggestion that the “role” reduction might occur when “[t]he attributor is less confident that the observed effect reflects the given cause” (p. 8, emphasis added). A straightforward way to see if an effect “reflects” a cause is when the cause may or may not be present. Less straightforward is when the cause is known to be present (and if applicable, its level is known), in which case whether and by how much the cause is reflected becomes difficult to reason about: the cause is there, and the effect is there, so what would it mean that the effect doesn’t reflect the cause?

To address these issues, two key distinctions must be reviewed: causal discounting vs. causal backgrounding, and causal ascription vs. causal induction. With these distinctions in mind, experimental scenarios are compared based on whether the certainty surrounding the candidate causes is the same for all causes, or different across causes.

### 2.2.1 Conceptual Distinctions

**Discounting vs. Backgrounding**

The number of causes for an event far exceeds the set of causes people typically consider. For instance, when determining the cause of an automobile accident, the invention of the automobile and the availability of fuel are certainly “causes” in the sense that they occurred prior to the event, and the event would not have occurred in their absence. However, it is rarely meaningful to mention such causes. Additionally, the fact that the automobile was invented should not reduce one’s confidence that slippery roads caused the crash. Still, in most conversations, mentioning the invention of the automobile when explaining a crash would not be informative, nor would it likely be considered a “cause” (cf. Einhorn & Hogarth, 1986).

To explain this seeming contradiction, it is useful to recognize a distinction between attribution, which can be seen as the solitary pursuit of causal understanding, and explanation, which is an inherently social act wherein “someone explains something to someone” (Hilton, 1990, p. 65; see also Hilton, 1991; Hilton & Erb, 1996). The presence of others leads to an attention to norms of conversation, such as those articulated by Grice (1975). Most relevant among them is his maxim of *quantity*, which says that conversation partners should reveal no more or less information that is essential to communicating an idea. Since both conversation partners presumably both know that the automobile has been invented, it is not relevant to mention in a conversational context.

There are now two reasons that a cause may not be mentioned. As Kelley discusses, a cause’s role may be discounted if the attributor changes his or her estimate of the presence or magnitude of a cause, which may then lead to its exclusion from an explanation of the event. Additionally, conversational considerations may lead to cause not being featured in an explanation of a series of events, while still being part of a person’s causal model. In this case, *causal backgrounding* is said to have occurred (Hilton & Erb, 1996). Hilton
(2007) distinguishes these according to whether the understanding of the event changes (discounting) or is merely tuned (backgrounding).

Hilton (2007) suggests that since causal backgrounding is a conversational phenomenon, it should only impact open-ended responses on experiments, and not affect rating scales. However, the design of the experiment and the nature of the question can also affect which phenomenon is tapped. Most obviously, some studies specifically designed to test causal backgrounding (and some that aren’t) ask participants to rate how well a particular cause explains the event, how good of an explanation for the event a cause is, or some variation thereof. While participants are not expected to be sensitive to technical distinctions between attribution and explanation, these sorts of questions are clearly different from attribution-related questions such as whether the cause was present, or if the “cause” is actually a trait or attitude, what level the actor possesses.

As will be discussed, some early results in the discounting literature might better be understood as backgrounding (see also McClure, 1998). Even in contemporary research the distinction is not universally recognized, with Hilton (2007) recently suggesting that distinguishing discounting and backgrounding is an important goal for future discounting research.

**Ascription vs. Induction**

Discounting and backgrounding differ according to whether the causal model is affected. However, another question is what kind of causal information is involved. Thus far the inference target has been one particular event, with the aim of understanding why that event occurred. This inference can be called causal ascription, and can be compared to reconstructing the causal scenario for that event, drawing upon background knowledge about each of the purported causes. This stands in contrast to causal induction, where the goal is to determine in general how strong a particular cause-effect relationship is.

Hilton (2007) notes that while induction concerns general relationships between events, ascription concerns particular, specific events. He also cites Hart and Honoré (1985) as differentiating induction and ascription by whether the goal is to formulate or apply generalizations. To illustrate the distinction, Morris and Larrick (1995) connect induction with research psychologists engaged in basic science, whose goal is to discover regular causal relationships, and ascription with clinical psychologists engaged in helping patients, whose goal is to determine which causal factors are at play in a particular case.

The previous distinctions differentiate induction and ascription by the goal of inferring something general or something specific. The two have also been differentiated according to the process involved in the inference. Morris and Larrick (1995) say that induction is the bottom-up determination of cause and effect relationships via the examination of covariation between cause and effect, while ascription is the top-down application of assumptions about cause-effect relationships to infer which causes were present in a single case. Despite their conceptual distinctness, induction and ascription are highly related in practice. For instance, the ascribed cause of an event becomes covariation data that will contribute to
a future inductive inference about the causal power of that event (Morris & Larrick, 1995).

Probabilities can be a useful way to understand the differences between induction and ascription. Induction is ultimately a question of whether and when a particular kind of cause can create a particular kind of effect. This corresponds to a prediction predicated on the cause's presence, i.e., \( P(E \mid C) \). Models of causal induction focus on the difference in the probability of an effect when the cause is and is not present, \( \Delta P = P(E \mid C) - P(E \mid \bar{C}) \) (Cheng, 1997). In contrast, causal ascription presupposes knowledge of causes and effects, and asks about the probability that a cause was present given that an effect occurred, \( P(C \mid E) \) (Ajzen & Fishbein, 1975; Morris & Larrick, 1995). These are clearly different probabilities, though they are also clearly related by Bayes' theorem.

The close relationship between induction and ascription in everyday reasoning has led to some confusion in theoretical and empirical work. Indeed, according to Morris and Larrick (1995), Heider (1958) does not recognize the distinction. Empirical examples can be seen in tests of Kelley’s covariation model. For instance, McArthur (1972) presents descriptions of a particular event (e.g., “John laughs at the comedian”, p. 174) along with information corresponding to different configurations of consensus, distinctiveness, and consistency. Participants are asked to judge “what probably caused the event to occur” (p. 174) and to predict the actor’s behavior in a similar situation. Both questions may require participants to infer the existence of a causal relationship between the “cause” (e.g., the person, the situation) and a category of effect (e.g., laughing at a comedian), which is induction. However, asking what caused the particular effect in the scenario just read is at least superficially ascriptive in nature: it is asking participants to first infer what causal relationships exist, \( P(E \mid C) - P(E \mid \bar{C}) \), and then to use that to ascribe the cause of one particular case, \( P(C \mid E) \). (McArthur’s second dependent variable, however, is a direct assessment of \( P(E \mid C) \).)

Kelley’s original intent was that discounting apply to cases where the attributor “has little information about a given effect and one or more possible causes”, not when the attributor possesses “cause and effect information for successive points in time” (Kelley, 1972a, p. 8). However, some researchers have extended the concept of discounting to inductive reasoning, sometimes acknowledging that doing so goes beyond Kelley’s original intent (e.g., Van Overwalle & Timmermans, 2005). Of course, though “framer’s intent” carries some weight in determining whether an extension to a theory is appropriate, a more important question is whether the extension is logically consistent with the original theory. In the case of inductive and ascriptive discounting, the question can be approached by asking whether the same normative standard would apply to both phenomena. This question will be returned to later. In the interim, this review will focus on ascriptive discounting.

### 2.2.2 Operationalizations

Studies have used different operationalizations for discounting, both in the types of information they provide and the dependent variables they have used. McClure (1998) recommends distinguishing questions asked according to whether they pertain to the prob-
ability, necessity, sufficiency, or explanatory relevance of a cause. Explanatory relevance is clearly related to causal backgrounding and not discounting. Causal necessity and sufficiency are judgments of $P(E \mid \overline{C})$ and $P(E \mid C)$. In studies of induction, these are relevant as dependent measures, but in studies of ascription these are only relevant as independent variables. Only the probability of a cause, $P(C \mid E)$, is a suitable dependent variable for ascription studies.

The information provided in a study also influences the meaning of the questions asked. Broadly, informational variations differ by the amount of certainty attached to the causal factors. An uncertain factor is one where either the presence of a binary cause is unknown, or where the strength and direction of a continuous cause (e.g., an attitude or trait dimension) is unknown. In contrast, a certain factor is one where the cause was known to be present and, if applicable, its strength and direction are also known. With these distinctions in mind, studies can be differentiated according to whether all of the causes have the same status, or whether the causes have mixed status.

**Same Status**

Kelley’s discussion of discounting allows for the case where all causes of an event are known to have been present, in which case their “role” is said to be reduced. This has led to a variety of study designs and dependent variables, some of which are not relevant to discounting.

When all causes are of known strength and direction, discounting may simply not be meaningful. Consider two experiments reported by Rosenfield and Stephan (1977), whose dependent variable is the degree to which a cause had a facilitative or inhibitory influence on the effect. What makes this choice peculiar is that they provide prior information about the direction and strength of both (Experiment 1) or one (Experiment 2) of the causes. They fail to find any discounting or augmenting when information is given about both causes, but do find discounting and augmenting when information is only given about one cause. Though they interpret this in terms of the anchoring effect of prior information, a more likely explanation is that the question is simply not very meaningful when the answer has in essence been given already. In contrast, when the question is asked without prior information about one of the cause’s direction and strength, people must first infer this information, and can sensibly answer the question.

Some studies where all causes are certain may also be tapping explanation, and hence backgrounding, rather than discounting. For example, Kruglanski, Schwartz, Maides, and Hamel (1978) presents scenarios where all of the candidate causes are present. Both studies in this paper include a dependent variable of explanatory value, which they find decreases as the number of causes increases, regardless of whether the causes are facilitative or inhibitory. In contrast, the second study also measures the perceived **strength** of a cause, which is found to be augmented when an inhibitory cause is present. Though Kruglanski et al. interpret this to mean that discounting and augmenting operate on separate judgments, a more contemporary understanding of these results is that one dependent variable tapped
explanatory value and another tapped causal structure.

Some dependent variables do not have an obvious meaning in relation to the probabilistic treatment of discounting. For example, the third study in Morris and Peng (1994) asks participants to rate the extent to which each of several factors was “a cause” of a particular murder. (This is not a study of discounting, but serves to illustrate the point.) This study may have been tapping the distinction between causes and enabling conditions (Einhorn & Hogarth, 1986; Cheng & Novick, 1991), in which case either explanatory value or an appraisal of the cause’s power is being assessed, neither of which is directly relevant to ascription. Still, participants may also have been asking an ascriptive question wherein they had some prior expectation of how powerful each of the causes is generally, but were attempting to determine whether these causes figured into the chain of events leading up to the effect to be explained. This latter question hearkens more to a view of attribution as story understanding (Read, 1987; Read & Marcus-Newhall, 1993), which is a more detailed form of attribution that requires more detailed information about the situation in question, and that goes beyond the scope of what this paper’s model is intended to handle. Reconciling these approaches is an important question for future research, but won’t be approached at this point.

A final possibility is when all of the possible causal factors are uncertain. This type of situation may occur in settings like attribution of success or failure, where effort, ability, and task difficulty can all contribute to performance (e.g., either effort and ability or an easy task are sufficient for good performance). Studies of this type often find phenomena such as conjunction effects (Leddo, Abelson, & Gross, 1984), where the presence of two causes is judged more probable than either alone. Studies that assess multiple causal factors where the status of each is uncertain are essentially asking for \( P(U_1 \mid E) \), \( P(U_2 \mid E) \), and so forth. Though questions about these probabilities can be analyzed formally, doing so will be left for future research.

**Different Status**

When discussing the causal schemata model, Kelley (1972b) notes that “[a] wide variety of experiments have been conducted according to the following paradigm: subjects are given information regarding an effect and the state of one possible cause for it and are asked to judge the magnitude (or presence versus absence) of another possible cause” (p. 154). This is the form of experiments that use the attitude attribution paradigm and the silent interview paradigm, and is also the relevant judgment when people notice situational causes of behavior (which are often concrete and known with greater certainty) and need to infer personal causes of behavior (which may be more variable). Judgments of this form can be handled via the probability \( P(U \mid E, C) \), where once again \( U \) is the uncertain cause and \( C \) is the certain cause. Because of the importance of this kind of judgment, this review will focus henceforth on judgments that match this format. Initially the focus will be on how to handle cases where \( U \) can be either present or absent, but then examples will be given of how to extend the judgment to cases where \( U \) is a continuous trait or attitude dimension.
2.3 Determining Correctness

Before introducing a normative standard that is intended to determine whether people’s judgments are correct, it is important to say what is actually meant by “correct”. After clarifying what kind of correctness is of interest, the theory’s scope will be limited, lest it be too broadly interpreted.

2.3.1 Correctness Criteria

There are three broad kinds of correctness criteria that can be applied to human judgment. Following Hammond (1996) and Dunwoody (2009), these can be called correspondence, coherence, and pragmatic criteria. Each will be reviewed in turn, and prior attribution theories will be discussed in relation to these criteria.

Correspondence Criteria

A judgment is correspondent to the extent that the conclusion reached matches the ground truth. For instance, a correspondent causal ascription for a person’s lateness would identify the actual factor that delayed the person’s arrival. Note that this use of the term “correspondent” differs from its use in Correspondent Inference Theory, where it means an attribution of a disposition that corresponds to (matches) the behavior shown (e.g., Jones & Davis, 1965; Reeder & Brewer, 1979). The concept of correspondence to ground truth has also been referred to as calibration (Baron, 2004, as cited by Dunwoody, 2009) and accuracy (Ajzen & Fishbein, 1975; Funder, 1987, though Funder’s use of accuracy also crosses into pragmatic criteria).

Determining correspondence is in general difficult to do. As Kelley and Michela (1980) point out, “the entire enterprise of psychology is directed toward specification of the true causes of behavior, and since these causes and their relative magnitudes are not yet known, it may be impossible to design a study to test unequivocally the accuracy [read, correspondence] of attribution” (p. 479). Still, there are sometimes ways around this problem. One simple solution is to take a judgment that is logically constrained, such as asking participants to estimate their ranking on some trait relative to the median of the population, and noting that more than half of respondents view themselves as above the median (Krueger, 1998). While this approach cannot say who is biased, it can detect an overall bias. There are also cases where criteria are available for each participant. For instance, if both the observer and observed are participants, the observed person can self-report on a particular characteristic, which can then be compared with an observer’s attribution of that characteristic (A. G. Miller et al., 1981; Reeder et al., 1989). All of the examples just cited have found that people’s judgments lack correspondence.
Coherence Criteria

Despite cases like those above, not all interesting psychological phenomenon can be compared against the ground truth under controlled conditions. Lacking correspondence criteria, an alternative approach is to assess the coherence of a person’s judgments with other beliefs and perceptions that the person has.

Attribution theories have tended to use coherence criteria. In presenting his covariation model, Kelley (1967) says that “the specified evidence provides a basis for subjective validity ... but not necessarily a basis for ... objective validity” (pp. 197–8). Likewise, later in the paper he discusses how phenomena such as cognitive dissonance and misattribution of arousal could be rational conclusions based on inaccurate information. Further aligning the theory with coherence, Kelley and Michela (1980) note that the theory should be applied to perceived, not actual, covariation. Likewise, Jones and Davis (1965) say that “correspondence [between behavior and traits] has nothing to do, necessarily, with the accuracy of the inference” (p. 228), potentially suggesting that (confusingly) Correspondent Inference Theory was itself meant to predict the correct attribution according to coherence criteria and not correspondence criteria.

Though reasoning from faulty premises doesn’t indict the reasoning process itself (Henle, 1962), justifications must be given for how premises should relate to conclusions. Kelley and Michela (1980) note that “judgments about accuracy [read, coherence] are being made implicitly and ... their unstated premises require examination” (p. 494; see also Harvey et al., 1981). Funder (1987) suggests that after becoming frustrated by difficulties assessing correspondence (or accuracy), attribution researchers turned their attention to studying how attributions were made. In so doing, they began to use their theories of how attribution happens as standards for how attribution should happen, and began to treat deviations from their theories as evidence of flawed reasoning.

One solution to the above problem (though not the one that Funder prefers) is to ground normative standards in an external logical system (Morris & Larrick, 1995). Bayesian theories are particularly apt for this purpose. For instance, Ajzen and Fishbein (1975) describe their theory as a normative standard of how an optimal observer would respond. They acknowledge that people’s attributions sometimes lack correspondence (which they refer to as accuracy), but claim that they are still largely coherent (see also Ajzen & Fishbein, 1978). When coherence can be established, problems with the judgment process itself can be ruled out, and lack of correspondence traced to faulty beliefs or perceptions (Hilton, 2007).

Pragmatic Criteria

Though correspondence and coherence may be desirable to researchers, they are not unequivocally useful to people in daily life. Pragmatic criteria evaluate whether a judgment helps the person who makes it achieve useful ends. These ends may go beyond the notion of people as intuitive scientists (Swann, 1984; Funder, 1987; Tetlock, 2002), and may go unnoticed under laboratory conditions (Swann, 1984; Funder, 1987; Andrews, 2001; McKenzie,
2.3. Determining Correctness

Note that while most classical attribution theories are concerned with coherence, their authors did occasionally speculate at pragmatic reasons that apparent biases might exist (e.g., Kelley, 1972a; Jones, 1979).

While it is certainly true that people’s attribution tendencies may suit a purpose that is not captured in laboratory studies, it is still useful to know whether or why these laboratory studies, which still contribute important knowledge about attribution, appear to show problems with the attribution process. Given that there is already sufficient evidence to indict the conclusions of the attribution process as non-correspondent in a variety of settings, the question shifts to why correspondence breaks down. Answering this question requires the ability to diagnose failures of coherence in the process, or to determine that coherence is achieved and that blame rests at the feet of beliefs or perceptions. Therefore, the remainder of this paper focuses on coherence criteria.

2.3.2 Product and Process

Models of human reasoning differ according to how closely they aim to match the actual cognitive processes that people use. Vicente and Wang (1998) and Sun (2008) differentiate product theories, which model expected patterns of inputs and outputs to a reasoning process, from process theories, which model the means by which the inputs are transformed into the outputs. In general, psychologists have been interested in understanding the reasoning process itself, and are particularly concerned with understanding why this process often produces incorrect conclusions. However, this aim presupposes the ability to differentiate what is “correct” from what is “incorrect”. Attempting to simultaneously model the nuances of human cognition while also specifying the judgment that is logically expected is more complicated than is necessary. Thus, this paper’s theory is concerned only with product (in particular, the normatively correct output for a given set of inputs), and leaves process considerations for future research. Other attribution researchers have had similar orientations. For instance, Jones and McGillis (1976) frame Correspondent Inference Theory as a “rational baseline model,” and not as a summary of “phenomenal experience” (p. 404), though their later discussion goes very much in a process direction.

Importantly, while a process theory must match empirical observations, a product theory need not, if the claim to be made is that the product theory prescribes the normative inference. Indeed, the fact that a product theory doesn’t match empirical observations is what gives rise to further investigation. This investigation will either conclude that the product theory is inadequate or inappropriate for the particular inference task, or that it is adequate and appropriate, and that people’s judgments are wrong. Thus, the use of terms like “normative standard” is not to imply infallibility. Rather, it is meant to connote objectivity and testability.

In the past, Bayesian models have been criticized for not adequately describing the human reasoning process. For instance, Fischhoff and Lichtenstein (1978) criticize work
by Ajzen and Fishbein (1975) on the grounds that complex processes can mimic the simple mathematical relationships that the latter authors specified, which Hoffman (1960) refers to as being “paramorphic”. While this criticism can apply to both product and process theories, it is only a limitation with a process theory. In fact, the goal with a normative standard (that happens to be simple) is to show that a possibly quite complex cognitive process does match the simple model.

2.3.3 Scoping

A product theory’s focus on inputs and outputs naturally leads to the question of which inputs and outputs the theory concerns. As in the original introduction of the discounting principle, this theory will assume that the effect to be explained, as well as the candidate causes for that effect, have all been identified, leaving out processes like spontaneous trait (Winter & Uleman, 1984) and causal (Hassin, Bargh, & Uleman, 2002) inferences, as well as more deliberate causal search (Shaklee & Fischhoff, 1982). Additionally, this theory assumes that the behavior that constitutes the effect has already been classified, leaving out considerations of how knowledge of the actor or situation can color how the same behaviors are perceived (Trope, 1986). Doing so is not meant to deny the importance of behavior classification, which plays an important part in trait inference (Trope, Cohen, & Alfieri, 1991) and attribution biases (Gawronski, 2004). However, determining a normative standard for how behavior should be classified is outside of the realm of understanding whether discounting and augmenting are themselves normatively correct.

2.4 Standard for One Uncertain Cause

The following standard specifies the normative confidence in the presence of an uncertain cause, $U$, given the certain presence of a cause $C$ and the occurrence of an effect $E$. Both $U$ and $C$ are potential influences on the observed event $E$, or on each other, and may either facilitate or inhibit $E$. With this notation, the attribution of interest is $P(U \mid E, C)$, or the probability that the uncertain cause was present, given both that the event occurred and that the certain cause was present (cf. Morris & Larrick, 1995).

Note that before $C$’s presence is known, the attribution of interest is $P(U \mid E)$, where according to Bayes’ theorem:

$$
\frac{P(U \mid E)}{P(U)} = \frac{P(E \mid U)}{P(E)}
$$

Writing the identity in this way highlights the symmetry between the two fractions. In particular, the left side of the equation expresses the confidence that $U$ was present given that $E$ occurred, as compared to the prior probability of $U$. The right side of the equation expresses the likelihood that $E$ would occur assuming that $U$ was present, as compared to how likely $E$ is to occur in general. Intuitively, upon witnessing $E$, a rational observer...
should increase (or decrease) his/her baseline confidence that \( U \) was present in proportion to how much \( U \) would increase (or decrease) the likelihood of \( E \) occurring.

The normative value for \( P(U \mid E, C) \) can be found by repeatedly applying Bayes’ rule,\(^7\) and arranging the terms in an intuitive way, resulting in:

\[
P(U \mid E, C) = P(U) \cdot \frac{P(C \mid U)}{P(C)} \cdot \frac{P(E \mid C, U)}{P(E \mid C)}
\]

The left side expresses the ascription for \( U \), given the occurrences of \( E \) and \( C \). The ascription depends on the prior probability of \( U \), and two ratios, which will be called the *cause-cause co-occurrence* and *relative effect likelihood* terms.

The cause-cause co-occurrence (or simply co-occurrence) term takes the ratio of the probability of \( C \) occurring given that \( U \) occurred to the unconditional probability of \( C \) occurring. If \( C \) and \( U \) are independent events, then by definition this ratio will be one. Otherwise, if \( C \) and \( U \) occur together, then the numerator will exceed the denominator, and if they tend not to occur together the denominator will exceed the numerator. Thus, this ratio is intuitively similar to the correlation of the two causes, though with a range of 0 to \( \infty \) rather than \(-1 \) to 1.

The relative effect likelihood (or simply relative likelihood) term concerns the likelihood of the effect given the presence of \( C \), and compares the case with \( U \) present to the case where \( U \)’s presence is unknown. This term captures the fact that the ability to infer that \( U \) did occur depends on the additional impact that \( U \) has over when only \( C \) is known to have occurred. In the simple case where both causes are sufficient, then \( P(E \mid C, U) = P(E \mid C) \), and so this term contributes no information. However, if \( U \) increases the chances of \( E \) occurring over \( C \), then this term will exceed one and make an inference of \( U \)’s presence more likely. On the other hand, if \( U \) and \( C \) operate antagonistically (i.e., they have opposite effects, or one inhibits the other’s effect), then the ratio may be less than one, making it less likely that \( U \) was present.

If both sides of the above identity are divided through by \( P(U) \), then the left hand side shows the change in confidence that \( U \) was present, after the effect occurs and given that the certain cause also occurred. The right hand side nicely parcels the effect of the two causes being related and the effect of the two causes occurring together.\(^8\) Confidence that \( U \) was present should increase if \( C \) and \( U \) tend to occur together, or if the effect is more likely when both causes are present. Confidence should decrease if \( C \) and \( U \) tend not to occur together, or if they make the effect less likely when they occur together. Other combinations of these two terms lead to different predictions depending on the relative sizes of each term.

\(^7\) Despite appearances, these terms are not completely orthogonal since \( P(C) = P(C \mid U)P(U) + P(C \mid \overline{U})P(\overline{U}) \) and \( P(E \mid C) = P(E \mid C, U)P(U \mid C) + P(E \mid C, \overline{U})P(\overline{U} \mid C) \). Nonetheless, it is both heuristically useful and mathematically correct to think of these terms as the incremental impact of knowing \( U \) was present compared to when \( U \)’s presence was unknown.

\(^8\)
2.4.1 Supporting Evidence

A Bayesian analysis helps to highlight the important categories of information that attributors may pay attention to (Fischhoff & Lichtenstein, 1978). Accordingly, the information categories identified in this model are not novel, and a great deal of prior research can be identified that speaks to the importance of each category. The following is a selection of this evidence.

Prior Probability

First, it should be noted that when the prior probability is in the denominator of the left hand side of the equation, the focus moves to the change in confidence that \( U \) is present. The latter case is similar to Ajzen and Fishbein’s prior and posterior odds approach, wherein they admonish researchers to ensure that the prior probabilities of a putative cause are equivalent across experimental conditions. It should be noted that this restriction only applies if, as Ajzen and Fishbein suggest, the attribution is operationalized as the likelihood ratio, and not if the posterior attribution is asked for.

Regardless of where \( P(U) \) is placed, the standard makes it clear that people should be less willing to infer a rare or unexpected cause than a common or expected cause. Indications of this tendency abound in the literature. When causes are traits, evidence that people are more willing to infer expected traits comes from the literature on stereotyping. For instance, Biernat and Ma (2005) compare the amount of evidence people believe is required to infer that an individual possesses a particular trait, and compare traits that are either stereotypic or non-stereotypic for the target. The authors find that people report requiring less evidence to infer that a target has a trait that is stereotypic compared to one that is not. If stereotypes are interpreted as the prior probability of a trait for the target, then these results demonstrate the fact that traits with higher prior probability are inferred more readily.

In the attitude attribution paradigm, the uncertain cause is the essay author’s attitude, the prior probability of which people’s attributions may appropriately reflect. Jones and Harris (1967) found that people attributed anti-Castro attitudes to participants constrained to write anti-Castro essays, but that people attributed neutral attitudes to participants constrained to write pro-Castro essays. The former result may simply reflect the high base rate of anti-Castro attitudes, and the latter result may in fact reflect participants’ having taken into account the low prior probability of pro-Castro attitudes, a complete neglect of which should have resulted in a pro-Castro attribution, not a neutral attribution. Chapter 3 uses this paper’s model to postdict this and other findings, showing that when the prior attitude distribution is skewed, people’s attitude attributions should be similarly skewed. With newly-collected data, Jennings measures people’s prior attitude attributions on an individual level, and finds that people make more neutral attributions when the essay expresses an attitude that they believe is uncommon.

People’s application of base rates in the attitude attribution paradigm appears to work
even when people anticipate how others will reason about their own behavior. Van Boven, Kamada, and Gilovich (1999, Study 1) ask participants to videotape counter-attitudinal speeches. They then asked each participant to guess what attitude an observer would ascribe to him or her, given that this observer was aware that the position to express was assigned. They then showed these videotaped speeches to observers. Both the speakers’ expectations and the observers’ attributions were in line with the speech when it expressed a position perceived to be common, but when the speech expressed a position perceived to be uncommon, both the expected and actual attributions were neutral.

Prior probability may also figure into more abstract causal ascriptions. For example, Hansen and Hall (1985, Experiment 1) ask participants to ascribe the force exerted by a player in a game of tug-of-war (and other similar situations) who is either alone or a member of a team. They find that while the force ascribed to one member of a team decreases linearly with the number of members of that team, the force ascribed to a lone performer increases to an asymptote well below the scale maximum as the number of opponents increases. Though their preferred interpretation is that people are more willing to “discount the many” than “augment the few”, they also acknowledge that their results could reflect people’s prior probabilities of how much force a person can exert, with extreme weakness being more probable than extreme strength. They report some incidental data that support this interpretation.

It should be noted that there is ample evidence that people neglect prior probability (or base rates) in a variety of contexts (Kahneman & Tversky, 1973; Nisbett & Borgida, 1975), possibly due to a conflation inverse probabilities, e.g., $P(A \mid B)$ and $P(B \mid A)$ (Dawes, Mirels, Gold, & Donahue, 1993; Villejoubert & Mandel, 2002). However, the above examples make it clear that prior probability is not universally neglected, and so it will be left for future research to untangle this contradiction.

### Cause-Cause Co-occurrence

Cause-cause co-occurrence refers to the tendency for the two candidate causes to occur together (or not), and not the covariation between either cause and the effect, which is the emphasis of other, inductive models (e.g., Kelley, 1967; Cheng & Novick, 1992). Though the relationship between multiple causes has been studied in the context of causal induction (e.g., Van Overwalle & Timmermans, 2005), such findings are not directly relevant to causal ascription (but, see the discussion in this paper’s conclusions).

Theories of causal ascription have tended to ignore cause-cause co-occurrence for the sake of simplicity, even though people are unlikely to believe that all causes are unrelated (Kelley, 1973). Kelley (1972b) explicitly acknowledged this as a limitation of his Causal Schemata Model, as did Reeder and Brewer (1979) in reference to their Implicational Schemata Model. Ajzen and Fishbein (1975) also limited their Bayesian analysis to independent causes. Fortunately, many studies of discounting and augmenting, being designed by conscientious researchers, have used statistically independent causes (perhaps out of habit of good research design, if not intentionally). However, there are exceptions,
both deliberate and unintentional.

One of the major contributions of Morris and Larrick’s (1995) model was its ability to handle correlated causes. To test this aspect of their model, Morris and Larrick replicated the attitude attribution paradigm, but manipulated the cover story so that authors were allegedly assigned which position to write in a way that was either negatively or positively associated with their attitudes, or that was independent of their attitudes. As expected, they found that people made more confident attributions of a corresponding attitude when the assignment was positively correlated with the author’s actual attitude, and less confident attributions when the assignment was negatively correlated.

An unintentional case of cause-cause co-occurrence can be seen in Rosenfield and Stephan (1977), where a personal cause (level of aggressiveness) is pitted against a situational cause (a discussion group’s attitudes about venting anger), with the target actor’s angry outburst being the effect to be explained. In their second experiment, the authors find that participants discounted (or augmented) the situational cause in accordance to the level of the actor’s disposition, but did not discount (or augment) the dispositional cause in accordance to the level of the group’s attitudes. They suggest that this may be due to a perception that the individual actor’s disposition can be predicted from the group’s attitude, which would counteract any discounting of the actor’s disposition due to the group’s attitude having also encouraged the same effect. They find support for this by controlling for the expected level of aggression for the actor (before the aggressive behavior was witnessed) in the correlation between the perceived strength of the dispositional cause and the perceived strength of the situational cause, in which case they find the expected negative correlation.

Cause-cause co-occurrence may play an important role in everyday causal reasoning. Chapter 4 uses this paper’s model to assess the coherence of trait attributions in the United States and China. The causes listed are taken from a previous study where participants listed possible causes for the target event. Interestingly, many of these causes are perceived to be related (either positively or negatively) with the target dispositional cause, and cause-cause co-occurrence appears to be reflected in people’s attributions.

Relative Effect Likelihood

Most theories of attribution have focused on the relation between knowing what people with particular traits do, and inferring from there what kind of person someone who does a particular thing is; that is, translating from $P(E \mid U)$ to $P(U \mid E)$. The main contribution of early Bayesian theories was to relate the posterior probability of a cause to the prior probability of the cause and the effect, via the likelihood (e.g., Ajzen, 1971; Trope, 1974; Ajzen & Fishbein, 1975). Both Trope (1974) and Ajzen and Fishbein (1975) took the ratio of these probabilities for two hypothetical causes, Trope comparing two traits, and Ajzen and Fishbein comparing the presence or absence of a cause. In the latter case, this leads to the conclusion that the posterior odds of a cause $P(U \mid E)/P(\overline{U} \mid E)$, equals the prior odds, $P(U)/P(\overline{U})$, multiplied by the likelihood ratio, $P(E \mid U)/P(E \mid \overline{U})$. Ajzen and Fishbein then go on to analyze a variety of attribution problems using verbal arguments about the
relative sizes of $P(E \mid U)$ and $P(E \mid \overline{U})$. For instance, they note that in forced compliance situations, the likelihood ratio will be one only if participants believe that the target has no decision freedom (presuming that $U$ would otherwise be relevant to $E$). Much later, Forsyth (2004) applied the likelihood ratio approach to the attitude attribution paradigm, and found that participants do indeed have likelihood ratios that depart from unity, adding to other evidence that people perceive that essay authors have some residual choice.

While verbal arguments about and direct assessment of likelihood ratios are useful, they obscure more substantial understanding of how multiple causes relate to each other. As already noted, theories like Kelley’s Causal Schemata Model and Reeder and Brewer’s Implicational Schemata Model usefully specify a variety of configurations of two causes and one effect, but fail to account for the inequivalence of posterior probability and likelihood. Once again, a major contribution of Morris and Larrick was to correct this mistake, and to treat causes as not deterministically related to effects. However, while many of their standards do involve the probability of an effect based on the conjunction of the presence and/or absence of both causes, their overall conclusions only highlight the causal sufficiency and association of the two causes. As this paper’s standard makes clear, the more important consideration is the likelihood of the effect when both causes are present, compared to the likelihood when one cause’s presence (or level) is unknown and another’s known.

Note that the standard makes it clear that in the right circumstances, discounting to below the prior is possible, and not, as Morris and Larrick (1995) suggest, extreme overdiscounting. For instance, Kun et al. (1980) find discounting to below a cause’s prior probability, which can be accounted for by the explanation that participants expected a different (e.g., stronger or qualitatively different) effect for two causes than for one cause, making the probability of the weaker effect given both causes lower than its probability when only one is known to be present.

The greater emphasis that this paper’s theory places on $P(E \mid C, U)$ as opposed to $P(E \mid U)$ and $P(E \mid C)$ (or $P(E \mid C, U)$ and $P(E \mid \overline{C}, U)$) highlights the fact that two causes in combination can do more than simply have an additive (or subtractive) impact on the probability of the effect. The latter view has been called a hydraulic view of causation (Gawronski, 2004), and is evident in research methodologies that operationalize causal attribution along a bipolar person and situation (or internal and external) scale. It is also evident in research such as Hansen and Hall (1985), which uses scenarios that directly invoke a competition between countervailing forces, such as a game of tug-of-war. Hansen and Hall frame the goal of discounting and augmenting as extracting the effect that an individual cause would have had were the actual effect not also influenced by multiple other causes. Their scenarios presume that each team member contributes a certain amount of force that corresponds to his/her strength, and that these forces facilitate or inhibit one team’s winning depending on which side of the rope the person is on. However, it is of course possible that people reading Hansen and Hall’s scenarios had an implicit understanding of social loafing (Latane, Williams, & Harkins, 1979), a case where a cause’s impact (one person’s contribution) changes given the presence of other causes (other people’s contributions). This is another alternative explanation for their finding that people more readily “discounted the
many” by estimating lower forces exerted by many teammates on one side of the rope than “augmented the few” by believing that a lone person opposing this team was extraordinarily strong (which was previously also discussed relative to the prior probability of extreme strength vs. extreme weakness).

A contrast to the hydraulic view of causation is the interactive view of causation (Gawronski, 2004). Like a statistical interaction, this view conceives of causes as working differently in combination than when alone. A simple example of interactive causation is Kelley’s Multiple Necessary Causes (MNC) schema, wherein an effect only occurs if both causes are present, such as in achievement tasks, wherein both effort and ability are necessary to achieve a good outcome. However, more complex relationships are also possible. Because of the special importance of interactive causation, this topic will be treated in greater detail later in the paper.

2.4.2 Comparison with Morris and Larrick

Morris and Larrick (1995) reconsider Kelley’s (1972b) causal schemata theory using a series of Bayesian arguments. They break their analysis down by the sufficiency of the causes, the independence of the causes, and the number of causes (i.e., whether there are other causes besides the two being considered). For details of the comparison, see Appendix A. In summary, this paper’s model agrees with the prescriptions of Morris and Larrick when it is assumed that \( C \) and \( U \) are sole causes (unless either \( C \) or \( U \) are jointly necessary, or one is a side effect of the other, or of the effect, \( E \)), or that they are sufficient and independent, in which case discounting is always implied. However, when there are causes other than \( C \) and \( U \), the models disagree. When all causes are independent, but not sufficient, the Morris and Larrick criteria incorrectly state that discounting is normative, where in fact that needn’t be the case. When all causes are sufficient but not independent, the Morris and Larrick criteria agree with this paper’s criteria, though this paper’s criteria are more intuitive. Finally, in the general case, the Morris and Larrick criteria fail to highlight the potential interactions between \( C \) and \( U \) in producing the effect.

In addition to being more intuitive, the present model has several other advantages. First, the judgments necessary for calculating normative attributions can be collected for any pair of causes, without needing to decide in advance which of the conditions (sufficiency, independence, and number) apply. Second, the model can make precise predictions for any kind of causal interaction, not just variations in necessity and sufficiency. Finally, focusing on prior probability, cause-cause co-occurrence, and relative effect likelihood leads to cleaner predictions than focusing on sufficiency, independence, and number. Thus, while credit must be given to Morris and Larrick for rigorously examining the ideas in Kelley’s (1972b) causal schemata theory, departing from Kelley’s approach appears to have been more fruitful.
2.5 Empirical Demonstration

This study aims to demonstrate the model’s aptness as a normative standard for causal ascription. While the model’s mathematical adequacy and suitability to the participant’s judgment task are logical rather than empirical questions (cf. Jones & McGillis, 1976, p. 404), establishing that the model is useful does require showing that it is possible to elicit the quantities required to assess coherence, and that people’s judgments cohere reasonably well according to the model. In so doing, the study will demonstrate approaches that other researchers can apply.

Importantly, this study is not designed to investigate when and why people fail to reason normatively. First, as mentioned previously, this model concerns coherence among beliefs, and not correspondence to ground truth. Given the ample evidence that people’s attributions lack correspondence, this model is more useful for determining whether failures of correspondence stem from flawed assumptions and perceptions, or from flawed reasoning (lack of coherence). Second, given the aim of showing the aptness of the model, this study is designed to enhance (though not necessarily maximize) people’s chances of showing coherence. Thus, the study’s results may reflect more coherent reasoning than would be expected under ecologically valid conditions. On the other hand, it may prove difficult for people to reliably estimate the quantities needed to apply the standard. Thus, this study is better viewed as a demonstration of the model than as a test of people’s reasoning, and it is not a “test” of the model itself.

2.5.1 Study Design

Biases in discounting and augmenting have been explained in terms of more general processes like anchoring and adjustment (Jones, 1979; Quattrone, 1982; Johnson, Jemmott, & Pettigrew, 1984). Given the robustness of the latter phenomenon, measuring an initial attribution and then an adjusted attribution may impair coherence, and would create ambiguity about the source of the problem (background beliefs vs. anchoring and adjustment). What’s more, the effect may be categorized differently before and after the certain cause is introduced. Given that discounting has been studied both with sequential and simultaneous cause presentation (McClure, 1998), and given that the primary aim is to demonstrate the applications of the standard for mixed certainty \( P(U \mid E, C) \), the initial attribution \( P(U \mid E) \) will not be measured in this study.

In order to demonstrate the flexibility of the model, the same uncertain cause, \( U \), will be pit against the same effect, \( E \), and several certain causes, \( C_i \). The certain causes will attempt to fully cross cause-cause co-occurrence and relative effect likelihood ratios that are less than one, equal to one, or more than one, creating a \( 3 \times 3 \) design. While this does make it feasible to analyze the results with a factorial analysis of variance (ANOVA), the fact that it is quite challenging to imagine causes where co-occurrence and relative effect likelihood have opposite directions (e.g., a certain cause that is less likely to occur given the uncertain cause, but that when paired with the uncertain cause is more likely to produce
the effect) suggests that this manipulation won’t be equally effective across conditions. Therefore, the more telling analysis will be to compare the ascription of the uncertain cause with the co-occurrence and relative likelihood (the prior probability of $U$ is the same across $C_i$). Attempting to cover the range of configurations should minimize problems associated with restriction of range when analyzing the data.

Using different stimuli, Chapter 4 applies this paper’s standard cross-culturally. In his study, participants estimate each of the probabilities in the standard directly, and where applicable, once for each of 22 causes. He finds very little coherence when one individual’s beliefs for one certain cause are compared to that individual’s ascription. However, correlations among each individual’s normative and actual ascriptions across all 22 causes are reasonably good (mean $r \approx .45$), and the aggregate predictions (averaging within cultural group) are also reasonably good ($r \approx .8$). Still, it would be better to be able to find coherence on an individual-by-individual and cause-by-cause basis. This study will make two modifications to the approach of the prior study in order to improve coherence.

The first difference between this study and Chapter 4 will be the use of a between subjects design to manipulate the certain cause. This decision may either help or hurt coherence. On the one hand, Kahneman (2003) suggests that within subjects designs encourage deliberate reasoning, which should help coherence. On the other hand, people may have difficulty keeping the effects of the certain causes separate in their minds, and may fatigue after making as many judgments as this would require, thereby hurting coherence. Though future work should systematically test the effects of between or within subjects elicitation, the present work will be done between subjects.

The second difference between this study and Chapter 4 is the elicitation approach. The latter study asked participants to directly estimate each of the probabilities in the model, which is not something that most people are used to doing, nor is it something that people can do with terrific accuracy (see O’Hagan et al., 2006). However, when both sides are divided by $P(U)$, the key judgments in the model are all relative, each ratio expressing a judgment about the relative likelihood of an event when there is more or less information available. By taking the log of both sides, each ratio becomes a judgment about whether the numerator or denominator is more likely, and by how much:

$$\log \frac{P(U \mid E, C)}{P(U)} = \log \frac{P(C \mid U)}{P(C)} + \log \frac{P(E \mid C, U)}{P(E \mid C)}$$

Because the log of a ratio is symmetric about zero, with zero indicating that the numerator and denominator of the ratio are equal, a Likert-type scale with a zero point that clearly means indifference between the two statements can feasibly be used to represent each ratio.

An additional advantage to using a log-relative format is that judgment reliability should not be affected by the base rate of the events. For example, if $C$ is a reasonably common event, then participants can express suitably subtle degrees of difference between $P(C \mid U)$ and $P(C)$ (e.g., 45% vs. 40%). However, if $C$ is a very rare event, then participants may run out of quanta to differentiate each, and may end up making the two
judgments equal (e.g., 1% vs. 1%) or overly distinct (e.g., 2% vs. 1%). As a result, coherence would be confounded with cause or effect base rates. While different authors have explored either increasing the number of quanta by increasing the denominator of the judgment (Yamagishi, 1994a, 1994b, 1997) or by using an inset “magnified” scale for smaller probabilities (Woloshin, Schwartz, Byram, Fischhoff, & Welch, 2000), the log-relative format avoids the need for such changes altogether.

In this approach, the target ascription is \( P(U \mid E, C)/P(U) \). This necessarily orients the participant to the importance of the base rate, and likewise controls for individual (or cultural) differences in the perceived base rate of the uncertain cause (cf. Chapter 4). As already stated, some authors have recommended that attributions be judged relative to the base rate or prior probability of the cause or trait in question (Ajzen & Fishbein, 1975; Jones & McGillis, 1976), and attribution has sometimes been operationalized this way (e.g., Ginzel, Jones, & Swann, 1987), though there are detractors (Fischhoff & Lichtenstein, 1978). Regardless, in light of evidence that people neglect base rates or rarely have stable base rate expectations (Koehler, 1996), the log-relative format should once again increase coherence.

Other authors have used log-relative formats to elicit subjective probabilities. For instance, Griffin and Buehler (1999, Experiment 3) replicate the lawyer-engineer problem (Kahneman & Tversky, 1973), and ask participants to rate the extent that a given description “matches” a lawyer or engineer, i.e., \( P(D \mid E) \) or \( P(D \mid L) \). Since by construction the population consists only of lawyers or engineers, the description must match one or the other. Using a scale with endpoints of “Exactly like my image of a lawyer” and “Exactly like my image of an engineer” and a midpoint of “Equally like a lawyer and an engineer” (p. 70), they divide people’s ratings by their complement (i.e., a 7 our of 10 is divided by 10 - 7 = 3); and treat this as the ratio of \( P(D \mid L) \) and \( P(D \mid E) \).

Kunda and Nisbett (1986) and then Roese and Morris (1999) used a somewhat similar approach to measure the perceived covariation between two variables. Their approach lists a fact of the form Person A exceeds Person B on variable one, followed by two predictions: Person A exceeds Person B on variable two, or Person B exceeds Person A on variable two. Participants then rate the relative likelihood of each prediction using a scale whose anchors are that prediction one or prediction two is very likely, and whose midpoint indicates indifference between the two. This response is treated directly as an index of perceived covariation between variables one and two.

Drawing on the above two examples, in this study each judgment will compare two hypothetical individuals, one of whom will be more specific and represent the numerator of the ratio, and the other of whom will be less specific and represent the denominator of the ratio. Participants will be asked to judge which person is more likely to fit a certain description, do a certain thing, etc., and to respond on a scale whose midpoint is clearly labeled “equally likely” and whose two directions indicate how much more one or the other person is to satisfy the given statement (see the materials section below).
Table 2.1: Characteristics of the samples in the empirical demonstration.

<table>
<thead>
<tr>
<th></th>
<th>East Asian Descent</th>
<th>European Descent</th>
<th>Other/Mixed Descent</th>
<th>Entire Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>98</td>
<td>51</td>
<td>85</td>
<td>234</td>
</tr>
<tr>
<td>Age</td>
<td>21.8 (4.18)</td>
<td>21.7 (2.52)</td>
<td>21.7 (2.80)</td>
<td>21.7 (3.39)</td>
</tr>
<tr>
<td>Years in US</td>
<td>11.67^1 (6.63)</td>
<td>9.62^2 (6.63)</td>
<td>11.84^3 (6.77)</td>
<td>—</td>
</tr>
<tr>
<td>Females</td>
<td>53%</td>
<td>59%</td>
<td>52%</td>
<td>54%</td>
</tr>
</tbody>
</table>

NOTE: 1 — Among 43 cases; 2 — Among eight cases; 3 — Among 23 cases

2.5.2 Methods

Participants

Participants were all members of an undergraduate Cultural Psychology class at the University of California, Berkeley, and completed the experiment for partial course credit. Participants received an e-mail inviting them to participate in the study, and reminding them of the course credit. They were given a five day period in which to complete the study. Out of 308 students in the class, a total of 250 people (81%) completed the study. A mistake in condition six made those data unusable (N = 28), and so another e-mail was sent to the class giving people who had not previously participated another opportunity, resulting in 24 additional participants over a 24 hour period. These two samples were then combined. After excluding two cases for having missing values on one of the four ratio estimates, the sample had N = 245.

Since the study was administered online from a location of participants’ choosing, there is a concern that some participants may have answered the questions thoughtlessly in order to obtain the course credit. In order to detect such cases, the time spent across the first three pages of the survey (the fourth page was a repetition of the first) was totaled, and the histogram examined. The median time was 160 seconds, with lower and upper quartiles of 120 and 219 seconds. Looking at a histogram of the lower quartile, there was a clear break at 60 seconds (20 seconds per page), and so a decision was made to exclude these cases (11 in total). These cases were fairly evenly spread across the scenarios, and were proportionally split across demographic categories. With these cases removed, the final sample was of size N = 234.

While this study is not cross-cultural in nature, the diversity of the student body and the topic of the class suggest that cultural effects could occur. Therefore, participants were subdivided into three ethnic groups, East Asian(-American) (42%), European-American (22%), and other or multiple ethnicities (36%). Overall and sub-sample characteristics are in Table 2.1.
Materials

The study materials consisted of a scenario wherein the target event \((E)\) is that a student in a large lecture class initiates a conversation on his/her mobile phone as a guest speaker is talking. The uncertain cause \((U)\) was that this student has a “rude” personality, which was clarified to mean that the student tends to behave in a rude manner across a variety of situations. The certain causes \((C)\) were varied in an attempt to fully cross co-occurrence and relative effect likelihood, as shown in Table 2.2. Kelley (1972a) cautioned that making between-subjects comparisons when studying discounting requires that the manipulation not affect how the behavior itself is seen, which would confound differences in the cause with difference in the behavior. The causes presented in this study will almost certainly change people’s impression of the behavior described. However, because this occurs at the beginning of the experiment, it should affect all of the judgments evenly, and hence not affect coherence.

For each judgment, participants were told about two hypothetical students from the class. The first student was always described more specifically, and the second more generally. For instance, when judging co-occurrence, the student representing \(P(C \mid U)\) was described with “You know that this person is considered to have a rude personality”, while the student representing \(P(C)\) was described with “You have no idea whether this student is considered to have a rude personality.” Each student was given a letter (e.g., Student A, Student B), and letters were not reused across judgments. Participants were told to think about which student was more likely to fit a particular description (e.g., “which student you think would be more likely to volunteer to call for help if the projector broke”), and then to respond using a Likert scale from \(-9\) to \(9\). The negative end of the scale was labeled with variants of “much more likely for Student B than Student A”, while the positive end was labeled with variants of “much more likely for Student A than Student B”. The midpoint of the scale was clearly labeled “equally likely for Students A and B”.

Procedure

Participants completed the study online, and as part of a larger collection of unrelated studies. Before this study, participants first completed one page of another study and a demographic questionnaire. Then, the software randomly assigned participants to one of the nine study conditions, plus one of two question orders for co-occurrence and relative likelihood. Given the ethnic diversity of the class, a decision was made to have the software automatically balance the nine conditions separately for the East Asian(-American), European(-American), and other/mixed ethnicity groups.

Before beginning the experiment, participants were informed of the number of questions, and were advised that the answers to some questions might seem blatantly obvious, while the answers to other questions might be very hard to determine. This was done to forestall confusion about co-occurrence ratings for some of the causes, which are plainly unrelated to the uncertain cause (cf. Wright & Wells, 1988; Schwarz, 1994). Participants
Table 2.2: Certain causes ($C$) used in the nine scenarios for the empirical demonstration, shown with their intended interpretations.

<table>
<thead>
<tr>
<th>No.</th>
<th>Co-occ.</th>
<th>Rel. Like.</th>
<th>Cause</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$&lt; 1$</td>
<td>$&lt; 1$</td>
<td>Before the talk, the student had agreed to help the speaker make a point about politeness.</td>
</tr>
<tr>
<td>2</td>
<td>$&lt; 1$</td>
<td>$= 1$</td>
<td>The student had volunteered to call for help after the projector broke.</td>
</tr>
<tr>
<td>3</td>
<td>$&lt; 1$</td>
<td>$&gt; 1$</td>
<td>The student wanted to make sure his/her roommate woke up in time for a big midterm.</td>
</tr>
<tr>
<td>4</td>
<td>$= 1$</td>
<td>$&lt; 1$</td>
<td>The speaker had asked all of the students to do something they wouldn’t ordinarily do.</td>
</tr>
<tr>
<td>5</td>
<td>$= 1$</td>
<td>$= 1$</td>
<td>The student received a text message from his/her dying mother.</td>
</tr>
<tr>
<td>6</td>
<td>$= 1$</td>
<td>$&gt; 1$</td>
<td>The speaker had asked all of the students to text message a friend with a website to visit.</td>
</tr>
<tr>
<td>7</td>
<td>$&gt; 1$</td>
<td>$&lt; 1$</td>
<td>The class had nominated the student as being particularly rude, and the speaker told the student to call someone he/she had offended and apologize.</td>
</tr>
<tr>
<td>8</td>
<td>$&gt; 1$</td>
<td>$= 1$</td>
<td>The student didn’t realize that the speaker had started talking.</td>
</tr>
<tr>
<td>9</td>
<td>$&gt; 1$</td>
<td>$&gt; 1$</td>
<td>The student wanted to tell a friend something embarrassing that the speaker had just said.</td>
</tr>
</tbody>
</table>
first read about the general scenario, and then read that one particular student had performed the behavior (making the phone call), and that the particular situation also applied. They then judged \( P(U \mid E, C)/P(U) \) by comparing the student just described to a randomly-chosen student in the class about whom nothing was known. On a separate screen, they then judged either \( P(C \mid U)/P(U) \) as described previously, or judged \( P(E \mid C, U)/P(E \mid C) \) by comparing a person who was in the situation and was known to have a rude personality to a person who was also in the situation, but about whom nothing else was known. These students were compared on how likely they would be to perform the target action. Participants then completed the remaining judgment on a third screen. For each judgment, it was made clear that the two people being considered were distinct from any of the other people who had been described.

Finally, as a way to see whether thinking about co-occurrence and likelihood affected people’s thinking about the attribution, participants were presented with the original rating task one more time. They were reassured that it was perfectly fine if their answer was the same as what they originally gave, and that it was also perfectly fine if their answer was different than what they originally gave. After answering this question, participants completed other unrelated studies. Participants were given a written debriefing after the survey website was closed at the end of the participation window.
2.5.3 Results and Analysis

Manipulation Checks

The intention of the nine scenarios was to manipulate the cause-cause co-occurrence and relative effect likelihood ratios. In order to test the success of these manipulations, both the appraisals of co-occurrence and likelihood were subjected to two tests. First, each test was run with a $3 \times 3$ ANOVA, which ideally would show a significant main effect of the relevant factor and no other significant effects. Second, ratings for each scenario were compared to zero with a series of $t$-tests.

For co-occurrence, the ANOVA results were mostly as desired: the main effect of the co-occurrence ratio manipulation was significant, $F(2, 225) = 12.37, p < .0001$, and the interaction was not significant ($p = .68$). However, there was a marginally significant effect for the likelihood, $F(2, 225) = 2.72, p = .07$. Though the direction of the manipulations relative to each other was generally successful, not all of the scenarios had the desired effect when compared to zero (i.e., log of one). As seen in Table 2.3, it appears that the effects were in the proper order, but shifted towards higher co-occurrence.

The manipulation of likelihood was less successful. Neither main effect was significant ($ps > .26$), but the interaction was significant, $F(4, 225) = 5.38, p < .001$. Examination of the individual means revealed that in fact every scenario except for scenario six resulted in a relative effect likelihood ratio that was significantly greater than one, and the exception for scenario six could be due to sample size.

Change in Attribution

The attribution was elicited both before and after the other two judgments. While the hope with such measurements is always that they will be the same, it is reasonable to expect that the attributions will change more when the two manipulations have opposite effects than when they have compatible effects. Thus, the scenarios labeled “?” in the table are expected to show the most change and the scenarios labeled “<< 1” and “>> 1” are expected to show the least change. To test for change, each person’s “before” attribution was subtracted from their “after” attribution. Overall, this difference trended toward significance, $t(233) = 1.62, p = .11$, Cohen’s $d = .11$. These differences were then compared to zero within each group, as shown in the table. As can be seen, the predictions were not supported, although given that the manipulations also did not work as intended, this is not surprising.

Attribution

Given that the manipulation of the likelihood ratio did not work as intended, the original predictions for the attributions in each scenario are likewise unlikely to work. On the basis of the results obtained in the manipulation check, new predictions were made by summing the effect of the two appraisals (see Table 2.3). The table also shows the results of per-scenario $t$-tests. As can be seen, the revised predictions (which were made after looking
Table 2.4: Aggregate model comparisons.

<table>
<thead>
<tr>
<th></th>
<th>Co-occ.</th>
<th>Rel. Like.</th>
<th>$R^2$</th>
<th>$F$</th>
<th>Co-occ.</th>
<th>Rel. Like.</th>
<th>$R^2$</th>
<th>$F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>0.712*</td>
<td>0.279</td>
<td>0.676</td>
<td>7.31*</td>
<td>0.498*</td>
<td>0.431*</td>
<td>0.856</td>
<td>20.86*</td>
</tr>
<tr>
<td>Coeffs Equal</td>
<td>0.419*</td>
<td></td>
<td>0.630</td>
<td>13.62*</td>
<td>0.453*</td>
<td></td>
<td>0.855</td>
<td>47.20*</td>
</tr>
<tr>
<td>Change</td>
<td></td>
<td>0.046</td>
<td>1.00</td>
<td></td>
<td></td>
<td>0.001</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>Without 5</td>
<td>0.610*</td>
<td>0.445*</td>
<td>0.831</td>
<td>14.76*</td>
<td>0.441*</td>
<td>0.524*</td>
<td>0.913</td>
<td>31.58*</td>
</tr>
<tr>
<td>Coeffs Equal</td>
<td>0.503*</td>
<td></td>
<td>0.825</td>
<td>32.98*</td>
<td>0.495*</td>
<td></td>
<td>0.911</td>
<td>72.02*</td>
</tr>
<tr>
<td>Change</td>
<td></td>
<td>0.006</td>
<td>0.22</td>
<td></td>
<td></td>
<td></td>
<td>0.002</td>
<td>0.13</td>
</tr>
</tbody>
</table>

NOTE: † $- p < .10$, * $- p < .05$. For $F$-tests, $df_{num.} = 2, 1, 1$ for constrained, unconstrained, and change models, and $df_{denom.} = 8$ for the overall set and $7$ when scenario five is excluded.

at the manipulation checks but before examining the attributions) held in both the before and after ratings for all but scenarios five, seven and eight, although scenarios seven and eight were marginally significant in the correct direction. Scenario five appears to have resulted in the intended attribution, both before and after eliciting the other beliefs, but these attributions are not coherent with the other beliefs. This will be discussed more later.

Aggregate Coherence

Though the eventual goal is to be able to detect coherence on an individual level, aggregate coherence (i.e., coherence among the mean judgments) is an easier goal to attain, and hence sets a limit on what should be expected at the individual level. Therefore, aggregate coherence is examined first.

If the response scales worked as intended, and if people’s judgments are coherent, then the sum of the co-occurrence and likelihood log-ratios should equal the attribution log-ratio. To test this, the mean co-occurrence and likelihood responses per scenario were subtracted from the corresponding attribution responses, and the results compared to zero. This approach failed to find coherence in both the before, $t(8) = -4.61, p < .01$, and after, $t(8) = -5.18, p < .001$, judgments. However, this is a highly stringent test that makes strong assumptions about how people used the response scale. A less stringent test is to predict the attribution from the co-occurrence and likelihood responses, omitting the intercept. This approach was much more successful. Table 2.4 shows linear models predicting the before and after attributions from co-occurrence and likelihood. In order to see whether people made differential use of co-occurrence and likelihood, each of these models is compared to a model where the regression coefficients are constrained to be equal. Finally, given that scenario five appears to have been unusual, the above models are also run with scenario five excluded.

The regression models in Table 2.4 show several things. First, overall, the before and after models all explain significant amounts of variance, which ranges from $R^2 = .63$ to
Table 2.5: Coherence of individual before and after judgments, by scenario and overall, as judged by the attribution minus the mean of the co-occurrence and likelihood.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Before</th>
<th></th>
<th></th>
<th>After</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$t$</td>
<td>$df$</td>
<td>$p$</td>
<td>$M$</td>
<td>$t$</td>
</tr>
<tr>
<td>1</td>
<td>0.77</td>
<td>1.01</td>
<td>27</td>
<td>0.32</td>
<td>0.84</td>
<td>0.41</td>
</tr>
<tr>
<td>2</td>
<td>-1.29</td>
<td>-1.93</td>
<td>27</td>
<td>0.06</td>
<td>0.44</td>
<td>0.67</td>
</tr>
<tr>
<td>3</td>
<td>-0.27</td>
<td>-0.37</td>
<td>25</td>
<td>0.71</td>
<td>-0.04</td>
<td>-0.06</td>
</tr>
<tr>
<td>4</td>
<td>1.00</td>
<td>2.08</td>
<td>24</td>
<td>0.05</td>
<td>0.10</td>
<td>0.92</td>
</tr>
<tr>
<td>5</td>
<td>-3.00</td>
<td>-3.71</td>
<td>26</td>
<td>0.00</td>
<td>-1.52</td>
<td>-2.29</td>
</tr>
<tr>
<td>6</td>
<td>1.07</td>
<td>1.27</td>
<td>20</td>
<td>0.22</td>
<td>0.64</td>
<td>0.88</td>
</tr>
<tr>
<td>7</td>
<td>-0.84</td>
<td>-1.09</td>
<td>24</td>
<td>0.29</td>
<td>-0.80</td>
<td>-0.95</td>
</tr>
<tr>
<td>8</td>
<td>-1.29</td>
<td>-1.62</td>
<td>25</td>
<td>0.12</td>
<td>-1.13</td>
<td>-1.56</td>
</tr>
<tr>
<td>9</td>
<td>0.29</td>
<td>0.35</td>
<td>27</td>
<td>0.73</td>
<td>0.39</td>
<td>0.57</td>
</tr>
<tr>
<td>All</td>
<td>-0.43</td>
<td>-1.67</td>
<td>233</td>
<td>0.10</td>
<td>-0.16</td>
<td>-0.70</td>
</tr>
</tbody>
</table>

$R^2 = .91$. Second, the models where the coefficients are constrained to be equal fit no worse than the unconstrained models. Third, removing scenario five appears to make a substantial improvement in fit, particularly for the before models. Finally, with scenario five removed, both the before and after constrained models appear to have regression coefficients of about .5, suggesting that people’s attributions are the simple average of their two other ratings.

**Individual Coherence**

The aggregate analysis strongly suggests that people use the scale in such a way that their co-occurrence and likelihood estimates average to form the attribution. While this does not comport with a strict interpretation of the scale as a log-ratio (the appropriate coefficient would instead be one, corresponding to an exponent of one on each term), it is quite reasonable to suppose that people adjusted their interpretation of the scale’s anchors according to the judgment (and likewise a bit unrealistic to assume that people think in a strictly uniform way about log-scales). Therefore, it seems reasonable to suppose that people’s attributions might cohere with the average of their co-occurrence and likelihood judgments.

To test this interpretation, the average of each person’s co-occurrence and likelihood ratings were subtracted from their before and after attributions. Then, these deviations were compared to zero, both overall and scenario-by-scenario. The results are shown in Table 2.5. As can be seen, scenario five lacked coherence (showed bias) in both the before and after judgments. In the after judgments, none of the other scenarios showed bias, while the before judgment for scenario four was significantly different from zero, and others trended in that direction. Also, there is a trend toward rejecting the null hypothesis for all judgments taken together in the before attribution, but not in the after attribution.
Cultural Differences

To test for cultural differences, the above deviations were subjected to a 3 (cultural group) by 9 (scenario) ANOVA. For the before judgments, there was a significant main effect of scenario (as should be expected based on the forgoing discussion), $F(8, 207) = 2.61, p < .01$, but no other effects were significant ($ps > .35$). For the after judgments, none of the effects were significant ($ps > .35$), including the culture main effect ($p ≈ .6$).

The above test shows no cultural differences in the (lack of) bias. However, there could also be cultural differences in the average amount of error. To test this, the above tests were repeated, but with the absolute values of the deviations. There were no significant effects for the before judgment ($ps > .22$). For the after judgment, there was a marginally significant effect of culture, $F(2, 207) = 2.53, p = .08$. This was examined using contrast codes. The first contrast compared the East Asian and European groups, and was significant, $t(207) = 2.00, p < .05$. The second contrast compared the average of the East Asian and European groups to the other/mixed ethnicities group, and was not significant ($p = .5$). The value of the first contrast coefficient was .94, which implies a larger average absolute deviation in the European group compared to the East Asian group. However, this effect was not anticipated and did not occur in the before judgments, so it is merely noted for future reference.

2.5.4 Discussion

Even though the manipulations of co-occurrence and relative effect likelihood did not work entirely as intended, the overall analysis was quite successful. First, the aggregate analysis strongly suggested that people’s attributions were based on the average of their co-occurrence and likelihood estimates. While averaging is not what would be predicted based on the model, it has a ready interpretation, and also suggests that people weighted co-occurrence and likelihood equally. The aggregate model’s explained variance was as high as 91% when the fifth scenario, an outlier, was removed.

At the individual level, using the norm suggested by the aggregate analysis found very little evidence of bias, particularly in the attributions that were completed after the co-occurrence and likelihood judgments. Likewise, there was no suggestion of cultural differences in the amount of bias. The amount of error (operationalized as absolute deviation from the standard) did not appear to differ across scenarios, and except for one unanticipated contrast in the after judgment, did not show cultural differences.

While the overall analysis leads to the conclusion that people’s judgments were largely coherent, two factors require further discussion. First, the choice of coherence standard requires further discussion. Second, the obvious lack of coherence for the fifth scenario, as well as the general failure of the likelihood manipulations, requires additional consideration.
Determining coherence

As mentioned during the analysis, the strictest interpretation of the identity would require simply adding the co-occurrence and likelihood terms to get the normative attribution. However, given that the response format used in this study is relatively novel, being so strict was not warranted. Instead, an aggregate analysis, which has the advantage of reducing error in each estimate, was performed first. This analysis fortuitously showed that both coefficients equaled approximately one half, which was applied as a standard in the individual analysis. While this decision may seem opportunistic, it is not, for two reasons.

The first reason that this was not merely an opportunistic choice is that the alternative, running a linear regression, would perforce have arrived at coefficients that fit the data better than did one half. Thus, any choice of coefficient would be worse than the regression coefficients. The fact that one half showed relatively little bias is more impressive than would have been a finding that the regression model showed relatively little bias—since it is by construction an unbiased estimator!

The second reason that this choice was justified is that one half corresponds to an averaging of the co-occurrence and likelihood judgments, which is a psychologically plausible interpretation of how people would use the scale, and certainly more plausible than coefficients of .4 or .8, etc. However, it is not defensible to simply apply the coefficients from the aggregate analysis whatever they may be; instead, the fact that the results were close to one half is what made this choice reasonable.

Moving forward, future research will need to replicate these findings, and more detailed statistical work needs to be done to understand how people use this response scale. In addition to trying different scenarios, variations to the elicitation context may also prove important. For instance, if the three judgments were elicited on one page, it could be that people would use more of an additive model. Another useful test would be to elicit attributions dependent on either one judgment (via the normal use of Bayes’ theorem) or three judgments, as follows:

\[
\frac{P(U \mid E, C)}{P(U \mid E)} = \frac{P(E)}{P(E \mid C)} \cdot \frac{P(C \mid U)}{P(C)} \cdot \frac{P(E \mid C, U)}{P(E \mid U)}.
\]

The above identity would be directly interpretable as the amount of discounting, and involves three judgments that all depend on the choice of \(C\). The expected outcome is that an aggregate analysis as done here would result in a coefficient of one for the simple use of Bayes’ theorem, and coefficients of one third (adjusting for sign) in the above equation.

Other variations to the response scale should also be tried. For instance, the scale is currently intended to be a large enough subset of what in fact is an unbounded range. Another approach would be to bound the scale, labeling the extremes with variants of “Definitely true for Person A and definitely not true for Person B” or the reverse. Mathematically, this would make the endpoints equal to \(\pm \infty\). People’s responses could then be scaled to \([-1, 1]\)

---

9 Any such work should keep in mind that restriction of range in the co-occurrence and likelihood manipulations may lead to attenuated parameter estimates.
and mapped to \((-\infty, \infty)\), using the arc-hyperbolic tangent (i.e., Fisher’s \(r\)-to-\(Z\)) or some other sigmoidal transform.

**Informational manipulations**

While this study did find reasonable coherence (i.e., very little bias, no scenario-by-scenario differences in error), the specific content of some of the findings is surprising. First, though co-occurrence was manipulated reasonably well, none of the manipulations succeeded in suggesting an action that was less likely for a rude person than for the average person. For instance, scenarios two (volunteering to call for help with a broken projector) and three (wanting to make sure a roommate doesn’t oversleep for an exam) are both courteous behaviors, and were intended to seem less likely for a rude person. Particularly if being rude is a rare trait, this should have resulted in a strongly negative (rather than null) rating. Though the ratings may coherently reflect people’s beliefs, it could also be that people failed to incorporate base rates properly when judging how likely the behavior was for an average student.

The likelihood manipulations were even less successful, which appears largely to be due to the failure to produce certain causes that would inhibit the effect. Scenario one (where the student is the speaker’s accomplice) appears simply to be an honest failure—the original thinking was that a rude person might refuse to go through with the demonstration, but people presumably believed that the rude person would relish in it. Scenarios four and seven resulted in relatively smaller likelihood judgments, but they still signified a positive relation. Scenario four, which was intended to evoke a logical calculus (this person has done something they wouldn’t ordinarily do by making the phone call) may have been overridden by other justifications (e.g., a rude person would use the speaker’s request as an excuse, and disregard the spirit of the request). Scenario seven was again subtle, with the thought being that a rude person would refuse to comply with the speaker’s request. However, participants may instead have focused on the power of a classroom full of people awaiting your compliance.

Based only on the above examples, one might simply conclude that the manipulations were poorly chosen. This was not for lack of trying, and so could suggest that there is an actual scarcity of causes that both co-occur with and countervail a trait (or a shortage of cleverness in imagining them). However, some vindication comes at the failure of scenario five, where the student receives a text message from his/her dying mother. The intent with this scenario was that people would (a) recognize that rudeness is independent of getting the text message, and (b) believe that the fact that the mother was dying would make everyone equally likely to make a phone call (the assumption being that the mother may be on death’s door). People’s actual attributions, both before and after the other judgments, fit with the intended manipulation: people did not perceive the actor as any more or less likely to be rude than anyone else. Likewise, the co-occurrence manipulation worked as intended, as the ratings were as expected. However, people’s likelihood ratings were significantly greater than zero, accounting for why people’s judgments for this scenario lacked coherence. The ques-
tion then becomes which judgment is responsible for the impaired coherence. One possible explanation is that people’s likelihood judgments relied on the representativeness of the action (making a call in the midst of the speaker’s talk) for the trait (rudeness) (Kahneman & Tversky, 1973), and failed to adequately consider the urgency of a dying mother’s message—something that appears to have been incorporated into the attributions themselves.

In summary, the initial attempt at a somewhat novel elicitation format for use with this paper’s standard was successful. However, questions remain about how to properly treat the scale responses vis-a-vis the logarithm of two judgments, and the choice of an averaging rather than an additive standard needs to be corroborated with more data. Additionally, common peccadilloes of probabilistic judgment (neglect of base rates, overreliance on representativeness) do seem to have crept in, so more tinkering and study is needed in order to create a format that will maximize people’s coherence. Thus, the empirical demonstration has served its purpose, and suggested new questions in the process.

2.6 Modeling Causal Mechanisms

Treating attribution as a probability judgment about a discrete event is a useful simplification that allows the application of Bayesian standards, but it is not without its problems.

The first problem with a discrete probabilistic treatment is that it does not match the judgments that people make in most experimental paradigms. For instance, though the attitude attribution paradigm can be approximated by asking participants to judge the likelihood that the author holds one particular attitude, the attitude expressed in the essay, or some variation thereof (e.g., Morris & Larrick, 1995; Forsyth, 2004), nearly all evidence from this paradigm asks participants to make their best guess of what the author’s attitude is, using some form of bipolar attitude measure. Likewise, the silent interview paradigm asks participants to rate how anxiety-provoking the situation is, or how anxious of a person the interviewee is, which are not discrete judgments. Judgments of personality, ability, or other person causes are likewise more easily understood in terms of continua than probabilistic events.

It is possible to coerce a continuous variable into a discrete event, such as by using a threshold event (i.e., defining the event to be all values of the range above a critical value). However, this still leaves another problem, which is that reduced confidence in the presence of a trait isn’t the same thing as uncertainty, as it may reflect confidence in the absence of a trait (Shaklee & Fischhoff, 1982). Using discrete probabilistic judgments can therefore confound the estimate of a cause or trait with the certainty in that estimate. Given that some experimental manipulations affect certainty in an estimate without affecting the estimate itself (e.g., Croxton & Morrow, 1984), this is an issue.

Another problem with coercing continuous ranges into discrete events is that while causes may be somewhat naturally thought of with thresholds, the observed effects are
concrete enough that thinking in terms of the effect observed along with all stronger effects would be unnatural. Kun et al. (1980) find that when one facilitative cause is known to be present, people discount below the prior probability of the other cause, even though both are facilitative. Their explanation is that perceivers expected a stronger or different effect than would be obtained with one cause. While this isn’t a problem mathematically, it is a counterintuitive result and isn’t naturally handled by a discrete probability treatment.

A problem for discounting and augmenting in general is that detecting either often occurs via changes in attitude, trait, or other attributions. This entails the assumption that these judgments were mediated by causal attributions (Kelley, 1972a), even though it is not always true that the mediating role of these attributions is established (Kelley & Michela, 1980). However, Bayesian formulae are agnostic regarding causality, and so it isn’t necessarily important to know whether people make causal judgments, so long as the probabilities adequately capture people’s beliefs. This leads to the final, and perhaps most serious problem, which is that both asking participants for or arguing about the relative sizes of different probabilities begs the broader question of how people make such determinations to begin with (Fischhoff & Lichtenstein, 1978). Though Bayes’ rule ensures that the different quantities exist in proper relation to one another, at bottom the overall argument still rests on possibly erroneous or imprecise intuitions, despite the appearances that invoking mathematics may create. Similarly, asking participants to directly estimate the probabilities to be plugged into a Bayesian identity may merely restate the fundamental judgment involved in making the attribution without actually revealing much about the core considerations that people’s judgments encompass. Here again, Bayes’ rule offers a way to compare the two sets of judgments, but it does not necessarily direct attention to the most important aspects of the attribution.

Two adjustments can address these problems. First, more detailed models are needed for the constituent judgments in the identity. Second, these models need to be adapted to continuous variables. Both are addressed in turn.

### 2.6.1 Detailed Causal Schemata

One solution to the above problems is to encode more complex theories of how multiple causes interact to create particular effects. In fact, this was the aim of early theories like Kelley’s Causal Schemata Model and Reeder and Brewer’s Implicational Schemata Model. However, as mentioned previously, these models incorrectly equated the likelihood of the effect and the posterior probability of the cause. With this paper’s standards, these complex theories of cause and effect can be encoded in a way that directly translates into probabilistic terms, thereby allowing the proper inversion of probabilities to take place.

The major contribution that Morris and Larrick (1995) made was to take Kelley’s multiple sufficient and multiple necessary causes schemata and relax sufficiency and independence assumptions while also properly inverting the effect likelihood into a posterior judgment of one of the causes. However, the multiple sufficient causes schemata is only one way that causes may interact. For instance, Kelley (1972b) also proposes a compensatory
causes schema, where the effect occurs when the summed strength of the two causes exceeds a threshold, and an additive effects schema, where the strength of the effect is the sum of the strengths of the two causes. He also suggests the possibility of irregular relationships between causal strengths and effects. Kun et al. (1980) explain their results in terms of a varient-effects schema, wherein the effect of two causes may be qualitatively, not just quantitatively different from the effect of either cause alone.

Reeder and Brewer likewise describe more complex relationships between causes and effects. Their model explicitly says that both causes and effects exist on ordered continua, and then discuss the conditions where a correspondent inference (“a decision that the actor’s behavior and the actor’s disposition are to be classified at the same point along their respective attribute continua”, p. 63) is warranted. They describe three implicational schemata, which specify the range of behaviors that a person with a particular level on the dispositional cause is capable of (with the particular points chosen possibly dependent on another cause, though they do not model this part very explicitly). In their partially restrictive schema, a person may perform behaviors at or near their level on the trait, but not throughout the entire range. In their hierarchically restrictive schema, a person may perform behaviors at or below their point on the range (as in intelligence or other skill-based domains, where ability places a ceiling, but not a floor, on performance). Finally, in their fully restrictive schema, disposition and behavior are very closely matched. If there is a range of behavior, then the disposition may be completely absent. (They suggest that this would apply to traits like “neatness” or “sloppiness”, which both imply one pattern of behavior, and neither of which would apply to people who are sometimes neat and sometimes sloppy.)

2.6.2 Application to Attitude Attribution

To illustrate how to adapt this paper’s standard to continuous causes and effects, an example is given using the attitude attribution paradigm. Apropos this paradigm, Reeder et al. (1989) proposed that people hold what they call a “central tendency assumption”, wherein observers expect that people will only express attitudes within a certain range of their own, personally-held attitude. Thus, when assigned to express a particular position, they may shift their essay in the direction of that position, but their own attitude will still anchor the extent of that shift. While people whose attitudes are near to the requested position may end up expressing the requested position, people whose attitudes are far from the requested position will have to meet the request halfway in between (as it were).

Reeder et al. express their model graphically (p. 170) and verbally, but do not formalize these notions. Drawing on their intuitions, Chapter 3 models a version of the central tendency assumption, using this paper’s normative identity in the process. The following is a brief summary of how Jennings does this, which is meant to model how the approach can be generalized to other schematic assumptions.

In the attitude attribution paradigm, the certain cause, \( C \), is the request made of the author, and the uncertain cause, \( U \), is the author’s attitude. The effect, \( E \), is the essay provided. For attitude attribution, if it is clear that the requested position is randomly assigned,
then $C$ and $U$ are unrelated, making the co-occurrence term one and hence not important for the model. To model the remaining terms, Jennings introduces the continuous variables $a$, $e$ and $p$ to express the author’s actual attitude, the attitude expressed in the essay, and the position requested, all on a pro-con continuum. He also uses a variable $s$ to represent the strength of the request. In a variation on the central tendency assumption, Jennings assumes that authors will attempt to express a position that lies in between their own attitude and the requested position, with the degree of compromise determined by the situation strength. If $s$ is limited to be between zero and one, then the expected essay position becomes $a(1 - s) + ps$. He then lets $p(e \mid a, p, s)$, or the expected distribution of essay positions for a person with attitude $a$ who was requested to express position $p$ with strength $s$, be normally distributed with mean $a(1 - s) + ps$ and a fixed variance. This functional relationship models the implicational schemata expressed in the central tendency assumption (though with the added flexibility of the parameter $s$).

In order to properly invert the likelihood of the essay into the posterior probability of the author’s attitude, Jennings notes that $p(a \mid e, p, s) \propto p(a) \cdot p(e \mid a, p, s)$. Varying $a$ throughout its range, and making assumptions about the prior distribution of attitudes in the population, Jennings is able to construct the posterior distribution of the author’s attitude. Using the mean of the posterior as the attitude to attribute and letting the variance of the posterior be inversely proportional to the confidence in this attribution, Jennings is able to postdict a variety of findings from the attitude attribution literature (for details, see Chapter 3).

The above approach allows the exploration of a variety of scenarios, and shows that there are multiple coherent ways that the same results can be obtained. In addition to varying the parameters, $e$, $p$, and $s$, the actual schema itself could be varied. For instance, Gawronski (2003) argues that people use an ability schema (i.e., Reeder and Brewer’s hierarchically restrictive schema) to make attitude attributions. By combining modeling to determine when the two schemata make different predictions with empirical work to test those predictions, future research can better understand how people reason about attitude attributions, and can be confident that the model’s assumptions about how authors behave ($p(e \mid a, p, s)$) are coherently reflected in their judgements.

**Extensions**

When people make causal attributions, they are more likely to reason in terms of specific causal mechanisms than covariation between cause and effect (Ahn, Kalish, Medin, & Gelman, 1995). The approach just demonstrated essentially models the expected distribution of outcomes for a mechanism (attempting to write an essay whose position compromises between what was requested and what the author believes), and traces the implications of this mechanism through to making an attribution. The approach can be extended to other schemata (based on other mechanisms), or to model causes that co-occur. In so doing, it must be understood that what is produced is a normative standard only insofar as the model matches people’s assumptions about the causal mechanism. For instance, if people’s at-
titude attributions deviate from what the above model predicts (or what an ability schema would predict), this could mean that they are reasoning in a biased way, or it could mean that they are reasoning with different schematic assumptions about cause-effect relationships. The utility of the modeling approach is that it allows researchers to communicate and explore their expectations explicitly (Gigerenzer, 2009), and that it properly inverts models that go from causes to effects into inferences from effects to causes.

### 2.6.3 Interactive Causation

As touched on previously, theory and research on discounting has often assumed that the two causes in question exist in hydraulic relation to one another. This assumption may trace back to Heider (1958), whose work helped establish social psychology’s interest in many things, including the role of person and situation causes, biases in invoking either kind of cause, and the assumption that person and situation causes are polar opposites that trade off against one another when producing behavior (Hilton, 2007). Of course, this hydraulic assumption is easily recognized as a form of discounting itself.

The assumption that person and situation causes are inversely related was so pervasive that early attribution studies often used bipolar response scales anchored with endpoints like person vs. situation, or internal vs. external cause. However, when separated, the attributions are often found to be relatively independent, meaning both that they should not be operationalized as one continuum, and that studies using one continuum provide weak evidence for discounting at best (McClure, 1998). Even absent problems of operationalization, discounting research has tended to use causes that are competing explanations for the same effect, rather than being compatible with each other (Morris & Larrick, 1995; McClure, 1998). This may have led to a view of discounting as an attempt to subtract or adjust for the effect of one cause in order to have a true understanding of the other cause (e.g., Hansen & Hall, 1985; Fiedler, 2007), which also tracks with the view of attribution as an initial dispositional attribution that must be adjusted for the situation (Gilbert et al., 1988). As Gawronski (2004) notes, it may be more appropriate to think of the second phase of attribution as complete recomputation rather than simple adjustment.

Krueger (2009) suggests that social psychologists have tended to neglect interactions between person and situation causes, both via the theories and the experimental paradigms in use. The standard in this paper, particularly when applied its continuous form, nicely highlights the potentially interactive nature of person and situation causes (or any two causes), and should be a useful tool for achieving a more nuanced understanding of people-in-situations.

### 2.7 Implications for Attribution Biases

In addition to an overriding concern with person and situation causes, another of Heider’s legacies on attribution is a concern with attribution errors and biases (Hilton, 2007).
The standard in this paper is useful for assessing the coherence of people’s attributions. As such, it can be used to differentiate actual biases in the attribution process from findings that contradict researchers’ intuitive yet incomplete expectations, and should be a useful tool in conducting attribution research.

Many relatively contemporary reviews have already weighed in on whether the fundamental attribution error is indeed an error or a problem, and whether it implies problems in reasoning or problems in beliefs (e.g., Sabini, Siepmann, & Stein, 2001; Gawronski, 2004; Krueger & Funder, 2004; Hilton, 2007). While none of these reviews had the advantage of this paper’s normative approach, neither have the papers that these reviews draw upon. Though it is certainly appropriate for reviews to suggest alternative explanations for previous results, empirical demonstrations are invariably more compelling. Thus, rather than attempting to re-review the attribution literature with an eye toward detecting previously unrecognized coherent ascriptions, this section will offer suggestions on how future empirical work can take advantage of this paper’s standard to test these possibilities directly.

### 2.7.1 Heeding Scope

The most important guideline for applying this standard is to recognize its intended range. Specifically, this standard applies when an effect has occurred, and there is one cause whose presence (and if applicable, strength and direction) is known, and one cause whose presence (or strength and direction) is unknown. Appropriate dependent variables are the probability of the cause’s presence, or an estimate of the cause’s strength and direction. Care should be taken that the dependent variables tap the causal model of the event, and not explanatory factors. Thus, questions about the cause’s explanatory value, impact, or importance are not covered by the model. Additionally, though the term “cause” is used casually throughout this paper, the identity at work applies to associations between quantities that are necessary but not sufficient to establish causality. This implies both that researchers need to remember this distinction, and that proper dependent variables should ask about the “cause’s” presence or level, and not, say, the probability that the “cause” was a cause of the effect, or that an effect happened because of the “cause” (see also Einhorn & Hogarth, 1986; Cheng & Novick, 1991). These distinctions may be too subtle for participants to notice (see, e.g., Chapter 4), but then again, they may not.

As already mentioned, this standard applies after behavior has been classified. However, knowledge of the situation where the behavior occurred can have a major impact on how it is classified, which may in turn skew people’s attributions. So long as the behavior is classified in the same way for each of the judgments being elicited, the standard should still be able to detect coherence. However, if it is possible that people’s opinion of the effect will change from judgment-to-judgment, then the standard may fail to detect coherence simply because the judgments relate to different appraisals of the behavior. These problems are less likely to occur when the behavior and certain cause are introduced simultaneously, but may present a major problem when the certain cause is introduced after the behavior, and after an initial attribution ($P(U \mid E)$) has been made.
It is common for attribution researchers to manipulate both informational elements of the judgment task and aspects of the participant’s situation, such as by making the participant cognitively busy. When judging coherence in such cases, it is important to remember that the background beliefs need to be elicited in the same state as the target ascription. For instance, people may make coherent attributions that underestimate the power of the situation if cognitive load affects their estimates of probabilities like $P(E \mid C)$, such as by altering people’s accessibility experiences (Higgins, 1996) when they try to imagine times when similar circumstances produced similar effects. Thus, if participants are cognitively busy when they make the target judgment, they should also be cognitively busy when they make the constituent judgments. Of course, the constituent judgments could also be elicited in non-busy conditions as a way to triangulate upon what changes when people are cognitively busy, but this would be asking a different question.

### 2.7.2 Meaningful Manipulations

In addition to applying the standard directly in order to assess coherence, the standard can be used to assist in the research design process. For instance, Kelley (1972a) points out that experiments claiming to find situational neglect may simply have failed to vary situational causes as strongly or clearly as behavioral causes, making it difficult to consider the apparently reduced response to the situation an error or bias. Shoda and Mischel (1993) claim that research participants want information that would help them look at person by situation interactions in trait attributions, but that study designs just don’t provide it. McClure (1998) faults discounting research for only testing alternative explanations for the same effect, rather that compatible explanations. As all of these examples show, researchers may fail to vary information in their scenarios in ways that test the range, relationship, and relative strength of various causes, and therefore may obtain only a partially representative understanding of people’s reasoning. This paper’s standard can be used both as a heuristic when selecting scenarios (as in the empirical demonstration, where an effort was made to systematically vary co-occurrence and relative likelihood), and as a manipulation check to ensure that the informational variations have the intended normative impact, both in direction and strength.

A related point to make is that it may in fact be difficult to vary the entire configuration of co-occurrence and relative likelihood. As a case in point, the manipulations in the empirical demonstration were not fantastically successful. It may simply be that trait terms have evolved in use in such a way that they naturally co-occur with other traits and causes that have similar effects. Indeed, this may be a reason that the distinction between co-occurrence and relative effect likelihood has not received greater emphasis before this paper. Of course, the deconfounding of variables that are naturally confounded is of great benefit to truly understanding how reasoning occurs.
2.7.3 Demonstrating Biases

Decades ago, Kelley (1973) wrote, “I believe social psychologists finally are realizing that their proper role is not to confound common sense but rather to analyze, refine, and enlarge on it” (p. 108). Only a few years ago, Krueger and Funder (2004) noted that judgment research has an unhealthy fixation on biases in reasoning. It may be that Kelley’s optimism was misplaced.

Krueger and Funder suggest that researchers should treat instances of apparent judgmental bias as a good Bayesian would, that is, not being so quick to reject the prior hypothesis that people can reason pretty adeptly after seeing one exception. However, if researchers do indeed have a fixation on finding flaws in judgment, then the result may be that researchers overestimate how common flawed judgments are in ordinary life. It may now be that a researcher who does not question the apparent flaws in yet another judgment task is in fact being a coherent Bayesian, though one who uses biased background assumptions. As a result, people’s judgments are considered “biased until proven normative” rather than “normative until proven biased”. Demonstrating significant differences between experimental conditions or cultural groups, or finding departures from normative expectations (informal or formal), are treated as prima facie evidence of a bias, and are added to the mounting pile of biases in reasoning. Only much later do voices come suggesting that the researchers’ thinking, and not the participants’ behaviors, may have been wrong.

The above is a cartoon characterization, and more than a little self-serving. However, the point does need to be made that researchers ought to be on the side of their participants. If apparently biased reasoning is a crime, then researchers ought to play defendant, and not prosecutor. One way to do this is to proactively measure the beliefs that make it possible to assess coherence. However, it will not always be practical to proactively measure these assumptions, particularly with process manipulations designed to short-circuit deliberate thought. If apparently biased responses are observed in such cases, researchers should use this or other normative standards to reverse engineer sets of assumptions that would make their participants’ responses coherent. Only if it can be convincingly demonstrated that people do not hold these assumptions should it be possible to convict people of biased (incoherent) reasoning.

2.8 Conclusions

This research has illustrated the use of a model that describes how to determine the internal consistency of beliefs about the causes of behavior. The model intuitively relates the normative confidence that a trait was present when an event occurred to the power of the trait to make the event happen. When it is learned that a circumstance capable of causing the event was also present, the normative attribution is revised according to whether the circumstance and the trait co-occur, and to whether the trait enhances (or hinders) the effect of the circumstance in either an additive or multiplicative way. Therefore, in contrast
to previous models (e.g., Morris & Larrick, 1995) and the general trend in social and personality psychology (Krueger, 2009), this model considers the possibility that personal and situational causes might have interactive effects.

In addition to introducing the above normative standard, this paper has added needed clarification to the literature on discounting and augmenting. First, several examples of older results in the discounting literature that had previously been treated as contradictory were shown to stem from differing operationalizations of the concept. Second, before introducing this paper’s standard, different ways of determining correctness were discussed relative to attribution research. The standard itself was shown to improve upon previous work, and in particular it was used to discover some oversights in Morris and Larrick’s (1995) otherwise excellent adaptation of Kelley’s theories. Additionally, the standard was shown to be flexible to apply in empirical work, although many questions remain about how best to elicit people’s beliefs in a way that maximizes their coherence. Finally, it was shown how the standard can be expanded to work with continuous models of behavior, avoiding the confusion of the inverse inherent in previous theories of how assumptions about cause and effect relationships should affect attribution (Kelley, 1972b; Reeder & Brewer, 1979).

Despite its strengths, this work is in its inchoate stages. The remainder of this paper discusses its limitations, and considers several interesting prospects for future research.

2.8.1 Limitations

This paper’s model is a powerful tool for advancing attribution research. Its main limitations are essentially “design choices”, limitations built in so that its scope would be manageable. Primary among these are the decisions to leave aside issues of behavior classification and causal search—in large part because it is not clear how either of these phenomena would be integrated into a normative rather than descriptive framework. As such, this model is also limited in its ability to directly encompass both normative and counter-normative aspects of reasoning, which is essential in order to have a comprehensive understanding of cognition (Krueger & Funder, 2004). However, in all of these cases, the model should facilitate the development of better descriptive models of cognition.

There are also limitations of the model that are not simply deliberate choices. Kelley (1972b) observed that causal schemata may be incomplete since people are likely to pay attention to more than just the presence or level of causes. Similarly, despite this model’s ability to be adapted to continuous causes and effects, people no doubt make much more nuanced appraisals of people and events, and this approach will quickly exhaust its ability to handle these factors. In fact, the simpler, discrete cause approach may be more facile in this case, though in the process it obscures exactly how such factors are encompassed within judgments of $P(E | C, U)$, etc.

It is also not immediately clear whether and how this model will be applicable to story understanding models of causal reasoning (Read, 1987; Read & Marcus-Newhall, 1993). It is possible that this model could be useful as a step within the process of determining the likely chain of events leading to an event. However, once again, the model’s limitation to
two causes may prove insufficiently facile for such use.

### 2.8.2 Future Directions

#### Eliciting Beliefs

The upper limit on how much coherence this model can demonstrate is set by the quality of the knowledge elicitation techniques in use. This paper’s empirical demonstration found generally good coherence (judged by low or nonexistent bias), but questions remain about how to interpret the log-relative format used, as well as how to help participants avoid common errors in probabilistic reasoning. Moving forward, it will be useful to explore several aspects of how to elicit the beliefs needed for this model in order to improve the ability to find coherence.

There are several variations on elicitation that should be explored. First, a systematic comparison of between- and within-subjects variation of cause candidates should be performed (cf. Kahneman, 2003). Second, attempts should be made to encourage an “outside view”, or distributional thinking, which entails encouraging people to move beyond one particular case to the class of all similar cases, something that has been shown to improve judgment quality (Lagnado & Sloman, 2004). Third, techniques should be attempted that are a more direct match to the probabilities involved, as opposed to the log-relative scale used in this paper. One possibility this would open up would be to elicit more concrete beliefs. For instance, rather than comparing $P(E | C, U)$, which is concrete, to $P(E | C)$, which involves uncertainty about $U$, $P(E | C)$ could be recomposed from $P(E | C, U)$ and $P(E | C, \overline{U})$ (and $P(U | C)$). Given the general difficulty that people have reasoning about uncertainty or abstract cases, this may also improve people’s coherence (Kleinmuntz, Fennema, & Peecher, 1996).

Another benefit of perfecting the techniques for eliciting the quantities involved in this model will be the establishment of guidelines about how much coherence can be expected. For instance, in personality psychology it was once generally thought that the upper limit on a correlation between a personality variable and a particular behavior tends to be around .3 (Mischel, 1968). While this was originally touted as a small correlation, it also set an expectation about the limits of what nature provides. Knowing how much coherence is possible when using this paper’s model would be similarly useful. Given that “rationality at best remains a null hypothesis that has failed to be rejected” (Krueger & Funder, 2004, p. 318), knowing the effect size for incoherent judgments and using it to determine the minimum sample size would bolster the case that failing to reject the null hypothesis is not just a Type II error.

#### Developing continuous models

While the straightforward application of the discrete model is useful and represents a major advance, continuous extensions of the sort demonstrated with the attitude attri-
bution paradigm have an even greater potential to advance social judgment research. As Gigerenzer (2009) observes, theories in psychology tend to be imprecise and capable of explaining seemingly any finding after the fact, but cannot generate predictions of their own. He recommends the development of objective computational and mathematical models that have the virtue both of generating new ideas, and of being provably wrong. Though the discrete form of this paper’s model highlights the information categories that attributors ought to pay attention to, it does not model and cannot rigorously predict how people will make these judgments. By theorizing and then modeling different assumptions about how continuous causes relate to continuous effects, and then using the computational model to generate unique and testable predictions, this approach can help to advance theory development. While the modeling techniques used here are still unfamiliar to most social psychologists, they are still more accessible than many other kinds of computational models.

**Reexamining prior work**

As noted earlier, previous research has not had the benefit of this paper’s standard. It may indeed be that many attribution problems that are currently considered to be biased will eventually be at least partially vindicated as coherent when tested in the proper way. While researchers are naturally more inclined to invent new theories and use compelling “big science” methods to test them, doing so without a proper understanding of findings so old that they are taught as fact in social psychology textbooks may only be forestalling the eventual heartbreak if the logical basis of these findings proves to be flawed.

New research methods always opportune a rush to pick the low hanging fruit of questions that were not previously answerable. There is a trove of past research in causal ascription with two causes that can be examined anew with this paper’s techniques. Far from rehashing old territory, investigating these questions promises to break new ground on how attribution is understood. Starting with familiar phenomena is a good way to begin the process.

**Probing the process**

McClure (1998) notes that Gilbert and Malone’s (1995) review of the causes of the correspondence bias fails to take into account “logical factors” (i.e., coherence) that may underlie people’s judgments. While many of the factors that Gilbert and Malone identify will no doubt retain their status as meaningful barriers to sound reasoning, at present this oversight renders some of their results ambiguous. Still, differences in people’s judgments as a result of cognitive load or other contextual factors suggests that either people are not always coherent in their judgments, or that they are coherent with respect to a shifting set of assumptions. Once the limits of coherence are known when people are given ideal conditions under which to reason, it makes sense to begin examining how that coherence breaks down. With a normative standard to separate the figure (faulty judgments) from ground (coherent if counterintuitive judgments), this line of research can much more rigorously
determine how the judgment process itself works.

**Multiple observations**

Many authors have suggested that the differentiating factor between induction versus ascription is the number of observations available (e.g., Morris & Larrick, 1995; Van Overwalle & Timmermans, 2005). However, while induction always requires multiple observations, multiple observations needn’t always imply that induction is at play.

Consider a simple case where a person’s attitude is being judged on the basis of an essay written under constraint. Suppose that after this judgment, another essay became available, also written under constraint. Would the latter inference, which by itself would clearly be ascription, suddenly become induction by its juxtaposition with the former inference? The answer is no, since the inference is still aimed at determining the level of one cause (the attitude) in producing an effect (well, two effects—the two essays).

The above scenario bears more than passing resemblance to scenarios that have been billed as inductive in nature. For instance, Van Overwalle and Timmermans present participants with the outcomes of tennis matches among various players, and asks participants to infer the extent that each player had an influence on the outcome—presumably because of their skill or performance. Though they view this as an inductive inference since skill can be thought of as a probability of winning \( P(\text{win} \mid \text{skill}) \), it is unlikely that participants approached these scenarios in the same way as they would have approached completely unfamiliar cause-effect contingencies. Instead, they most likely used schematic assumptions about how skill, motivation, and luck conspire to determine the outcome of tennis matches. Thus, participants may really have been ascribing a continuous skill parameter conditioned on the outcome and other information, and via some causal schema for \( P(\text{win} \mid \cdots) \).

The confusion in the above example may stem from a lack of differentiation between general cause categories (skill) and specific manifestations of those categories (e.g., Andre Agassi’s skill). Inductive inferences are required to learn how general causes produce general effects. This knowledge will in turn be applied to understand one particular win or loss. That ascription may then update the general knowledge about how wins and losses are obtained. However, the proximal inference is still ascription, and should still be judged according to normative standards for ascription. Before this can be done, however, normative standards for ascription across multiple observations will need to be devised.

**Both causes uncertain**

When both causes are uncertain, this paper’s standard is not directly applicable. In general, the probability of an uncertain cause \( U_1 \) is expressed by Bayes’ theorem, i.e.,

\[
\frac{P(U_1 \mid E)}{P(U_1)} = \frac{P(E \mid U_1)}{P(E)}
\]

When there is another cause \( U_2 \) that may also be present, implicitly nothing changes about the above expression. However, it is illustrative to consider the normative value of \( P(E \mid \)
in relation to the more concrete cases where \( U_2 \) is or is not present, which can be done by using the following expression: \(^{10}\)

\[
P(E \mid U_1) = P(E \mid U_1, U_2)P(U_2 \mid U_1) + P(E \mid U_1, \overline{U}_2)P(\overline{U}_2 \mid U_1)
\]

A similar expression applies for \( P(E \mid U_2) \). As applied to discounting (i.e., the possibility of multiple causes reducing the confidence in any one), contrasting (say) \( P(U_1 \mid E) \) before and after the possibility of \( U_2 \) was mentioned would only involve a change if \( P(U_2) \) changed as a result of its mention.

**Conjunctive explanations**

A topic of research related to discounting is conjunctive explanations, where the presence of two causes is seen as more probable than the presence of either one (Leddo et al., 1984). McClure (1998) frames conjunctive explanations as a contradiction to the discounting principle, in that the discounting principle suggests a single cause explanation is preferable due to its simplicity, while conjunctive explanations involve multiple causes, and therefore a more complex causal scenario. If the criterion of interest is whether (say) \( P(U_1 \mid E) < P(U_1, U_2 \mid E) \), then conjunctions are always irrational, just as with the conjunction fallacy, the belief that \( P(A, B) > P(A) \) (Tversky & Kahneman, 1983). However, if the criterion is whether \( P(U_1, \overline{U}_2 \mid E) < P(U_1, U_2 \mid E) \), then there could clearly be a rational basis for preferring conjunctive explanations, and one that is compatible with discounting. It may prove useful to use models of \( P(E \mid U_1, U_2) \), to which the above probabilities are related via Bayes’ theorem, to explore these questions.

**Culture**

In psychology, social cognitive processes are the most studied kind of cultural difference, which is fitting given how influential the social cognitive approach has been in psychology. Much cross-cultural work in social cognition has used intuitive standards to determine what is normative. When the researcher and the participant inhabit the same culture, the source of these intuitions is consistent. However, when the researcher and participant live in different cultural realities, there might be a mismatch of intuitions. Without a rigorous way of determining what is normative, fully explainable and internally consistent mismatches can be mistaken for fundamental differences in cognition. This standard can help highlight the relevant background beliefs that might differ across cultures, and can ensure that research stimuli are equivalent across cultures, not just in their surface meaning, but in the normative conclusions they imply (see Chapter 4). Normative models like this one should become an integral part of cross-cultural work in social cognition, just as cross-cultural comparisons should be come an integral component of social cognition research in general.

\(^{10}\)\(P(E \mid U_1)P(U_1) = P(E, U_1) = P(E \mid U_1, U_2 + \overline{U}_2) = P(E \mid U_1, U_2)P(U_1, U_2) + P(E \mid U_1, \overline{U}_2) + P(U_1, \overline{U}_2) + P(E, U_1, U_2, \overline{U}_2)\)
2.8.3 Prospects for Attribution Research

Attribution studies are deceptively simple to design: write a scenario, vary a fact or two, and ask a question about it. Beneath this simplicity is a realm of nuance that is difficult to appreciate at first. This makes sense: human behavior is delightfully nuanced, and so it would be surprising if the thought processes that we have to understand it were any less complicated. Simple assumptions about reasoning won’t suit this reality well. For instance, contrary to received wisdom, constrained behavior can still be a meaningful indicator of personality (Gawronski, 2004; Hilton, 2007). At odds with the idealized view of experimental control, participants go beyond the knowledge provided in scenarios and make use of their general background knowledge when answering attribution questions (Hilton, 2007). Studying attribution may not require researchers to go beyond simple written scenarios, and it may not require expensive equipment that peers into the hardware of the brain. It will, however, require a complex understanding of reasoning, and it will require “technology” in the form of explicit theories of what attributions are normative and how attributions occur. This work shows how background knowledge can be properly taken into account, at least with ascriptive inferences to one cause when the other cause’s status is known. Its contribution may therefore be relatively small, but it is a useful move in the right direction.
Chapter 3

Application to Attitude Attributions

People’s tendency to neglect situational influences on behavior has been a subject of long-standing interest to social psychologists. Many of the earliest and most famous demonstrations of this error make use of the attitude attribution paradigm (Jones & Harris, 1967), wherein participants read an essay that expresses an opinion on an issue, and must estimate the author’s attitude. Complicating this judgment is the fact that the author was assigned what position to express, which pits two competing explanations—holding the attitude, or complying with the request—against each other. Participants tend to make attitude attributions in line with the essay even when the position was assigned, which is called the correspondence bias (Gilbert & Jones, 1986; Gilbert & Malone, 1995).

It is not straightforward to say whether people’s responses in the attitude attribution paradigm are in fact biased. On the one hand, evidence suggests that people’s attributions are incorrect: when participants rate essays that other study participants wrote under constraint, the attributed attitudes are more in line with the essay than with the authors’ self-reported attitudes (Snyder & Jones, 1974; Reeder et al., 1989; Fletcher, Reeder, & Bull, 1990; A. G. Miller, Ashton, & Mishal, 1990; see also A. G. Miller et al., 1981). On the other hand, if people’s attributions are internally consistent with their own perceptions and assumptions, it is hard to call their attributions completely biased (Jones, Worchel, Goethals, & Grumet, 1971; Ajzen & Fishbein, 1975; Kelley & Michela, 1980; A. G. Miller & Rorer, 1982; Morris & Larrick, 1995; Gawronski, 2003; Forsyth, 2004). These two views concern two different standards for correctness, known as correspondence and coherence, respectively (Hammond, 1996; Dunwoody, 2009). The former view, correspondence, concerns the external validity of the judgment, while the latter view, coherence, concerns the internal consistency of the judgment. To avoid confusion between correspondence criteria and the correspondence bias, the terms external validity and internal consistency will be used. Though people’s judgments probably lack external validity, it is not clear whether they are at least internally consistent.

Checking internal consistency requires knowing what information is relevant to a judgment, and how that information determines the correct answer. This paper develops a Bayesian model relating assumptions and perceptions to attitude attributions. Since the
model in grounded in mathematics, the steps between premises and conclusions can be more readily verified than with previous verbally justified standards (Morris & Larrick, 1995). Additionally, the model is agnostic to what process people might use to make judgments, helping researchers advocating different mechanisms at least agree on the expected outcome (Sun, 2008).

### 3.1 Normative Model

In the attitude attribution paradigm, observers know what essay was written, and the circumstances under which it was written. Their judgment of whether the essay author holds the expressed attitude is (Morris & Larrick, 1995):

$$P(\text{attitude} | \text{essay}, \text{circumstances})$$

Letting $A$, $E$, and $C$ stand for the attitude, essay, and circumstances, and applying Bayes’ rule in a strategic way,\(^1\) this equals:

$$P(A) \cdot \frac{P(C | A)}{P(C)} \cdot \frac{P(E | A, C)}{P(E | C)}$$

Intuitively, these three terms express the prior probability of the attitude, the co-occurrence of the circumstances of the attitude, and the relative likelihood of a person writing the essay, comparing someone with the attitude to the average person.

\(^1\) $P(A | E, C) = P(A, E, C) / P(E, C) = P(E | A, C)P(A, C) / P(E, C) = P(E | A, C)P(C | A)P(A) / [P(E | C)P(C)]$. See also Chapter 2.
The model can be applied in two ways. First, it can be interpreted schematically in order to draw conclusions about the general pattern of inferences that should be expected. For instance, the standard makes it immediately clear that the conventional wisdom that the essay communicates no information about the author’s attitude in light of the circumstances is correct only if two conditions are met. First, the co-occurrence term must be one, meaning that the position the author was assigned must be unrelated to the author’s attitude. Simply specifying that the position was assigned by the professor is insufficient to make the author’s attitude meaningless, since the observer might reasonably assume that the professor considered the author’s attitude when making the assignment. Second, the likelihood term must also be one, meaning that the constraint must be seen as equally compelling regardless of the author’s attitude (with completely compelling being a special case of this). As other researchers have argued (e.g., Jones et al., 1971) or found (e.g., A. G. Miller, 1974), observers do not seem to hold this belief.

Though these conclusions are powerful, it is possible to do better. The second way to apply a model such as this one is to use it to quantitatively assess the internal consistency of people’s judgments, which is done by measuring quantities on both sides of the equation. Previous authors have done this using alternative Bayesian standards (Trope, 1974; Morris & Larrick, 1995; Forsyth, 2004). However, every previous model has required participants to think in terms of discrete or comparative probabilities (e.g., the likelihood of the essay with and without the attitude, as in Trope [1974] and Forsyth [2004]), while nearly all other studies of attitude attribution ask participants to estimate the author’s true attitude on a Likert-type scale. Achieving a match to what participants actually estimate requires switching from the probabilities of dichotomous events (e.g., writing the essay or not) to probability densities over continuous variables, as follows:

- “Pro” and “con” attitudes are generalized to real-valued attitudes along a “con” (negative) to “pro” (positive) continuum, with attitudes further from zero being more extreme. The variable $a$ will refer to the author’s attitude, while $e$ will refer to the position expressed in the essay.

- The circumstances ($C$) are decomposed into two things: $p$, the position that the author was asked to express, and $s$, the strength of that request. The variable $p$ can vary as discussed above, while $s$ can vary between zero (for no inducement at all) and one (for a completely compelling inducement).

Using the above variables, $P(A \mid E, C)$ becomes $p(a \mid e, p, s)$. Converting the prior, co-occurrence, and relative likelihood terms into probability distributions and multiplying the three over the range of $a$ gives the probability of each possible attitude. The expected value of this distribution will be the attitude attribution, and the confidence in this attribution will be proportional to the distribution’s standard deviation.

Completing the normative model requires specifying the forms of the three terms. The prior distribution, $p(a)$, is just the assumed distribution of attitudes in the population that the author comes from. The co-occurrence term expresses how the circumstances vary with
the author’s attitude. Assuming a random assignment process, then this term will be one throughout its range, and can be ignored. This leaves the likelihood term, \( p(e \mid a, p, s)/p(e \mid p, s) \). Since the denominator does not involve \( a \), the expression can be written:

\[
p(a \mid e, p, s) \propto p(a) \cdot p(e \mid a, p, s)
\]

These terms will be called the posterior, the prior, and the (essay) likelihood, respectively.

The final task is to specify a form for the likelihood, \( p(e \mid a, p, s) \). This can be done by determining the distribution of essay positions that an author with attitude \( a \) would write when asked to express position \( p \), facing an inducement of strength \( s \). Instead of requiring participants to estimate this themselves (as in Trope, 1974 and Forsyth, 2004), the form of the function will be specified mathematically. Past research has found that observers expect constrained authors to express an attitude somewhere in between their own attitude and the attitude that was requested (A. G. Miller & Rorer, 1982), which Reeder et al. (1989) refer to as the central tendency assumption. Thus, an author with (say) a strong con position who was asked to express a strong pro position would attempt to write a neutral essay. This expectation can be modeled by saying that when \( a \), \( p \), and \( s \) are known, \( p(e \mid a, p, s) \) is a normal distribution, with:

\[
\mu = a \cdot (1 - s) + p \cdot s
\]

With no inducement \((s = 0)\), \( \mu = a \), the author’s own attitude. With a completely compelling inducement \((s = 1)\), \( \mu = p \), the requested position. For other values of \( s \), \( \mu \) is a weighted compromise between \( a \) and \( p \). While one could imagine ways that the distribution’s standard deviation might depend on \( a \), \( p \), and \( s \), for parsimony it will be assumed to be constant.

The above model of authors’ responses specifies the distribution of \( e \), given that the other variables are known. However, when applied, \( e \) is known but \( a \) unknown. This does not present a problem, as illustrated in Figure 3.1. The left graph shows the essay distributions for two values of \( a \) (-2 and 2), where \( s = .5 \) and \( p = 2 \). The right graph shows the likelihood distribution over the range of \( a \), where \( e = 2 \). The middle image shows how the two graphs are related, with two points shown on all three graphs (\( e = 2 \) and \( a = \pm 2 \)) for reference.

As already mentioned, this model improves upon previous Bayesian models of attitude attribution in that its output is the same kind of variable as participants actually estimate. In addition, the model’s inputs correspond to perceptions that are relatively straightforward for participants to reason about (the position expressed in the essay, \( e \), whether the essay is weaker or stronger than was expected of the author, \( e - p \), and how constraining the situation was, \( s \)). When testing scenarios schematically, this makes it possible to continuously vary the model parameters, or to test specific combinations of parameters, rather than having to make verbal arguments about, say, the relative sizes of \( P(Essay \mid \text{Attitude}) \) and \( P(Essay \mid \overline{\text{Attitude}}) \) (cf. Ajzen & Fishbein, 1975). When testing the internal consistency of participants’ actual judgments, it becomes possible to directly ask for the relevant quantities. The tradeoff is that the model does assume that people believe constrained authors
Figure 3.2: Model of the choice (left) and no choice (right) conditions for Jones and Harris (1967), Study 1. In the bottom panel, black arrows show original results, and black lines show the model’s results.

will express a position between their own attitude and the request. However, future research could specify other models of how authors respond (e.g., Gawronski, 2004) and translate this into a likelihood function as was done here.

3.2 Illustrations

3.2.1 Choice and Prior Probabilities

Correspondent inference theory (Jones & Davis, 1965) was intended to be a normative standard for how people should make attributions, and aims to specify which behaviors justify the inference of information about a person that would not have been assumed previously (Jones & McGillis, 1976). Jones and Harris (1967) was an attempt to show that while both constrained behavior and expected behavior do not contribute new information, an expected behavior performed under constraint will still lead to a corresponding attribution, simply because the underlying disposition would be expected anyway. It is for this reason that they used advocacy for “Castro’s Cuba”—a highly unexpected behavior in 1960’s America—as the critical test. As predicted, they found that people made attributions corresponding to the constrained behavior when the behavior was expected (arguing against Castro). What they were surprised to find was that though people did not make completely corresponding attributions when the behavior was unexpected (arguing for Castro), their attributions did not revert to the level that would be obtained had the behavior been completely disregarded. This is the result that triggered the volumes of research on the correspondence bias that continues to this day.
Table 3.1: Model-predicted attributions for strong and weak essays under weak and strong situational constraint. Weak situation shows no reversal for the weak essays, but strong situation does. The same pattern can be obtained by keeping situation strength constant but making the weak essays less weak.

<table>
<thead>
<tr>
<th></th>
<th>Con</th>
<th></th>
<th>Pro</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Strong</td>
<td>Weak</td>
<td>Weak</td>
<td>Strong</td>
</tr>
<tr>
<td>Requested position ((p))</td>
<td>-2</td>
<td>-2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Essay position ((e))</td>
<td>-3</td>
<td>-1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Attribution ((a)), weak constraint ((s = .25))</td>
<td>-2.78</td>
<td>-0.56</td>
<td>0.56</td>
<td>2.78</td>
</tr>
<tr>
<td>Attribution ((a)), strong constraint ((s = .75))</td>
<td>-2.16</td>
<td>0.72</td>
<td>-0.72</td>
<td>2.16</td>
</tr>
<tr>
<td>Essay position ((e))</td>
<td>-3</td>
<td>-2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Attribution ((a)), strong constraint ((s = .75))</td>
<td>-2.16</td>
<td>-0.72</td>
<td>0.72</td>
<td>2.16</td>
</tr>
</tbody>
</table>

This paper’s model can reproduce the pattern of results that Jones and Harris obtained. Doing so requires two reasonable assumptions. First, assume that the prior attitude distribution was strongly right skewed (i.e., very few people supporting Castro), which is supported by anecdotal comments Jones and Harris make about participants’ self-reported attitudes. Second, for parsimony, assume that both the pro and con essays were equivalently strong, and no weaker or stronger than requested.

To reproduce the “choice” condition, the model is run with strength set to zero \((s = 0)\), which is illustrated in the left half of Figure 3.2. The top panel of this graph shows the prior distribution, \(p(a)\), while the middle panel shows the likelihood functions for the con (dashed line) and pro (dotted line) essays. These lines show \(p(e \mid a, p, s)\), where \(e = \pm 3, s = 0\) since authors could choose what to express (making the requested position, \(p\), irrelevant), and \(a\) varies across the \(x\)-axis to encompass the range of attitudes shown. The bottom panel shows the posterior distributions, \(p(a \mid e, p, s)\), which are the result of multiplying the prior distribution by either likelihood distribution. In this case, the prior distribution has only a small effect on the posterior distributions, and the expected values of the distributions (shown by the black, vertical lines) are a very close match to the results that Jones and Harris originally obtained (shown by the black arrows).

Jones and Harris found nothing counterintuitive about their results for the choice condition, but were surprised by the results in the no choice condition, which can be replicated by choosing an appropriate value for \(s\). Not shown in the figure is the case where the situation is seen as completely constraining \((s = 1)\). Under conditions with no behavioral freedom, everyone is equally likely to have written the requested essay, and so the two likelihood functions are flat lines. As such, both posterior distributions are equal to the prior distribution, making the normative attribution for both essays equal to the mean attitude in

\(^2\text{Note that attitude values are always rescaled to a } -4 \text{ [con] to } 4 \text{ [pro] for consistency of comparison across studies.}\)
the population. This result is what Jones and Harris were expecting to find. Since this is not what they obtained, values of \( s \) less than one must be tried.

A good fit to the original results was obtained with \( s = .6 \), meaning that the situation is seen to be constraining, but with room for some individual choice. The right half of Figure 3.2 shows this case, where it can be seen that though the prior distribution is the same and the likelihood functions have the same locations, the likelihood functions are also more spread out (since constrained behavior is less informative than freely chosen behavior). Even though the pro and con likelihood functions are symmetric, the posteriors are not, which is a result of multiplying by the asymmetric prior.

As the bottom panel shows, the expected values of either posterior (black, vertical lines) are quite close to the results that Jones and Harris obtained (black arrows). In particular, for the “con” essay, the model-derived and actual attributions are still in the direction of the essay. For the “pro” essay, however, multiplying by the prior probability has brought the model-derived results closer to the midpoint, and like the actual results, still somewhat correspondent with the essay itself. It is also worth noting that the “pro” posterior is more spread out than the “con” posterior, just as Jones and Harris found greater variance in this condition than in the other conditions of their study.

As the above shows, the model can reproduce the important features of the original demonstration of the correspondence bias, with only one parameter varying between the choice and no choice conditions. As such, it establishes that the results in Study 1 of Jones and Harris (1967) could be the result of an internally consistent reasoning process, given the assumption that the participants did not believe that the author’s situation in the no choice condition was completely constraining. In fact, according to the model, the only internally consistent ways for perceivers to make attributions other than to the mean attitude in the population are if they believe that the situation leaves room for choice, or if attitude and assigned position don’t interact. Though these attributions are probably externally invalid, the possibility that they are internally consistent suggests that defects in observers’ reasoning processes are not necessary to explain these results. Likewise, people may make perfectly reasonable assumptions about how situational constraints in general would influence essay authors. The source of “bias” may simply be that people applied those assumptions using an insufficiently strong appraisal of the power of the author’s particular situation.

### 3.2.2 Prior Probabilities Across Populations

In addition to illustrating the effect of the strength parameter \( (s) \), the previous example shows how important of a role the prior attitude distribution plays in determining what attribution people should make. Failing to recognize the role of the prior attitude distribution can lead to mistakes in interpreting results when the prior distribution might be expected to differ across experimental conditions, such as when culture is an independent variable. This can be illustrated by reexamining Study 2 of Miyamoto and Kitayama (2002), which found that American participants made more correspondent inferences than Japanese participants. Like other authors, Miyamoto and Kitayama measure bias via the difference in the mean
pro and con essay attributions, with larger distances indicating a larger bias. However, the means they report for the Japanese participants are asymmetric (i.e., the pro and con attributions are not equally extreme), whereas the Americans’ results are symmetric. As the model of Jones and Harris (1967, Study 1) shows, asymmetric attributions can result from an asymmetric attitude distribution. Likewise, symmetrically distributed attitudes should produce symmetric attributions.

With this in mind, the model was applied to reproduce their results, keeping every parameter except the prior attitude distribution consistent across cultures. The issue in this study was capital punishment. In order for differences in priors to cause their results, the American participants would have needed a symmetric prior (e.g., the normal distribution), while the Japanese sample would have needed an asymmetric prior, with more people favoring the death penalty. Though this might contradict stereotypical notions about either culture, it is worth noting that the study was run at a Japanese university, using American exchange students as the Western sample. This would give the Japanese students a clear population (students at that university), but leave the Americans unclear about the population, and hence more likely to assume a symmetric distribution. In addition, Japan does have the death penalty, and in data reported later, there was stronger support for the death penalty among East Asians. For these reasons, it is reasonable to speculate about whether prior attitude distributions could account for the results.

Model fitting is illustrated in Figure 3.3. The left graph represents the Japanese data, and the right graph represents the American data. As can be seen, a left-skewed distribution (i.e., mostly pro) is used for the Japanese data, and a symmetric distribution is used for the American data. Given that equals values were used for every other model parameter, the likelihood functions (middle panels) are the same in the Japanese and American models. However, when multiplied through by the different prior distributions, the Japanese and American posteriors produce very different results.

The model’s predictions match the results obtained by Miyamoto and Kitayama reasonably well. This suggests that the Japanese and Americans could have interpreted every aspect of the scenario and essays equivalently (suggesting no differences in the correspondence bias itself), but normatively arrived at results that make the Americans appear more biased simply due to different assumptions about the populations essay authors were drawn from.

Though this does not disprove Miyamoto and Kitayama’s interpretations and does not rule out cultural differences in the American and Japanese attribution processes, this demonstration does suggest an alternative explanation for Miyamoto and Kitayama’s results that bears further investigation. More generally, the model makes plain the fact that there are many possible ways in which people in different populations might arrive at different attributions while still reasoning in internally consistent ways. Absent a model like this, it can be difficult to know what assumptions and perceptions people’s judgments likely depend

---

3 Among 107 European-Americans, there was a mean attitude of -.75 (negative indicating opposition to the death penalty) with a median of -1, and among 103 Asian-Americans, there was a mean attitude of .38 with a median of 1. The difference in means is significant, $t(206) = 3.64, p < .001$. 
Figure 3.3: Model of Japanese (left) and American (right) results for Miyamoto and Kitayama (2002), Study 2. In the bottom panel, black arrows show original results, and black lines show the model’s results.

upon, and even more difficult to know how those assumptions and perceptions should be synthesized when determining whether a particular response is normative. With the model, it becomes possible to set up rigorous cross-cultural comparisons that distinguish between superficial and fundamental differences.

3.2.3 Degree of Compliance

Thus far, it has been assumed that perceivers believe that the essay written was no weaker or stronger than was requested. However, compliance needn’t be all-or-nothing. In a followup study to Jones and Harris (1967), Jones et al. (1971) manipulate the strength of the essay in order to understand how behavioral extremity affects attributions. One of their key results is that when people read an essay written under constraint and expressing a weak position, they attribute the opposite attitude to the author as was expressed in the essay. When the essay position was strongly argued, they attribute a corresponding attitude. In an attempt to replicate this result, A. G. Miller (1974) found that people made less extreme attributions when reading a weak essay than when reading a strong essay, but did not find any reversal. In both cases, however, the degree of compliance affected the attributions.

The model is able to reproduce these result patterns by varying the situation strength parameter, \( s \), and leaving everything else constant. Model-predicted attitude attributions for weak and strong levels of constraint are shown in the top and middle of Table 3.1. Reversal occurs for the strong constraint, but not for weak constraint. Intuitively, this is because stronger constraints make it less likely that a person would deviate from the requested po-
Figure 3.4: Illustration of strong and weak “con” essays for weak constraint (left, $s = .25$) and strong constraint (right, $s = .75$). In both cases, the requested position, $p$, is -2 and the essay positions, $e$, are -3 and -1 for the strong and weak essays.

sition. Therefore, when someone does deviate from the requested position by writing a weaker-than-expected essay, it is reasonable to conclude that this person must hold an attitude very different than what was requested. This is illustrated in Figure 3.4. As can be seen, the likelihood functions are both further from the requested positions and more spread out at higher constraint. When multiplied by the priors, the resulting attributions are more tempered.

Varying situation strength is not the only way to replicate the different patterns of results. The bottom two rows of Table 3.1 show what happens when the weak essays are made less ambivalent, but the strong level of constraint kept consistent. This change removes the reversal that had been obtained with the weak essays. In speculating on the failure to replicate the Jones et al. (1971) weak essay reversal, A. G. Miller (1974) does in fact note that his weak essays were not as weak as those in Jones et al. While both Jones et al. and Miller speculate that strong essays “engulf the field” (in the parlance of Heider [1958]) whereas the weak essays allow the percever more latitude to notice the situation, the model suggests that no such perceptual metaphors are necessary. Instead, both outcomes are reasonable conclusions of an internally consistent logic that does not depend on any distortions in perception, failure to notice the situation, or alteration in the underlying behavioral model being used to make the attribution. This example also makes clear that there are often multiple internally-consistent ways to obtain the same pattern of results. The model makes it possible to explore many sources of a result, thereby suggesting hypotheses for behavioral research.
3.3 Empirical Results

In addition to fitting previous research results, the model fits new data (collected for a different purpose).\footnote{These data are part of an in-progress replication of Miyamoto and Kitayama (2002), and use their essays as stimuli. In addition to varying essay position, the study varies essay length. As in Miyamoto and Kitayama’s data, this produces no effect, nor is there an interaction, $F(1, 242) = 2.37, p = .13$. Therefore, the length manipulation is not discussed further.} Participants ($N = 246$) read essays for and against the death penalty, and then learned that the author was randomly assigned the position to take. Participants then rated what they thought the author’s attitude was, how confident they were in their answer, and other perceptions (detailed next). Replicating past results, there was a significant difference between the pro and con essay attributions ($M = -0.90$ vs. $M = 0.98$, $t(244) = -8.44$, $p < .0001$). Model-based predictions were then tested, after reversing all of the relevant quantities (including those described next) for participants in the “con” essay condition.

As the model of the Jones and Harris (1967) and Miyamoto and Kitayama (2002) results showed, a skewed prior attitude distribution should result in skewed attitude attributions. In particular, attributions for essays expressing rare opinions should be closer to the midpoint that attributions for essays expressing common positions. Additionally, as judged by the variance of the posterior distributions, people should be less confident in their attributions when the expressed position is rare. This was tested by looking at participants’ self-reported prior attitude distributions, which were elicited by having people apportion 100 percentage points to three equal-sized intervals encompassing the measurement scale. On the basis of these measurements, a “skew” was calculated for each participant by taking the log ratio of the lower and upper intervals of their priors. Negative ratios imply more probability mass near the “pro” end of the scale, and positive ratios imply more probability mass near the “con” end of the scale, matching the meaning of positive and negative skew. Supporting the model’s predictions, the correlation of attribution and skew was $r = -.14$ ($p < .05$), and the correlation of confidence and skew was $r = -.20$ ($p < .01$). (Recall that after reversing the numbers in the “con” condition, positive skew indicates less probability mass near the essay’s position, accounting for the negative correlations.)

Next, the co-occurrence between the situation and attitudes was examined. As mentioned at the outset, if assignment is non-random, people might reasonably believe that the essay author’s own attitude and the assigned position are related. To test this, people were compared by whether they indicated (as intended) that the author had no control over assignment. There was a significant difference ($M = 0.73$ vs. $M = 1.21$, for no control vs. control, respectively, $t(244) = 2.16, p < .05$).

Finally, the likelihood model predictions were examined. As shown with the modeling of the Jones et al. (1971) and A. G. Miller (1974) result patterns, the model predicts that attribution extremity and overcompliance should be positively related, and that attribution extremity and perceived situation strength should be negatively related. Participants estimated overcompliance via a question asking how much weaker (or stronger) the essay
was than what they believed was expected, and strength was measured via a question about how much overall choice the author had (reversed). After partialing out the effects of skew and strength, attribution and overcompliance were positively related, \( pr = .14 \ (p < .05) \). After partialing out skew and overcompliance, attribution and strength were negatively related \( pr = -.16 \ (p < .05) \). Because higher strengths lead to more spread out likelihood functions, the model also predicts that confidence and strength should be negatively related, which was supported \( r = -.24 \ (p < .001) \).

Though these effect sizes are not large, it is none-the-less impressive that they were found at the individual level, and in a study designed to produce bias (rather than minimize it). More controlled tests in situations intended to minimize bias may produce even stronger results.

### 3.4 Conclusions

Using a simple yet plausible model of how people respond to instructions to advocate a particular opinion, this work derives a model that can postdict prior attitude attribution results, and that fits newly-collected data. Though the correspondence bias can be seen when people’s attributions are compared to the ground truth, this work suggests that these attributions could be internally consistent with people’s other beliefs and perceptions. Future work should investigate why these beliefs (e.g., about how people respond to requests) and perceptions (e.g., of the request strength or the essay extremity) don’t match reality.

Early in the history of correspondence bias research, Jones et al. (1971) conceded that correspondent inferences for constrained behavior are only wrong if every person in that situation would comply. Short of this extreme, they say that “it would be very difficult if not impossible to determine whether [a correspondent inference] should be judged as attributional distortion” (p. 77). The model presented here helps answer this question by encoding a set of assumptions mathematically, and then using the logic of Bayes’ rule to understand the implications of those assumptions. It is likely that many findings using the attitude attribution paradigm can fruitfully be reexamined in light of the added precision that this model provides.
Chapter 4

Cross-Cultural Application

Demonstrations of cultural differences are often more compelling that demonstrations of similarities, particularly when data suggest that a cognitive bias found in one culture is absent in another. This tendency is evident in cross-cultural attribution research, which, broadly speaking, has found that people in non-Western cultures do not share the Western habit of neglecting situational influences on behavior. However, such asymmetric findings pose an important question: if the attributions that people in Western cultures make are thought to be wrong, does that mean that the attributions that people in non-Western cultures make are right?

Of course, neither culture’s attributions are likely to be completely wrong or completely right, and suggestions that one way of thinking is better than another are often meticulously avoided since the independent variable is culture. However, skirting the question altogether is just as antithetical to theoretical progress as false dichotomies and other oversimplifications. Thus, for the sake of argument (and perhaps a little bit of guilty pleasure), this paper asks quite plainly, “Who’s right and who’s wrong?” and, relatedly, when it comes to the process of making attributions, “Who’s better?” The hope is that posing these questions in such a stark manner will lead to the gray area in between where the truth most often lies.

Two principal difficulties thwart attempts to answer these “correctness” questions. First, attribution occurs via several interrelated processes, each of which is subject to contextual influences. Differences in response could simply be the natural outcome of using the same cognitive processes under different contextual demands, and not indicative of differences or flaws in the processes themselves. Second, contrary to the assumption that participants in attribution experiments reason based only on the information provided by the experimenter, attribution processes draw heavily on background knowledge (Hilton, 2007). Like contextual influences on the attribution process, this background knowledge is likely to vary across cultures, meaning that the correct attribution might vary as well. Thus, while superficial cultural differences in response can readily be observed, the depth of these differences often is not or cannot be ascertained.

This paper compares the correctness of attributions in a Western culture (the United
States) and a non-Western culture (China). Rather than attempting to provide a definitive answer to “Who’s right, who’s wrong, and who’s better?”, the paper seeks to demonstrate an approach to making such comparisons responsibly. In so doing, the two difficulties just mentioned are addressed, the former via an examination of prior cross-cultural attribution results vis-a-vis contextual influences on the attribution process, and the latter via the use of a Bayesian standard for testing the coherence of people’s attributions with their perceptions and background assumptions (see Chapter 2). Though this paper’s results are just a launching off point for future exploration, some suggestions at an answer to the correctness question are given.

4.1 Background

The fundamental attribution error is the “general tendency to overestimate the importance of personal or dispositional factors relative to environmental influences” (Ross, 1977, p. 184). While also identified by other names (e.g., overattribution [Jones, 1979], the correspondence bias [Gilbert & Jones, 1986; Gilbert & Malone, 1995]), and while discussed long before having a catchy name (e.g., Heider, 1944, 1958; Ichheiser, 1949; Jones & Davis, 1965), the fundamental moniker highlights the centrality and ubiquity of the effect. Indeed, Jones (1990) suggested that it was the most robust and repeatable phenomenon in social psychology. It is no surprise, then, that the findings that people in the non-Western cultures of India (J. G. Miller, 1984) and China (Morris & Peng, 1994) did not appear to make this error should attract great attention.

Successfully avoiding one error does not mean that all possible errors were avoided. Accordingly, Morris and Peng (1994) noted that “our studies demonstrate that each culture was biased relative to the other culture but not that either was biased relative to the truth” (p. 968). Thus, possible answers to the question of who is right and who is wrong are that “both are right”, “both are wrong”, or, as usually happens, “it depends”. Regardless, resolving this conundrum requires not just comparisons between cultures, but also comparisons to a standard of correctness.

Correctness is a matter of perspective, and three such perspectives have been identified (Dunwoody, 2009). The most straightforward is “correspondence”, which requires that a person’s answer match the objective, ground truth. Given that determining the causes of people’s behavior (the ground truth) is the aim of psychological research in the first place, it is seldom possible to assess this standard in attribution research (Hilton, 2007). A more readily applied standard is “coherence”, which requires that a person’s answers be internally consistent with other beliefs that the person holds. Though it may still be troubling if people’s attributions are coherent without always being correspondent, such a finding can at least orient researchers toward the inputs to the reasoning process, and away from problems with the process itself (Hilton, 2007). Still, though coherence is often desirable

1Not to be confused with the use of “correspondence” in Correspondent Inference Theory (Jones & Davis, 1965).
to researchers, it may not be as intrinsically useful to people in everyday life. A pragmatic stance determines “correctness” according to whether people’s behaviors help them attain useful ends. (See Chapter 2 for a more detailed discussion of the correspondence, coherence, and pragmatic standards in attribution.)

Regardless of what form of correctness one aspires toward, the fact that people in different cultures respond differently to the same experimental conditions is puzzling. As a way to decide what standard for correctness is relevant to the “Who’s right and who’s wrong?” question, it helps to understand when and why attributions for the same events differ. This understanding requires looking at the social context of attributions, and at the attribution process itself.

4.1.1 Explanations: Socially-Oriented Attributions

Before examining the attribution process and culture’s role in it, attribution must be understood within the broader context of explanation. While often used interchangeably, here attribution will be identified with the cognitive process of reconstructing the chain of events leading to a particular outcome, whereas explanation will be identified with the social process wherein “someone explains something to someone else” (Hilton, 2007, p. 233, emphasis in original). Explanations provide accounts for why one or more events occurred, and while they draw on causal chains of events, they needn’t be limited to causal factors, and not all causal factors need be included (Keil, 2006). Being intended for an audience, several considerations affect which factors are included in an explanation, including: knowledge shared between conversational partners (Slugoski, Lalljee, Lamb, & Ginsburg, 1993), which kinds of factors are good explanations for intentional vs. unintentional behaviors (Malle, 1999), and social goals that the explanation itself may serve (Malle, Knobe, O’Laughlin, Pearce, & Nelson, 2000). Casual attribution is one of many cognitive processes that people use to produce an explanation, and is conceptually only responsive to the above considerations to the extent that pursuing a particular explanation dictates which attribution questions are undertaken.

Both J. G. Miller (1984) and Morris and Peng (1994) include results pertaining to explanation. J. G. Miller (1984) compared American and Indian children and adults, and found that as Americans aged, they made more use of dispositional information in explaining events than did Indians, but that there was no cultural difference in the ability to make dispositional attributions per se. Her results suggest a process of socialization wherein Americans learned to preferentially offer dispositional explanations, while Indians did not. Morris and Peng (1994) use a variety of methods, some of which relate to explanation. Their second study is an archival analysis of Chinese- and English-language newspaper accounts of two mass murders, and broadly-speaking finds that the Chinese-language accounts featured more situational factors and that English-language accounts featured more dispositional factors. (Lee, Hallahan, and Herzog [1996] replicated this finding.) Morris and Peng’s third study takes both types of factors from the newspaper accounts and asks participants to rate the extent to which each was a cause of the murder (a judgment having
to do with explanation rather than causal attribution; see Chapter 2). American participants rated dispositional causes more highly, and Chinese participants rated situational causes more highly.

These results suggest culturally-contingent preferences for dispositional or situational explanations, which, as is often the case, could be interpreted as reflections of culture-specific biases in the attribution process. However, explanations serve goals besides communicating completely accurate knowledge, making it possible that these differences result from other goals that people have when offering explanations. For example, people can successfully tailor explanations in order to convey specific impressions about themselves (Malle et al., 2000). This suggests that cultural norms of self-presentation (or other aspects of social etiquette) may be satisfied in part through how people explain events to others. Along these lines, Fry and Ghosh (1980) found that Indian children took personal responsibility for failure and credited success to luck, whereas White children took responsibility for success and assigned failure to luck, which Fry and Ghosh link to different patterns of socialization in the two groups. Tailoring explanations in this way may help a person gain social approval. In Hong Kong, for instance, Bond, Leung, and Wan (1982) found that confederates who offered self-effacing explanations for their performance on an intellectual task were better liked than confederates who offered self-enhancing explanations, and that unlike in studies in Western contexts, people who appeared more competent were not better liked.

These and other results suggest that, independent of the causes people actually believe, people in non-Western cultures may have external incentives to think of situational causes for their own and others’ behaviors, and that analogously people in Western cultures may have external incentives for focusing on dispositional causes. In fact, the desire to satisfy perceived social norms may directly influence a variety of culture-contingent behaviors. Zou et al. (2009) found that the degree to which a person believes that culture-specific values are held by most people in their culture predicts their performance on tasks that have previously shown cultural differences (including attribution), and that the predictive power of perceived consensus is as good as a person’s own stance on these values. By priming bicultural individuals with one of the two cultures the participant identified with, Zou et al. also demonstrate that perceived cultural consensus may cause, and not just correlate with, these differences. This work suggests that people tune their explanations in order to conform with social norms specific to their culture, and that at least in bicultural individuals, this tuning is flexible rather than ossified.

To summarize, while people in Western cultures broadly prefer dispositional explanations over situational explanations, people in non-Western cultures either do not show this preference, or directly prefer situational explanations over dispositional explanations. Though this difference could reflect different attribution biases, it could also reflect the fact that people tune their explanations in order to achieve social goals, some of which are no doubt culture-specific. In non-Western cultures this may entail offering more situational explanations in a number of contexts where Western cultures may reward dispositional explanations. This tuning is mediated by the values that people in one’s culture are perceived
to hold as much as it is by one’s own values, suggesting that ingratiation is a motivating factor.

The above discussion demonstrates that explanations serve a pragmatic purpose. Thus, even if a person explains his or her (say) failure in a way that does not correspond with the actual sequence of events leading to that failure, and even if a person’s explanation lacks coherence due to focusing on certain causes in excess of how powerful or important they are actually believed to have been, the explanation may help the person win the approval of others, and hence serve a different goal than veridicality. Different cultures will value different goals, and while one could in principle determine what those goals are and compare how well people in different cultures satisfy them, doing so would depart from the commonly understood nature of the fundamental attribution error as an error in social judgment (rather than a rational response to contextual demands). Thus, if an answer to the correctness question is to be found, it will be found by looking at how people construct causal models for events, which explanation only taps indirectly.

### 4.1.2 Inferential Goals and Two-Stage Attribution

Attribution has been characterized as subordinate to explanation (Anderson, Krull, & Weiner, 1996), and so it must be understood how the explanation process can affect the course of the attribution process. Two possible ways are discussed here. First, as is most often discussed, explanation processes will filter the elements of a causal model that people communicate. In this case, the causal model is not thought to be affected (Hilton & Erb, 1996). However, explanatory concerns may sometimes drive the attribution process. For instance, a person attempting to offer a face-saving explanation for a superior’s failure may need to search for reasonable situational causes. In this, the second way that explanation processes can affect attribution, explanatory concerns could direct causal search. Since people do not exhaustively explore every possible causal scenario when reconstructing a chain of events (Shaklee & Fischhoff, 1982) this may lead people to settle upon a causal model than they might not otherwise have, and that model may lack correspondence.

A mental process that can operate in different ways in response to contextual demands is likely to operate by default in the way that these demands most often push it. This could lead people to guide their search for causes in a way that matches their culture’s preferred type of explanation, even if there is no incipient strategic need to do so. In fact, there is evidence that these cultural differences in default orientation exist. For instance, Morris and Peng’s first study asks participants to rate the degree to which the focal actor in animations of social scenarios acted as a result of internal and external causes. While this sort of judgment may reflect having constructed a causal model of the event and then classified the strengths of the various causes, it may instead simply reflect a general causal orientation that would drive subsequent causal search (see Chapter 2). Regardless, Morris and Peng found that Americans rated internal causes more highly and that Chinese rated external causes more highly, and in a replication of this paradigm, Zou et al. find that this preference was mediated by perceived consensus of cultural values and responsive to cultural identity primes.
Additional evidence of default causal orientations comes from research that manipulates whether participants are cognitively busy. In general, research with this paradigm finds that cognitively busy and unbusy participants give different attributions for the same events, leading to the theory that attribution occurs via two stages: an initial, automatic, and potentially biased attribution, followed by an optional, effortful, and potentially less biased attribution (Gilbert et al., 1988). As originally studied, the initial step in the attribution process was thought to be a dispositional inference (trait attribution), and the secondary step was thought to be situational correction (noticing situational factors, and adjusting the trait attribution accordingly). For instance, Gilbert et al. (1988) found that cognitively busy participants did not make use of situational information when judging the trait anxiety or attitude of a videotaped speaker, while control condition participants did. However, Krull (1993) found that if observers are told to infer the role of situational influences on behavior, then cognitively busy participants did not incorporate information about the actor’s personality. Thus, it appears that what is automatic and what is effortful depends on the observer’s inferential goal.

Culture has been shown to affect an observer’s default inferential goal. Chiu, Morris, Hong, and Menon (2000, Study 2) found that American participants under time pressure made more dispositional attributions to individuals, but not to groups, while Chinese participants under time pressure made more dispositional attributions to groups, but not to individuals. There were no differences in the no time pressure conditions. The authors account for the cultural differences in the time pressure conditions by noting that East Asian culture conceives of groups as being more autonomous, whereas Western culture conceives of individuals as being more autonomous, thus giving American and Chinese observers different inferential goals in the group and individual conditions.

Possible support for the impact of default inferential goal comes from Knowles, Morris, Chiu, and Hong (2001), who found that American participants’ attributions of an author’s attitude were affected by a cognitive busyness manipulation, only reflecting the role of the situation when not cognitively busy. Chinese participants’ attributions, on the other hand, were not affected by the manipulation, reflecting the role of the situation in both cases. Knowles et al. interpret these results as showing that East Asians can automatically correct for situational influences on behavior, though they admit that it isn’t clear whether there was automatic correction, or whether the initial inference was not in need of correction. Another untested possibility is that Chinese participants had a default goal to diagnose the situation, and that if the American participants had been given that goal, as well, the differences would have been eliminated.

If cultural differences are the simple result of different default inferential goals, then explicitly telling participants what their inferential should be ought to eliminate any differences. Lieberman, Jarcho, and Obayashi (2005) tested this, telling American and East Asian participants to adopt either a dispositional or situational inferential goal, and presenting them with information and videotaped behavior designed to either make situational or dispositional attributions more likely. When participants in both groups were under cognitive load, they made attributions to the focus of their inferential goal. When not under
cognitive load, American participants made attributions in the direction that Lieberman et al. claim that the information and behavior ought to imply. However, East Asian participants made more situational attributions even when the behavior and information warranted a dispositional attribution, regardless of cognitive load. After ruling out the possibility that the East Asians interpreted the videotaped behavior differently, they concluded that East Asians have a “situational causality heuristic” that leads them to err on the side of situation causes.

The fact that manipulations of inferential goal and cognitive load can produce differences within both cultures clearly shows that, as judged by correspondence criteria, people in neither culture are correct all of the time. They also may or may not be correct according to coherence criteria. However, there is uncertainty regarding coherence due to the fact that different authors have judged correctness in different ways, some of which may have missed important elements of coherence. Still, even independent of coherence criteria, the effectiveness of cognitive load manipulations may suggest that people in both cultures lack coherence at least some of the time. Before that can be confidently concluded, however, it would need to be established that people perceived the same information in the same way, both in the load and no load conditions, and cross-culturally.\(^2\) This is the topic of the next section.

### 4.1.3 Attention to Situational Information and Subtle Behaviors

Making attributions requires gathering the necessary information and then processing that information. The results from the inferential goal and cognitive load studies do not make it clear whether the observed differences are due to attending to different information, or to processing that information in different ways. For instance, the “automatic correction” that Knowles et al. found in China could have been due to Chinese participants having retained information about the situation more readily than the American participants, not to their judgment process being less impaired. This view would fit with other studies that have been interpreted as showing that East Asians are more adept than Westerners at noticing situations or at perceiving subtle behavioral cues. Since a judgment is only as good as the information available, comparing coherence across cultures requires knowing whether Westerners simply do not notice this information, or whether they notice it but fail to make use of it when they should.

In general, researchers do not find major cultural differences when they run straightforward replications of paradigms that have been used to argue that people in Western cultures make biased attributions. For instance, Krull et al. (1999) found no cultural differences using the quizmaster (Ross, Amabile, & Steinmetz, 1977) and attitude attribution (Jones & Harris, 1967) paradigms. In fact, with attitude attributions, Choi and Nisbett (1998, Study 1)

\(^2\)Lieberman et al. find that while cognitive load affects how anxious the target’s behavior is seen to be, culture does not. However, other perceptions and background assumptions that influence coherence may also have varied.
found stronger correspondent inferences\(^3\) in Korea than in the United States. Likewise, Miyamoto and Kitayama (2002, Study 2) and Masuda and Kitayama (2004, Study 2) found correspondent attitude attributions in both Japan and the United States, though the pro-con difference was less in Japan than in the United States. (As will be discussed later, many of these differences may be explainable in terms of the assumed prior attitude distributions in either culture [see Chapter 3], and hence may not be indicative of cultural differences in the reasoning process.)

Differences do begin to emerge when researchers make efforts to either highlight the role of the situation or to subtly show that the actor's compliance was begrudging. For instance, Choi and Nisbett (1998, Study 2) find that while Koreans make less correspondent attributions when they must write their own essay under similar constraints as the author whose essay they later read, Americans do not respond to these manipulations. Likewise, Miyamoto and Kitayama (2002, Study 1) find that Japanese attributions are more tempered when reading a short rather than long essay, while Americans do not respond to differences in length. Finally, Masuda and Kitayama (2004, Study 1) find that Japanese participants make more tempered attributions when they believe that they themselves assigned what position another person should express, compared both to when they observed someone else make this assignment, and when the assignment process was not highlighted. In contrast, American participants responded equivalently when they either assigned or watched someone assign the position, which in both cases was only slightly less correspondent than when the process was not highlighted.

These results suggest two things. First, to the extent that it is an error to make attributions in line with constrained behavior, then this error can occur in both cultures. Second, people in East Asian cultures appear to be more responsive to certain manipulations designed to highlight the role of the situation. However, this does not mean that Americans are \textit{un}responsive to such manipulations, all of which were first attempted only in a Western context. For instance, Choi and Nisbett's manipulation of having participants write their own essay and of providing them with arguments are taken from Jones and Harris (1967, Studies 2 and 3) and Snyder and Jones (1974). In Snyder and Jones, it is not the case that people did \textit{not} respond to the manipulations, but rather that their response was not strong enough relative to the researchers' expectations. Likewise, though Miyamoto and Kitayama find that Americans did not respond to the length of the essay, other studies have found that Americans respond to essay persuasiveness (e.g., Jones & Harris, 1967; Jones et al., 1971; A. G. Miller, 1974) and to other paralinguistic cues in the essay (e.g., Blankenship & Craig, 2007), meaning that Americans may simply not have interpreted the essay length as a meaningful cue to the author's enthusiasm for the topic. Similarly, Masuda and Kitayama (2004) deliberately showed participants videos where the speaker was bland and expressionless. As they themselves note, Ishii, Reyes, and Kitayama (2003) found that Asians attend spontaneously to word tone more than content, while Americans respond spontaneously to word

\(^3\)Here, "correspondent" means an attribution in line with behavior, as in Correspondent Inference Theory (Jones & Davis, 1965).
content more than tone. The fact that the speaker was expressionless may have been (unintentionally) a cue to the speaker’s true attitude in Japan more than in the United States, which could amplify the effects of the other manipulations in Japan and mask them in the United States.

The above results can be summarized as follows. First, the attributions of people in both cultures are affected by manipulations both of their own situation when judging another person’s behavior, and of subtle variations in the behavior itself. Second, how people respond to these manipulations appears to be culturally contingent. Though East Asians may respond to some manipulations more readily than Westerners, there is ample evidence that Westerners do respond to other manipulations (which have not been tested cross-culturally), and there are reasons to suspect that some of the manipulations with which differences are found may have been more meaningful to East Asians than to Westerners. None of this means that there are not cultural differences in how readily people notice situational causes (and there are analogous cultural differences in the tendency to notice contextual and relational elements of visual scenes [Masuda & Nisbett, 2001]). What it does mean is that existing results do not definitively separate whether people noticed situational cues from whether they found those cues meaningful. What’s more, even if this distinction were regularly made, it wouldn’t be clear how variations in degree of meaning across culture ought to affect variations in the normative attribution.

4.1.4 Toward Determining Coherence

While the forgoing has not been a comprehensive review of every relevant cultural difference in attribution, it has been sufficient to suggest how to begin answering the “correctness” question. First, it was suggested that attributions often are made in service of explanatory goals, which may dictate (in ways that are culture-dependent) that people offer one kind of explanation or another. Thus, contrary to the common assumption that people are motivated to make correct attributions in order to be able to predict and control their world, it may be that people undertake causal reasoning for purposes other than learning the truth, and as such their responses may be pragmatically correct even if factually wrong (lacking correspondence).

While some apparently biased explanations may be pragmatically correct, there certainly are situations where accuracy is the primary consideration, and where cultural differences still emerge. Importantly, both cultures are affected by manipulations of things like cognitive load. Even without having correspondence criteria, this necessarily implies that neither culture makes correspondent (to ground truth) attributions all of the time. While these cognitive load results could also imply that neither culture makes coherent attributions all of the time, it isn’t clear whether cognitive busyess affects what information is attended to, or whether it affects how information is processed. What’s more, studies that purport to show cultural differences in how readily situational and subtle behavioral cues are noticed are generally ambiguous about whether the cues were not noticed, or were noticed and not used. In the event that the cues were noticed and not used, it may still be the
case that the cues were disregarded not carelessly, but because they were not considered to be meaningful.

In summary, cultural differences in explanation may be pragmatically correct, and cultural differences in attribution lack correspondence in both cultures at least some of the time. What cannot yet be answered definitively is whether there are cultural differences in coherence. Doing so requires a standard for assessing coherence, which is the topic of the next section.

4.2 Determining Coherence

Any psychological research that goes beyond mere description has a normative standard in mind. Often, this standard is trivial: if a theoretical construct is meaningful, then there should be some kind of difference in the dependent variable when a manipulation operationalizing that construct occurs. As more manipulations are included, the standard becomes more complex. Still, in most cases the expected outcomes can be justified using the language of main effects and interactions, which only involves specifying relative values.

Studies of the attribution process provide information designed to make a particular judgment normatively correct, and manipulate elements of the participant’s situation in order to enhance or disrupt part of the presumed cognitive process behind that judgment. Statistical differences are interpreted as evidence that the process was affected, and hence of the construct’s importance. However, such interpretations presume that the manipulation did not also affect the normatively correct answer. Regardless of whether such effects are thought of as interesting mediators or noisome confounds, their presence requires that the normative standard in use become more sophisticated than a mere description of relative values. Instead, the standard must be able to cope with continuous covariates. While any covariate can be controlled for statistically, claims about normativity obviously require that only the proper covariates be included, and that their impact be accounted for in the right way.

The above problem is compounded when culture is included as an independent variable. First, culture may color perception, meaning that the provided information might be understood in different ways (e.g., a behavior may be seen as stronger in one culture, or it might be irrelevant in another). If the provided information is perceived differently, then the normative judgment will likewise vary. When differences in perceptions are not accounted for, they will be mistaken for differences in the judgment process itself.

A second problem that arises in cross-cultural work is that, as Hilton (2007) notes, research participants draw on the background knowledge that they bring with them to the laboratory when answering experimenters’ questions, including assumptions about specific cause and effect relationships. A theory that fails to account for these assumptions will be incomplete, but if the background knowledge is relatively consistent across participants, then the theory may at least still be useful. However, since these assumptions likely develop
via the participant’s social ecology (Fiedler, 2007), assumptions may vary rather widely across cultures. Including culture as an independent variable makes the incomplete but previously adequate normative standards quickly become inadequate. As before, failure to recognize this change may lead to mistaken conclusions about cultural differences in the judgment process.

In summary, responsible work in attribution, particularly if done cross-culturally, requires a normative standard that can account for differences in perceptions and assumptions when specifying the rational judgment. Without such a standard, experimenters may mistake rational responses for irrational deviations, and falsely conclude that the judgment process itself is flawed. While every paradigm will have a slightly different standard, many common paradigms have the same basic form, as discussed next.

### 4.2.1 Core Judgment: Ascriptive Discounting and Augmenting

As discussed in the background, experiments that analyze open-ended explanations or categorical person and situation attributions may not directly reflect the causal attribution process, and so do not provide good evidence for the “correctness” question. However, other classical attribution paradigms are suitable. Common to these paradigms is the pitting of two causes against each other, one of which is personal, and the other of which is situational. For instance, the attitude attribution paradigm pits the author’s attitude (person) against the assignment of the position (situation), and the silent interview paradigm (Snyder & Frankel, 1976) pits the interviewee’s trait anxiety (person) against the content of the interview questions (situation).

In paradigms like these, participants are asked to attribute the level or presence of one cause (e.g., an attitude, or trait anxiety) on the basis of observed behavior (e.g., an essay, acting nervously). This judgment is made less straightforward by the presence of an additional cause (e.g., instructions, interview questions) that could also produce the behavior. In these situations, it is often rational to *discount* the uncertain cause in light of the certainty that the second cause was present, meaning either that the perceived level of the uncertain cause is less extreme (and less in line with the behavior), or that confidence that the uncertain cause was present is diminished (Kelley, 1972a). Discounting might occur in any sort of situation where a person’s behavior is initially attributed to a related personality trait, but where learning of a situational cause for the same effect might make the person less confident that the personality trait was in fact present. In any of these cases, the situation might make the person *more* confident in the dispositional cause, which is called *augmenting*.

As described here, the perceiver is making an *ascriptive* inference, meaning that the goal is to determine how one particular event happened, based on prior knowledge of possible cause and effect relationships in the world. This stands in contrast to induction, where the goal is to determine whether a particular cause and effect relationship exists, and if so, how strong it is. Causal induction involves generalizing from many particular cases to an abstract rule that applies across all similar cases, while causal ascription applies that knowledge to one particular case (Morris & Larrick, 1995; Hilton, 2007). While both causal
induction and causal ascription involve causality and relate to a person’s causal knowledge, and while both may be required to complete some causal inferences, they present distinct inferential problems (see Chapter 2).

### 4.2.2 Normative Standard

Though not all cross-cultural studies of causal attribution involve ascriptive discounting (or augmenting), a great many do, making it a good place to start when answering the correctness question. Kelley (1972a) suggested that it is rational to discount when both causes in question facilitate the effect, and that if one cause facilitates the effect and the other cause inhibits it, then it is rational to augment the facilitative cause. Kelley (1972b) expanded on these ideas by proposing that people possess causal schemata that specify how two causes combine to produce an effect. He claimed that discounting was rational with a “multiple sufficient causes” schema, while augmenting was rational with a “multiple necessary causes” schema. He also presented more complex schemata with varying implications for discounting.

While intuitive, Kelley’s ideas have some problems, most notable being his assumption that the two causes are statistically independent, and his assumption that causes deterministically lead to the associated effects. Morris and Larrick (1995) presented an adaptation of Kelley’s ideas that relaxes both assumptions. They present a series of Bayesian arguments about when it is and isn’t rational to discount, concluding that “[i]t is rational to discount one cause of an effect given an alternative cause, except when these two causes tend to occur together and they are minor, not major, causes of the effect” (p. 347). They also offer more precise standards, with the choice of standard contingent upon the sufficiency, independence, and number of causes. Because of this contingency, these standards are difficult to apply without a priori knowledge of people’s assumptions.

Chapter 2 presents an alternative Bayesian standard that is flexible enough to be applied for any two causes. Like Morris and Larrick, this standard involves three events:

- **E** An event that has occurred, and for which a cause is sought
- **T** A dispositional (trait) cause that could have led to the event
- **C** A situational cause (circumstance) that could also have led to the event

The trait and circumstance, **T** and **C**, could in fact be any causes, which is why Morris and Larrick simply refer to them as **A** and **B** and Chapter 2 refers to them as **U** (for uncertain presence) and **C** (certain presence). However, this research focuses on person vs. situation causes, and so these more suggestive symbols are used.

According to Morris and Larrick, there are two probabilities that are involved in discounting (and augmenting). The initial (dispositional) attribution is \( P(T \mid E) \), that is, the probability of the trait being present, given that the event has occurred (and implicitly, that
one cannot say whether any other circumstances were at play). The subsequent (situationally corrected) attribution is \( P(T \mid E, C) \), or the probability that the trait was present, given both that the event occurred, and that the circumstance was at play (though again, implicit in this, other circumstances may or may not have been present, as well). For \( P(T \mid E) \), Bayes’ theorem states that the normative value is \( P(E \mid T) P(T) / P(E) \). For \( P(T \mid E, C) \), Chapter 2 repeatedly applies Bayes’ rule to yield the following:

\[
P(T \mid E, C) = P(T) \cdot \frac{P(C \mid T)}{P(C)} \cdot \frac{P(E \mid C, T)}{P(E \mid C)}
\]

The three terms of the right hand side have intuitive interpretations. The first is simply the prior probability (or base rate) of the trait. The second term expresses how much more (or less) likely the circumstance is for a person with the trait, compared to the average case (cause-cause co-occurrence). Likewise, the third expresses how much more (or less) powerful the circumstance is if the person also has the trait, compared to when the circumstance is present but the trait may or may not be (the relative event likelihood). Thus, the normative trait attribution given the presence of a situational cause depends on how common the trait is to begin with, whether the trait makes the situational cause more (or less) likely, and whether the trait and the situational cause work differently in concert compared to when the trait’s presence is unknown but the situational cause is present.

The above identities provide a standard for determining whether a person’s attribution is coherent with perceptions of the event and background assumptions. It can be applied both schematically, by making assumptions about the relative sizes of the three terms, or empirically, by assessing all of the relevant quantities. The latter is what is done in this research. However, before doing so, the next section discusses how culture may affect the three constituent beliefs of the identity.

### 4.2.3 Culture and Background Assumptions

While many cultural differences in background assumptions may be idiosyncratic to the specific events, circumstances, and traits involved, there may also be “main effect” differences in some of this knowledge. Indeed, prior probability, cause-cause co-occurrence, and relative event likelihood may be good ways to operationalize more abstract concepts like “lay dispositionism” and “lay situationism” (Morris & Peng, 1994; Choi et al., 1999), or to measure things like “situation theories”, knowledge of how situations affect behavior (Gawronski, 2004). While not a primary purpose of this paper, looking for these differences now may lay the groundwork for interesting future research.

---

4 In this paper \( P(A, B) \) means \( P(A \cap B) \).

5 \( P(T \mid E, C) = P(E, C, T) / P(E, C) = P(E \mid C, T) P(C \mid T) P(T) / [P(E \mid C) P(C)] \)
Prior Probability

Some theorists have suggested that attributions ought always to be judged relative to the prior probability (or base rate) of the trait in question (Ajzen & Fishbein, 1975; Jones & McGillis, 1976), and at least in the attitude attribution literature, it is somewhat common to use the assumed mean attitude in the population as a benchmark against which to assess how much the perceiver discounted the possibility that the author believes what was expressed in the essay. It is less common to consider the base rate (or average level) of traits across cultures. However, as the standard makes clear, people in one culture might rationally make superficially more extreme attributions if the disposition in question is more common in that culture.

The prior probability can also have subtle effects on the rational attribution. Chapter 3 presents a continuous variable adaptation of this paper’s Bayesian standard, in which case the probability of the trait becomes the prior probability distribution of attitudes in the population. He shows that unless perceivers believe that the essay author was completely constrained about what position to express, the shape of the prior attitude distribution affects both the spread between the pro and con attributions, and how symmetric the attributions are about the scale midpoint. As such, the distance between the pro and con attributions is not a good measure of the degree of correspondent inference when comparing populations with two different prior attitude distributions. His results suggest that some of the cultural differences that previous authors have found may simply be the rational reflection of different prior attitude distributions in the East Asian and Western cultures being compared.

Cause-Cause Co-occurrence

Many attribution paradigms assume, but do not always specify, that the situation occurs independently of the person’s personality. If this independence isn’t made clear, then perceivers might substitute whatever assignment process they wish. However, though this is a problem for these paradigms in general, there isn’t a good reason to think that people in one culture or the other will assume any more or less random of a process, and so it’s unlikely to be an alternative explanation for many (if any) previously-observed cross-cultural differences.

More broadly, it is possible that culture may affect perceptions of how strongly everyday traits and situations are related. Gawronski (2004) argues that rather than showing that people lack an understanding of how situations affect behavior, attribution research consistently shows that people expect situations to affect behavior, and that these beliefs affect their causal judgments (though sometimes in ways that lead to more dispositional attributions, such as when knowledge of situations biases how behavior is perceived). One way that beliefs in individual responsibility for behavior and in the power of the situation could be reconciled is if people believe that individuals bring about many of the situations that they face. For instance, an anxious person might put himself or herself into more anxiety-producing situations than a person who isn’t considered dispositionally anxious. On the
other hand, another possibility is that people in cultures that favor situational attributions might see relevant situational causes as more common in general, and hence hedge their dispositional attributions more.

**Relative Event Likelihood**

The relative event likelihood term encompasses both perceptions of the extremity of the observed event and assumptions about the power of the situational and personal causes to produce that event. In situations where the behavior in question must be observed and interpreted by study participants, assessing the beliefs in this term can help to equalize cultural differences in those perceptions.

More generally, it could be that culture affects how powerful situation and person causes are, that is, how likely they are to produce a particular event. Choi et al. (1999) argue that both East Asian and Western cultures believe that personal dispositions powerfully determine behavior, but that East Asians also think in terms of situations powerfully determining behavior. In two studies of behavioral prediction, Norenzayan, Choi, and Nisbett (2002) found essentially no overall cultural differences in predicting behavior from traits, but did find that East Asians made stronger predictions of behavior from situational information when the situational information was highlighted. These findings would correspond to no cultural differences in $P(E_j|T)$, and to higher (or lower) estimates of $P(E_j|C)$ in non-Western cultures for causes that facilitate (or inhibit) the effect.

**4.3 Study**

This study is an initial attempt to make a cross-cultural comparison of the coherence of people’s attributions. Pursuant to the foregoing discussion, this study attempts to eliminate all influences on the attribution process that are peripheral to the core causal judgment, ascriptive discounting (or augmenting) of a dispositional cause in the presence of a situational cause. Coherence will be assessed by asking participants to directly estimate all of the probabilities in the normative standard, and then comparing the estimated values of $P(T|E)$ and $P(T|E,C)$ to the benchmark values calculated from participants’ other probability judgments.

This study pits one dispositional cause (actually a disjunction of two similar causes) against several situational causes, both of which could have caused one particular event. Each situational cause is assessed in a separate judgment, and the causes are presented within-subjects. These causes were taken from an earlier study in which participants in both the United States and China listed as many possible causes for the target event as they could think of. From these responses, 22 were chosen that were commonly mentioned in both countries. This study is likewise conducted in the United States and China, and the American sample is subdivided into European-American and Asian-American (born abroad and first-generation native born) subsamples. The subculture samples may help to
distinguish differences that are due to broad cultural assumptions about social causality from differences that are due to social ecology.

This study is designed to directly tap the coherence of people’s ascriptive inferences. Since the primary dependent variable is a probability judgment about the presence of the dispositional cause, this study should tap participants’ causal models directly. Participants presumably have a goal to be accurate, and are explicitly directed by the question to think about the actor’s disposition, equalizing cultural differences in inferential goals. Participants will be encouraged to reason deliberately, both by the use of quantitative probability scales as the dependent measure (Windschitl & Wells, 1996) and by the within-subjects manipulation of the circumstances (Kahneman, 2003). Since all information is presented on paper and there is little of it, it is unlikely that any cues will go unnoticed. Though the same descriptions may be interpreted by people in different ways, this should affect each judgment, and hence not affect coherence.

Hypotheses

The primary purpose of this research is to demonstrate the value of normative standards when making cross-cultural comparisons of attribution. It is largely exploratory in nature, meaning that the emphasis is more on finding new questions to ask than on definitively answering old questions. Still, a few tentative hypotheses can be suggested.

The primarily expected difference is that the European-American participants will give higher estimates of \( P(T \mid E) \) than the Chinese participants. No mean-level differences are expected for \( P(T \mid E, C) \), though there may be differences at the cause level. Any such differences should be accounted for by the normative model, and no differences are expected in how coherent any groups’ judgments are.

Since background assumptions are being measured, some subsidiary hypotheses can be suggested on the basis of the earlier review. First, no differences are expected in \( P(T) \). Second, if there are differences in \( P(C \mid T) / P(C) \) (cause-cause co-occurrence), the European-American participants’ estimates are expected to deviate from one (indicating greater association between the trait and the circumstances) more often than in the other groups. Third, if there are differences in the mean level of \( P(C) \), it will be higher in China. Finally, on the basis of past research, there should be no differences in \( P(E \mid T) \), though if differences do occur it should be higher in the European-American group. If different, the mean level of \( P(E \mid C) \) should be higher in China for facilitative causes and lower for inhibitory causes.

4.3.1 Methods

Participants

This study was run in both the United States and China. In the United States the experiment was included as the first portion of a separate experiment completed for course credit, and for which there were no recruiting criteria. Cultural groups were formed from
Table 4.1: Sample criteria and characteristics.

<table>
<thead>
<tr>
<th></th>
<th>Eur.-Amer.</th>
<th>First Gen.</th>
<th>Foreign-Born</th>
<th>Chinese</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parents Born</td>
<td>US, Can., W. Eur.</td>
<td>E. Asian country</td>
<td>E. Asian country</td>
<td>China</td>
</tr>
<tr>
<td>Born</td>
<td>United States</td>
<td>United States</td>
<td>E. Asian country</td>
<td>China</td>
</tr>
<tr>
<td>Lived</td>
<td>United States</td>
<td>United States</td>
<td>US after age 5</td>
<td>China</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$M = 10.26$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$SD = 4.84$</td>
<td></td>
</tr>
<tr>
<td>Ethnicity</td>
<td>Eur.-Amer.</td>
<td>Asian(-Amer.)</td>
<td>Asian(-Amer.)</td>
<td>Han (except 1)</td>
</tr>
<tr>
<td>$N$</td>
<td>31</td>
<td>15</td>
<td>15</td>
<td>40</td>
</tr>
<tr>
<td>Age</td>
<td>19.6 (1.26)</td>
<td>19.8 (0.86)</td>
<td>20.1 (0.58)</td>
<td>20.9 (2.26)</td>
</tr>
<tr>
<td>Female</td>
<td>49.8%</td>
<td>51.2%</td>
<td>48.4%</td>
<td>59.5%</td>
</tr>
<tr>
<td>Psych. Majors</td>
<td>53.9%</td>
<td>49.3%</td>
<td>64.8%</td>
<td>49.9%</td>
</tr>
</tbody>
</table>

this larger sample. In China, the experiment was conducted by itself, and participants were recruited from an undergraduate Psychology class and paid 10 yuan for their participation. The sample descriptions are in Table 4.1.

Materials

Participants responded to a scenario in which the event to be explained was that a student was late to an in-class presentation that he or she was scheduled to deliver (see Appendix B.1). The putative dispositional cause was that the person was “irresponsible or absent-minded”.\(^6\) The 22 circumstances are shown in abbreviated form in Table 4.2 and in full form in Appendix B.2.

Procedure

Participants completed the study on a computer. In China, participants were in a room with up to four other people, but were seated in a manner that made it impossible to see other screens. In the United States, participants either completed the experiment in private booths, or in a room with up to five other people, again seated in a way that made it impossible to see other screens. Participants were only allowed to begin when all scheduled participants had arrived and completed their consent forms.

Once participants began, they were led through the entire experiment by the software. In the United States, all participants completed the experiment in English, while in China all participants completed the experiment in Chinese. After completing a demographic questionnaire, participants responded to a series of probability judgments, which are reproduced in Appendix B.3. These were asked for on five separate screens as follows:

1. $P(T \mid E)$ and $P(T)$

\(^6\)Including both possibilities was done to encompass the broadest range of dispositional attributions. The disjunction is not a problem mathematically but may have created other problems, discussed later.
2. \( P(T \mid E, C) \) for each circumstance

3. \( P(C) \) and \( P(C \mid T) \) with each circumstance as a row and the two judgments as columns; participants were explicitly told that the second quantity should be lower if the trait makes the circumstance less likely, higher if the trait makes the circumstance more likely, and the same otherwise

4. \( P(E \mid T) \) and \( P(E) \)

5. \( P(E \mid C) \) and \( P(E \mid C, T) \), again with circumstances in rows and judgments in columns, and with explicit instructions regarding their relative values

On each screen where per-circumstance ratings were necessary, the circumstances appeared in a different random order.

### 4.3.2 Results

In the following, comparisons will be made across all three groups with omnibus tests. Additionally, comparisons will be made between the European-American and Chinese groups alone, as these should theoretically show the most extreme differences. As in any exploratory study, alpha inflation is a concern. Since this study will no doubt require followup, error correction procedures are not used, leaving the task of sorting out unanticipated differences for future work.

#### Prior Probability

The mean prior probabilities, \( P(T) \), in the Chinese, Foreign-Born, First-Generation, and European-American groups were 29.4\% \((SD = 16.2\%)\), 27.5\% \((SD = 21.7\%)\), 12.5\% \((SD = 11.6\%)\), and 19.3\% \((SD = 11.9\%)\) respectively. Contrary to expectations, there are significant differences, \( F(3, 97) = 5.63, p < .01 \), including the difference between the Chinese and European-American groups, \( t(69) = 2.90, p < .01 \).

#### Dispositional Attribution

The mean dispositional attributions, \( P(T \mid E) \), in the Chinese, Foreign-Born, First-Generation, and European-American groups were 36.9\% \((SD = 28.7\%)\), 48.8\% \((SD = 31.1\%)\), 38.7\% \((SD = 33.1\%)\), and 56.9\% \((SD = 27.6\%)\) respectively. However, given the cultural differences in \( P(T) \), the more relevant quantity is \( P(T \mid E) - P(T) \). The mean values for this difference were 7.4\% \((SD = 29.2\%)\), 21.3\% \((SD = 33.9\%)\), 26.2\% \((SD = 26.7\%)\), and 37.6\% \((SD = 28.4\%)\), in the same order as before. There are significant differences, \( F(3, 97) = 6.27, p < .001 \), including the Chinese and European-American groups, \( t(69) = 4.36, p < .0001 \). In line with previous research, European-Americans made stronger dispositional attributions, while in the Chinese group the difference only approached being different from zero \( t(39) = 1.61, p = .11 \).
Normativity

The normative value for the initial dispositional attribution is $P(E \mid T)P(T)/P(E)$. The normative value was calculated for each participant, and was capped at one where applicable (about 30% of the time). Then, the difference between each person’s reported value and normative value was taken. In the Chinese, Foreign-Born, First Generation, and European-American groups, the mean differences were $40.6\% \ (SD = 39.2\%), \ 21.1\% \ (SD = 42.3\%), \ 2.8\% \ (SD = 32.1\%), \ and \ 5.3\% \ (SD = 42.0\%)$, respectively. (Positive values indicate that the actual attribution was less extreme than the normative value.)

The overall mean was non-zero ($F(1, 97) = 16.3, p < .001$), and there were group differences ($F(3, 97) = 5.90, p < .001$). Using Tukey’s Honestly Significant Difference (HSD) test with a familywise error rate of .05, the Chinese group is different from the First Generation and European-American groups, with no other differences significant. Each group’s difference scores were also compared to zero. The difference was significant in the Chinese group ($t(39) = 6.54, p < .0001$) and approached significance in the Foreign-Born group ($t(14) = 1.94, p = .07$), but is clearly not significant in the First Generation ($t(14) = .34, p = .74$) and European-American ($t(30) = .71, p = .49$) groups.

Source of Difference

Taken together, these results suggest that while European-Americans make dispositional attributions and Chinese do not, the European-Americans are making coherent attributions and the Chinese are not. The reasons for this apparent difference must be determined.

First, it should be noted that there was a potentially important translation issue. The English phrasing for $P(T \mid E)$ is “What percentage of the people in this class who are late for their presentations do you suppose are irresponsible or absent-minded?” The Chinese translation is similar, except that it asks how many are late because they are irresponsible or absent-minded, a subtle difference that escaped notice during the backtranslation process. The fact that the Chinese and Foreign-Born East Asian groups perform so similarly casts doubt on this having mattered. However, in order to rule it out entirely, $N = 48$ participants in China completed the dispositional attribution portion of the experiment, with half using the original translation and half using the correct translation. There was no difference in the dispositional attribution, $P(T \mid E) - P(T)$, $t(46) = .25, p = .80$, with a mean value of $13.1\% \ (SD = 29.0\%)$, which was itself significantly different from zero ($t(47) = 3.14, p < .01$). The normative values were calculated as before, and while there were no differences due to phrasing ($t(46) = 0.46, p = .65$), the mean difference was non-zero ($t(46) = 4.43, p < .0001$), with a mean of $22.5\% \ (SD = 34.9\%)$. Thus, the replication sample made a slightly stronger dispositional attribution (though still smaller than the European-American group), and this attribution was significantly less strong than the normative value.

Of course, the estimates that make up the normative values are not infallible. Differ-
ences in $P(T)$ have already been noted. There are no strong differences in $P(E)$, $F(3, 97) = 1.86, p = .14$, with the Chinese and European-American groups not different ($t(69) = 0.84, p = .40$). Likewise, though there are group differences in $P(E | T)$ ($F(3, 97) = 2.89, p < .05$), the Chinese and European-American groups are not different ($t(69) = 0.51, p = .61$). (The omnibus difference appears to be due to lower estimates in the First Generation group.)

Given that both $P(T | E)$ and $P(T)$ show cultural differences, whereas the other probabilities are essentially the same, the apparent non-normativity in the Chinese (and Foreign-Born) samples could be due to a dispositional attribution that is too weak, or a prior probability that is too high. (The average value of $P(T)$ in the Chinese replication sample was 26.9%, and didn’t differ by condition.) For comparison, the “normative” value of $P(T)$ was calculated as $P(T | E)P(E)/P(E | T)$ using the reported values, and capped at one. There were no group differences in this value ($F(3, 97) = 0.94, p = .43$), which had a mean of 16.4% ($SD = 19.5\%$). (In the replication sample, the value was 15.5%). This may suggest that the Chinese group overestimated $P(T)$. Regardless, it is clear that the separation between the two probabilities in the Chinese data is not normative.

### Situational Attribution

Without a normative standard, attributions would be compared across cultures directly. In order to illustrate the value of the standard, the raw values of $P(T | E, C)$ were compared cross-culturally. First, the per-participant mean attribution was calculated across the 22 circumstances, and compared between groups. The overall difference was significant, $F(3, 97) = 4.75, p < .01$, as were the Chinese and European-American groups, $t(69) = 3.51, p < .001$. The mean attribution was higher in the European-American group, 47.8% vs. 32.6%. For illustration, the attributions in the European-American and Chinese groups were compared on a cause-by-cause basis, and were significantly different in 13 of 22 cases (59% of the time).

The simplest way to compare these attributions cross-culturally is to account for cultural differences in the base rate of the trait. The raw values of $P(T | E, C) - P(T)$ were analyzed as above. The overall mean difference was significant, $F(3, 97) = 7.11, p < .001$, as were the Chinese and European-American means, $t(69) = 4.54, p < .0001$, with the Chinese mean (3.2%) lower than the American mean (28.5%). The Chinese mean was not significantly different from zero, $t(39) = 1.07, p = .29$. On a cause-by-cause basis, 20 of 22 causes produced different attributions.

### Situational Correction

Another way to compare cultures is to look at how much discounting (or augmenting) occurred, which can be measured with $P(T | E, C) - P(T | E)$. If this value differs significantly from zero (and is negative), then there was discounting. As above, the mean across the 22 causes was calculated per participant. There were no overall cultural differences,
Table 4.2: Discounting (negative numbers) and augmenting (positive numbers) by group.

<table>
<thead>
<tr>
<th>Event</th>
<th>All (^1)</th>
<th>Ch (^2)</th>
<th>For (^3)</th>
<th>1st (^4)</th>
<th>E-A (^5)</th>
<th>Omni. (^6)</th>
<th>Ch–E-A (^7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>traffic is bad</td>
<td>-24.4**</td>
<td>-18.6***</td>
<td>-23.1*</td>
<td>-19.4+</td>
<td>-20.5***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>last minute prep</td>
<td>0.5</td>
<td>-14.7**</td>
<td>1.8</td>
<td>7.9</td>
<td>-2.8</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>urgent phone call</td>
<td>-27.1***</td>
<td>-13.9**</td>
<td>-27.7**</td>
<td>-29.2**</td>
<td>-26.9***</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>slow getting dressed</td>
<td>7.5</td>
<td>1.4</td>
<td>6.5</td>
<td>11.8</td>
<td>9.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bus was delayed</td>
<td>-23.7***</td>
<td>-11.6*</td>
<td>-17.7</td>
<td>-20.8*</td>
<td>-27.3***</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>misread the clock</td>
<td>-7.9</td>
<td>3.1</td>
<td>5.2</td>
<td>-10.2</td>
<td>-7.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hit snooze button</td>
<td>6.5*</td>
<td>10.3</td>
<td>5.3</td>
<td>12.8</td>
<td>8.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>got very lost</td>
<td>-8.2**</td>
<td>-12.2**</td>
<td>-11.1</td>
<td>-19.7+</td>
<td>-4.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>alarm not set</td>
<td>9.0***</td>
<td>14.6*</td>
<td>15.1</td>
<td>19.4+</td>
<td>11.5+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>minor car crash</td>
<td>-33.3***</td>
<td>-21.9***</td>
<td>-34.7**</td>
<td>-30.2**</td>
<td>-44.1***</td>
<td>*</td>
<td>**</td>
</tr>
<tr>
<td>prior class late</td>
<td>-29.0***</td>
<td>-22.7***</td>
<td>-28.7**</td>
<td>-31.6**</td>
<td>-30.7***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>printer breaks down</td>
<td>-23.1***</td>
<td>-11.5*</td>
<td>-16.5</td>
<td>-20.6*</td>
<td>-21.4***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>slides are misplaced</td>
<td>9.0*</td>
<td>7.7</td>
<td>4.5</td>
<td>17.6</td>
<td>11.6+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>talking with friend</td>
<td>12.9***</td>
<td>17.8*</td>
<td>16.5</td>
<td>16.3</td>
<td>17.6*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>forgot talk time</td>
<td>18.4***</td>
<td>20.8**</td>
<td>18.7</td>
<td>24.8*</td>
<td>18.8**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>feels too nervous</td>
<td>-25.7***</td>
<td>-21.6***</td>
<td>-12.7</td>
<td>-12.7</td>
<td>-32.9***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>more pressing task</td>
<td>-15.7**</td>
<td>-4.2</td>
<td>-14.8</td>
<td>-16.2</td>
<td>-14.0+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mishap on way</td>
<td>-27.5***</td>
<td>-12.2*</td>
<td>-16.3+</td>
<td>-30.7**</td>
<td>-30.7***</td>
<td>+</td>
<td>*</td>
</tr>
<tr>
<td>left stuff behind</td>
<td>-0.4</td>
<td>-4.8</td>
<td>-2.0</td>
<td>9.8</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>roommate needs help</td>
<td>-23.2***</td>
<td>-16.2**</td>
<td>-18.8*</td>
<td>-16.0</td>
<td>-16.4*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>just became sick</td>
<td>-27.0***</td>
<td>-16.4**</td>
<td>-27.3**</td>
<td>-25.1**</td>
<td>-19.4**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>don’t really care</td>
<td>14.7***</td>
<td>33.2***</td>
<td>11.8</td>
<td>9.3</td>
<td>21.3**</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: All numbers are percentages. Ch = Chinese, For = Foreign Born, 1st = First Generation, E-A = European-American; 1 - \(F(1, 97)\), 2 - \(t(39)\), 3 - \(t(14)\), 4 - \(t(14)\), 5 - \(t(30)\), 6 - \(F(3, 97)\), 7 - \(t(69)\); + - \(p < .10\), * - \(p < .05\), ** - \(p < .01\), *** - \(p < .001\)

\(F(3, 97) = .17, p = .9\), nor were the Chinese and European-American means different, \(t(69) = .73, p = .47\). On a cause-by-cause basis, only three differences were significant. These results, as well as the comparisons of the overall and group means with zero (i.e., no discounting or augmenting), are shown in Table 4.2.

As can be seen, there are clear cases of both discounting (e.g., “traffic is bad”, “just became sick”) and augmenting (e.g., “forgot talk time”, “don’t really care”). There are also cultural differences (“bus was delayed”, “minor car crash”, and “mishap on way”), with the European-Americans discounting more. This could be due to the generally stronger initial dispositional attribution in the European-American group.

Normativity

The normative value of \(P(T \mid E, C)\) depends on the value of \(P(T)\), which is potentially problematic if the Chinese participants did indeed report this value with bias. In order
Section 4.3. Study

86

to eliminate the influence of $P(T)$, the log ratio of $P(T \mid E, C)$ and $P(T \mid E)$ is taken. Negative values correspond to discounting and positive values correspond to augmenting, and because the log is taken, the degree of difference is symmetric about zero. The normative value of this ratio is:

$$\log \left[ \frac{P(C \mid T)}{P(C)} \cdot \frac{P(E \mid C, T)}{P(E \mid C)} \cdot \frac{P(E)}{P(E \mid T)} \right]$$

In order to avoid division by zero or an infinite log, all zeros were changed to .5%.

Ideographic fit was first determined by doing a paired samples $t$-test for each participant. The fit was generally not very good, with 76% of participants showing a significant difference ($p < .05$). An alternative approach is to correlate the empirical and normative values. (Doing so eliminates the role of $P(E \mid T)/P(E)$ and $P(T \mid E)$ since they are constant across circumstances, making this test of $P(T \mid E, C)$ alone.) This approach was more successful, with a mean correlation of .46, which is different from zero ($t(100) = 11.56, p \approx 0$). The omnibus ANOVA was not significant, $F(3, 97) = 1.85$, $p = .14$, while the Chinese and European-American groups were marginally different, $t(69) = 1.82, p = .07$, with the Chinese group less normative on average ($\bar{r} = .40$) than the European-American group ($\bar{r} = .54$).

Given that each person made 114 probability estimates, it would be surprising if unreliability was not a problem. One way to correct for unreliability is to average several people’s estimates, in particular, all participants in each cultural group. This was first done with the above correlational approach, using nonparametric bootstrap resampling with 10,000 replications to generate confidence intervals. The median correlations (and 95% confidence intervals) for the Chinese, Foreign-Born, First Generation, and European-American groups were .88 (.78, .94), .76 (.40, .90), .72 (.51, .85), and .86 (.73, .93). None of the differences is significant. The higher correlations using averaged scores is because each individual’s correlation is attenuated by error, whereas the values based on averaging estimates are relatively error-free. The wider confidence intervals and lower median scores in the middle two groups can be attributed to the smaller sample sizes, which do less to mitigate error.

Strong correlations are still possible in the context of a few outliers, and so it is worth checking whether there are individual causes that appear to lead to non-normative attributions. This was done by predicting $P(T \mid E, C)$ from its normative counterpart, and looking at the residuals. In this model, $P(T)$ was left out, meaning that the coefficient in the bivariate regression model includes both $P(T)$ and a multiplicative scaling factor.

The results are shown in Table 4.3. Given that the Foreign-Born and First Generation groups are appreciably smaller than the other two groups, only the Chinese and European-American results are shown. Only one cause (“traffic is bad”) shows a cultural difference, with the Chinese participants being less dispositional and the European-Americans being more dispositional than is predicted. Several causes (“slow getting dressed”, “minor car crash”, “mishap on way”) show the same pattern of deviation (or nearly so), while two

---

7 All aggregations and statistical tests with correlations reflect having used Fisher’s $r$-to-$Z$ transformation.
Table 4.3: Residuals for predicting $P(T \mid E, C)$ from its normative value, excluding $P(T)$.

<table>
<thead>
<tr>
<th>Cause</th>
<th>Chinese 2.5%</th>
<th>Chinese 50%</th>
<th>Chinese 97.5%</th>
<th>Eur.-Amer. 2.5%</th>
<th>Eur.-Amer. 50%</th>
<th>Eur.-Amer. 97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>traffic is bad</td>
<td>-10.6</td>
<td>-4.8</td>
<td>-0.1</td>
<td>2.9</td>
<td>8.0</td>
<td>13.2</td>
</tr>
<tr>
<td>last minute prep</td>
<td>-0.1</td>
<td>4.3</td>
<td>9.0</td>
<td>-10.8</td>
<td>3.1</td>
<td>14.6</td>
</tr>
<tr>
<td>urgent phone call</td>
<td>-7.8</td>
<td>-1.6</td>
<td>4.4</td>
<td>-11.5</td>
<td>-3.0</td>
<td>4.3</td>
</tr>
<tr>
<td>slow getting dressed</td>
<td>1.9</td>
<td>11.3</td>
<td>20.2</td>
<td>-2.2</td>
<td>10.4</td>
<td>21.1</td>
</tr>
<tr>
<td>bus was delayed</td>
<td>-4.6</td>
<td>0.5</td>
<td>5.7</td>
<td>-8.8</td>
<td>-1.5</td>
<td>6.0</td>
</tr>
<tr>
<td>misread the clock</td>
<td>-10.3</td>
<td>-2.2</td>
<td>5.4</td>
<td>-7.7</td>
<td>2.4</td>
<td>12.1</td>
</tr>
<tr>
<td>hit snooze button</td>
<td>-1.2</td>
<td>6.9</td>
<td>14.5</td>
<td>-2.9</td>
<td>11.6</td>
<td>22.2</td>
</tr>
<tr>
<td>got very lost</td>
<td>-12.4</td>
<td>-5.6</td>
<td>0.6</td>
<td>-6.3</td>
<td>6.2</td>
<td>17.8</td>
</tr>
<tr>
<td>alarm not set</td>
<td>-22.0</td>
<td>-12.9</td>
<td>-0.9</td>
<td>-12.6</td>
<td>-1.3</td>
<td>11.0</td>
</tr>
<tr>
<td>minor car crash</td>
<td>-15.8</td>
<td>-9.6</td>
<td>-3.8</td>
<td>-33.1</td>
<td>-22.6</td>
<td>-13.3</td>
</tr>
<tr>
<td>prior class late</td>
<td>-10.9</td>
<td>-6.3</td>
<td>-1.5</td>
<td>-8.1</td>
<td>-1.3</td>
<td>4.6</td>
</tr>
<tr>
<td>printer breaks down</td>
<td>-7.8</td>
<td>-1.0</td>
<td>5.9</td>
<td>-13.3</td>
<td>-2.9</td>
<td>4.4</td>
</tr>
<tr>
<td>slides are misplaced</td>
<td>-8.6</td>
<td>2.1</td>
<td>11.2</td>
<td>-19.7</td>
<td>-7.6</td>
<td>7.3</td>
</tr>
<tr>
<td>talking with friend</td>
<td>-13.3</td>
<td>2.7</td>
<td>18.9</td>
<td>-19.3</td>
<td>-1.1</td>
<td>17.1</td>
</tr>
<tr>
<td>forgot talk time</td>
<td>-1.3</td>
<td>12.0</td>
<td>24.9</td>
<td>-8.2</td>
<td>7.7</td>
<td>20.1</td>
</tr>
<tr>
<td>feels too nervous</td>
<td>-7.8</td>
<td>-1.6</td>
<td>4.3</td>
<td>-15.2</td>
<td>-5.9</td>
<td>3.1</td>
</tr>
<tr>
<td>more pressing task</td>
<td>-3.1</td>
<td>4.8</td>
<td>12.0</td>
<td>-2.8</td>
<td>7.6</td>
<td>15.8</td>
</tr>
<tr>
<td>mishap on way</td>
<td>-12.4</td>
<td>-4.9</td>
<td>1.1</td>
<td>-13.8</td>
<td>-7.5</td>
<td>-2.0</td>
</tr>
<tr>
<td>left stuff behind</td>
<td>-4.8</td>
<td>1.5</td>
<td>7.4</td>
<td>-14.3</td>
<td>-4.6</td>
<td>5.0</td>
</tr>
<tr>
<td>roommate needs help</td>
<td>-5.2</td>
<td>0.6</td>
<td>6.2</td>
<td>-23.6</td>
<td>-8.4</td>
<td>2.5</td>
</tr>
<tr>
<td>just became sick</td>
<td>-12.4</td>
<td>-4.3</td>
<td>3.5</td>
<td>-12.0</td>
<td>-1.1</td>
<td>7.1</td>
</tr>
<tr>
<td>don’t really care</td>
<td>-4.3</td>
<td>8.8</td>
<td>22.8</td>
<td>-2.7</td>
<td>16.0</td>
<td>29.2</td>
</tr>
</tbody>
</table>

NOTE: Table shows bootstrap confidence quantiles, based on 10,000 replications. Negative values indicate that the actual value is less dispositional than the predicted (normative) value. Italicized causes indicate confidence intervals that exclude zero in one or both groups.
causes ("alarm not set", "prior class late") are less dispositional than is normative in the
Chinese group only. No readily apparent explanations present themselves, and so the results
are merely noted for future reference.

As another benchmark, the Chinese and European-American groups are compared on
the normative amount of discounting or augmenting, using the log ratio (base 2) of the
normative values for $P(T \mid E, C)$ and $P(T \mid E)$. The results are shown, with 95% bootstrap
confidence intervals, in Figure 4.1. Comparing with Table 4.2, the cases where cultural
differences are found ("bus way delayed", "minor car crash", "mishap on way") do not
appear to show normative differences. Of the three clear normative differences ("last minute
prep", "roommate needs help", "left stuff behind"), one also shows a marginal difference
in Table 4.2 ("last minute prep"), but the other two do not. Again, it is not clear why these
mismatches between the normative standard and the actual answers occur, but they are noted
for future reference.

**Subsidiary Predictions**

It was predicted that European-Americans might see traits as more predictive of situations. This can be assessed by looking at the average value of $P(C \mid T)/P(C)$. Values further from one indicate greater predictive power. To assess this, the absolute value of the log ratio of each person’s estimates of $P(C \mid T)$ and $P(C)$ were calculated and then averaged per participant across the 22 circumstances. These averages were compared across groups. There was a marginally significant overall difference, $F(3, 97) = 2.48$, $p = .066$, and a marginally significant difference between the European-American and Chinese groups, $t(69) = 1.86$, $p = .067$, with a higher mean in the European-American group.

It was also predicted that Chinese participants might have higher average estimates
for $P(C)$. This was tested by computing each person’s average value of $P(C)$ across the
circumstances, and then comparing these averages across groups. Again, the overall test
was marginally significant, $F(3, 97) = 2.68$, $p = .051$, and the difference between the
European-American and Chinese groups was marginally significant, $t(69) = 1.68$, $p = .10$,
with the mean estimate higher in the Chinese group.

As already discussed, though there was an overall difference in $P(E \mid T)$, the Chinese
and European-Americans were not different, which was as predicted. It was also predicted
that the Chinese participants would have higher estimates of $P(E \mid C)$ for facilitative causes
and lower estimates for inhibitory causes. All of the causes in this experiment appear to be
facilitative (both intuitively, and since $P(E \mid C, T)$ exceeds or equals $P(E \mid C)$ for all
causes), which makes sense given that they originate from an experiment where people
listed situational causes that would facilitate being late. Thus, higher values are expected
in China for $P(E \mid C)$ across all causes. This was not supported overall, $F(3, 97) = 0.63$,
$p = .60$, nor for the Chinese and European-American groups, $t(69) = 1.26$, $p = .21$. There
were two cause-by-cause differences (adjusting for alpha inflation), one in each direction.
Figure 4.1: Comparison of the normative values for $\log_2[P(T \mid E, C)/P(T \mid E)]$ for European-American (upper, dashed line) and Chinese (lower, dot-dashed line) participants, ordered by mean value. Lines indicate 95% confidence intervals based on bootstrap resampling.
4.3.3 Discussion

Overall, previously found cultural differences were replicated. Chinese and Foreign-Born Asian participants made less dispositional initial attributions, both when judged by \( P(T \mid E) \) alone, and when adjusted against \( P(T) \), in which case Chinese attributions were barely different from the base rate. European-American participants made strong dispositional attributions, as did First-Generation Asian-Americans. When situational causes were introduced, there were again cultural differences, though the number depended on how differences were judged: using raw attributions, about 60% of the circumstances showed differences (with European-Americans the highest), and using attributions adjusted for the base rate nearly all circumstances showed differences. The number of differences greatly diminished when the situational attribution was compared to the initial dispositional attribution.

Results for coherence were mixed. Unexpectedly, the European- and First-Generation Asian-American groups appeared to make more normative dispositional attributions. Given that there were no cultural differences in the probability of the event given the trait, or in the base rate for the event, it appears that the lack of coherence in the Chinese and Foreign-born groups may stem from an overestimate of the base rate of the trait, or an overly conservative dispositional attribution. Though there was a small translation error in the Chinese version of the study, it was not the cause of this difference. For the situational attributions, coherence was harder to assess. The difference between people’s normative benchmarks and their actual attributions was usually significantly non-zero. However, correlations between the benchmark and actual values were reasonably good at the individual level (\( r \approx .45 \)) and when within-group aggregate probabilities were used (\( r \approx .8 \)). The Chinese group may have been less coherent at the individual level, but there was no difference in the coherence of the aggregate measures.

On a cause-by-cause basis, results were mixed. There were causes where people’s attributions departed (on aggregate) from the normative benchmark, sometimes across cultures, sometimes in one culture only. Additionally, there were cases where the normative benchmarks were different across cultures, but where the actual attributions were identical. These anomalies may disappear with replication, and if not will require future study to understand.

Background assumptions themselves showed some differences, and a fair amount of similarity. As already noted, there was an unexpected difference in the base rate of the trait. Additionally, as hypothesized, the European-American group perceived slightly higher correlation between the trait and the circumstances than did the Chinese group, who themselves may have had slightly higher base rates for the circumstances. However, there were no regular cultural differences in the causal power of the trait or circumstances, in contradiction to previous research (Norenzayan et al., 2002).
4.4 General Discussion

4.4.1 Assessing Normativity

Overall, the study illustrated that more cultural differences will be detected without considering background beliefs than with, though making within-subjects comparisons was more facile than using per-person normative benchmarks. Unfortunately, many process manipulations make within-subjects designs impossible, and within-subjects designs do not reveal which perceptions and assumptions differ across cultures and conditions. Thus, normative benchmarks are still necessary, which is why the biggest disappointment of the study was that people’s judgments did not cohere very well at the individual level, particularly on a per-circumstance basis. However, before concluding that people’s judgments are not coherent, more effort needs to be put into learning how to properly elicit the relevant beliefs.

One problem specific to the study itself was the use of the disjunctive trait “irresponsible or absent-minded”. While the mathematics of the Bayesian identities do not technically change when using $T$ as opposed to $(T_1 \cup T_2)$, it may be that participants focused on one member of the pair more than the other, and that which trait was focused upon depended either upon the circumstance being considered, or, more vexingly, the probability judgment being made. The latter possibility in particular would tend to reduce the coherence of the results.

Another limitation of this particular study design was the sheer number of judgments participants had to make. While manipulating the circumstances within-subjects may have increased coherence by highlighting the importance of the manipulation (Kahneman, 2003), fatigue may have erased these gains. Morris and Larrick (1995) also elicit probability judgments and find reasonable coherence, but they only ask each participant to make one judgment, plus its associated background judgments.

Eliminating the above problem might improve fit, but additional measures will likely be necessary. Future studies should take better account of the literature on eliciting expert beliefs (see O’Hagan et al., 2006). For instance, the study could have done a better job encouraging an outside view (Lagnado & Sloman, 2004), which generally increases the quality of people’s judgments. It also seems likely that people had trouble making judgments where the state of one cause was uncertain ($P(C)$ and $P(E \mid C)$), so future research might attempt to estimate these quantities by marginalizing (e.g., $P(C) = P(C \mid T)P(T) + P(C \mid \overline{T})P(\overline{T})$). Of course, this would compound the effect of the cross-cultural difference in $P(T)$, itself an anomaly requiring followup.

In addition to solving the above issues, future studies should use different scenarios, and should include both facilitative and inhibitory causes. The paper’s normative identity can also be adapted to specifically match the inference behind attitude attributions (see Chapter 3) or other paradigms (see discussion in Chapter 2). The advantages of such an approach are that the model can be made to directly match the response that participants typically give, e.g., an attitude estimate on a bipolar scale rather than the probability of holding one particular attitude. The background assumptions that the response depends on
Section 4.4. General Discussion

can also be elicited with more straightforward questions (e.g., how much weaker or stronger than was requested do you think this essay is?) that are more familiar to participants than probability judgments, and less prone to error.

It will not always be practical to proactively measure the assumptions necessary to assess coherence, particularly with process manipulations designed to short-circuit deliberate thought. If apparently biased responses are observed in such cases, Chapter 2 recommends that researchers assume the stance of a defendant’s lawyer in a criminal trial, and remember that participants are “normative until proven biased” rather than “biased until proven normative”. Thus, rather than assuming a bias because participants don’t reason in the way that seems normative to the researcher, researchers should apply normative standards in reverse to ask what beliefs participants would need to hold in order for their stated judgment to be coherent, and then attempt to determine if those are the beliefs people do hold. Of course, many researchers do this already, and so their reasoning should hopefully be improved by the more explicit standard used here.

4.4.2 Cross-Cultural Methodology

It is all too easy to find cultural differences. It is less straightforward to understand why they occur. However, if cross-cultural research is to expand theories in psychology rather than just catalog differences, understanding the overriding unity that produces apparent differences is essential. This may place a special burden on cross-cultural researchers; the use of vague and intuitive normative standards is endemic to social judgment research (cf. Gigerenzer, 2009), but more problematic when background assumptions begin to vary across cultures.

One way to improve the state of research is with more explicit, quantifiable, and objectively-disprovable models (Gigerenzer, 2009). Such models can be separated into process models, which attempt to account for how the inputs to a process are transformed into the outputs, and product models, which specify the relationship between inputs and outputs without being concerned about psychological plausibility (Sun, 2008). While process models are most researchers’ end goal, proper product models have the advantage of being able to encode and test normative expectations, which may include counterintuitive implications. Product models thus help separate the “figure” of aberrant judgments from the “ground” of coherent answers, and are particularly key in cross-cultural applications where the researcher’s own folk psychology doesn’t match the folk psychology of half of the research sample.

By separating figure from ground, the implication is not that the “ground” is uninteresting. As Krueger and Funder (2004) note, models of cognition should encompass both rational and irrational judgments, and focusing only on one or the other necessarily results in an incomplete theory. In cross-cultural research, this should be extended to say that theories must account both for cultural differences and similarities in order to be considered comprehensive.

To build comprehensive theories, more empirical research will of course be necessary.
The most emphatic recommendation to be made is that normative assumptions be checked meticulously in this process. Cross-cultural researchers are accustomed to translating and backtranslating as a way to ensure that research stimuli are identical across cultures. However, particularly with social judgment tasks, not all important information is contained in the words that participants read or hear. Instead, the normative impact of the information people take in depends heavily on what is already in their heads, and this is something that backtranslation procedures are not designed to handle. To conduct social judgment research responsibly, researchers should build outward, starting with the words and stimuli, and then moving to the judgments that people make under the most ideal and controlled conditions—with equivalent goals, ample time, and no differences in what is noticed or remembered—before beginning to tinker with the judgment process itself.

**4.4.3 Future Research Directions**

The core of this research should continue to examine how to elicit the beliefs needed to assess coherence, and then to test coherence with a broader variety of paradigms and scenarios. Only then should work begin to assess how process manipulations (differentially) affect these judgments. Work that manipulates inferential goals and cognitive busyness should be particularly careful to determine whether, as a result of the manipulation, different information is perceived, the same information is perceived differently, or the information is processed differently.

In addition to coherence, a number of topics were touched upon in the review section of this paper that deserve dedicated followup. One ripe area of study is to examine how explanation drives attribution, and in particular to look at how social norms about modesty, face, etc., might affect the explanatory (and hence attribution) goals that people adopt. Such research would go a long way toward moving away from culture as a set of values and toward culture as procedural and declarative knowledge (Chiu & Hong, 2007) and as a social process.

Along the lines of knowledge, the normative identity used here can be used to guide how concepts like lay dispositionism or lay situationism can be operationalized. The finding that the Western participants saw slightly higher relations between traits and circumstances was also interesting, and should be followed up upon for its potential to help resolve the seeming contradiction between understanding the power of the situation yet believing in the primacy of individual. Work could also be done to determine how people assess the background beliefs used in this research, and in particular whether the explanations people hear might drive the judgments that they make, particularly as regards frequency of and correlations between causes (cf. Fiedler, 2007).
4.5 Conclusion: So, who’s right and who’s wrong?

This paper began by asking a provocative question about cultural differences in attribution: if cultures disagree, then who’s right and who’s wrong? Predictably, the answer is “it depends”. Cultural differences in explanation may be expedient responses to social demands, and hence be pragmatically correct across cultures. Laboratory paradigms indicate that when judged by correspondence between people’s answers and the ground truth, people in both cultures are sometimes wrong. The study conducted for this paper showed that when it comes to coherence under ideal conditions, there were no strong cultural differences in how well dispositional attributions in the presence of situational causes were made. For dispositional attributions with situational causes uncertain, East Asians were more tempered, but also less normative. And of course, these results demand replication and extension, and even then they may say nothing about how often people in different cultures are wrong or right outside of the laboratory (McKenzie, 2003).

Unsatisfying though the answer “it depends” may be, it is satisfying to be asking the question to begin with rather than diplomatically avoiding it. The aim of this work has been to elevate reasoning about what’s correct beyond the level of intuitions and syllogisms, but it is only an incomplete first step. So begin even the longest of journeys.
Chapter 5

Conclusions

This work demonstrated the use of a Bayesian identity for determining the coherence of people’s attributions regarding one cause for an effect, given that another cause for the effect was also present. Empirically, it was revealed that people’s attributions are largely coherent, and that it cannot be strongly concluded that there are major cross-cultural differences in the degree of coherence. However, more work will need to be done to improve the research instruments used, particularly as they begin to be applied in settings that are either more true to life, or that attempt to short-circuit the attribution process.

The model was also adapted from dichotomous judgments to continuous relationships between variables, and then applied with the attitude attribution paradigm. The model was able to postdict several classical findings that had previously been considered biased, and model-predicted relationships were also discovered in newly-collected data. These results suggest that people’s attitude attributions may themselves be coherent, though more work must be done to encompass more of the existing literature, and to test the model’s empirical predictions in a more deliberate manner.

Perhaps of greater importance than the empirical contributions of this work are its theoretical ramifications. First, the identity is a useful tool to researchers who either wish to verify that a manipulation in their study will have the intended normative implications, or who wish to understand why participants in a previous study did not respond to a manipulation as was intended. It could be that upon further examination, many previous research findings that have been treated as evidence of flaws in people’s causal reasoning will be vindicated as coherent. Additionally, the continuous application of the identity holds great promise for the explicit specification of different forms of cause and effect relationships. Moving forward, this approach holds particular promise with the attribution of morality and ability, which both follow Reeder and Brewer’s (1979) hierarchically restrictive schema.

For all the promise in this work, it should not be forgotten that it must eventually be reconciled with the remainder of work in the attribution field. At present, the model elides issues such as the attribution of intentionality (Malle, 1999) or motives (Reeder, 2009), and neglects distinctions like that between causes and conditions (Einhorn & Hogarth, 1986; Cheng & Novick, 1991). While these distinctions are mathematically irrelevant to the
identity, some of the anomalous study results in this work may be explainable under more semantically rich theories. At the mathematical level, the distinction between causal ascription and causal induction requires further clarification. While it is probably true that the two processes require different normative standards, it is also probably true that they are difficult to actually separate in the laboratory, let alone in everyday life. This broader context must be kept in mind when interpreting the results obtained using this work’s Bayesian identity.

Limitations notwithstanding, this work has the potential to spur major advances in the understanding of attribution in social and personality psychology. The use of explicit, quantitative theories should be seen as a key step in improving the status of psychology as a cumulative science. This work is but a simple example of the quantitative, but hopefully one that will prove accessible enough to the broader field to eventually help popularize quantitative approaches in general.
References


& N. Harvey (Eds.), *Blackwell handbook of judgment and decision making*. Malden, MA: Blackwell.


Miller, A. G., & Rorer, L. G. (1982). Toward an understanding of the fundamental at-


Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attri-


Appendix A

Comparison with Morris and Larrick, 1995

This appendix compares this paper’s model with Morris and Larrick’s (1995) model, which is stratified by the sufficiency of the causes, their independence, and the number of other possible causes of the effect. A cause $C$ is sufficient if and only if $P(E | C) = 1$. Two causes $C_1$ and $C_2$ are independent if and only if $P(C_1 | C_2) = P(C_1)$. These causes are sole causes if and only if $P(E | \overline{C_1}, \overline{C_2}) = 0$.

Morris and Larrick are primarily interested in when discounting is warranted, which is to say when $P(U | E, C) < P(U | E)$. Stated in terms of the present standard, this is equivalent to:

$$P(U) \cdot \frac{P(C | U)}{P(C)} \cdot \frac{P(E | C, U)}{P(E | C)} < P(U) \cdot \frac{P(E | U)}{P(E)}$$

As can be seen, the inequality does not depend on $P(U)$. This section reconsiders Morris and Larrick’s findings as summarized in Table 7 on page 347 of their paper. In all of the following, assume that $P(E) < 1$.

A.1 Sole Causes

Morris and Larrick find that discounting is always implied for sole causes, i.e., when $P(E | \overline{C}, \overline{U}) = 0$. Translating the conditional probabilities into the ratios of conjunctions, the criterion becomes:

$$\frac{P(C, U)}{P(C)P(U)} \cdot \frac{P(E, C, U)P(C)}{P(C)P(E, C)} < \frac{P(E, U)}{P(U)P(E)}$$

Consider the Venn diagram in Figure A.1. The above inequality is equivalent to:

$$(a + b + d + e)(e) < (b + e)(d + e)$$
Figure A.1: Venn diagram to illustrate probabilities in Morris and Larrick (1995).

\[ ae + be + de + e^2 < bd + be + de + e^2 \]
\[ ae < bd \]

By the assumption of sole causes, \( P(E \mid \overline{C}, \overline{U}) = 0 \), and so \( P(E, \overline{C}, \overline{U}) = a = 0 \). Therefore, discounting is implied so long as \( b, d > 0 \), which is to say \( P(E, C, \overline{U}) > 0 \) and \( P(E, U, \overline{C}) > 0 \). These conditions will hold unless \( C \) only causes \( E \) with \( U \) also present (or the reverse). There are two cases where this might occur. First, both causes might be jointly necessary, in which case observing one most certainly should not reduce confidence that the other was present. Second, one “cause” (say, \( C \)) might appear to be a cause of \( E \) due to its association with \( U \), which could happen if \( C \) is in fact a side effect of \( U \) or \( E \). Clearly observing a side-effect of a cause or its effect shouldn’t reduce confidence that the cause was present. However, aside from these two boundary cases, discounting is always rational for sole causes.

### A.2 Non-Sole Causes

When \( C \) and \( U \) are not sole causes, the conclusions are more complicated. The exception is when sufficiency and independence are assumed, in which case \( P(E \mid C) = P(E \mid U) = P(E \mid C, U) = 1 \). In this case the criterion reduces to \( P(E) < 1 \), which is true by assumption.

Suppose that sufficiency is assumed, but that independence is not. In this case the criterion becomes \( P(C \mid U) / P(C) < 1 / P(E) \). If the correlation is less than zero, \( P(C \mid U) < P(C) \), which given that \( P(E) < 1 \) shows that the criterion holds. If the correlation is more than zero, \( P(E) \) must be small enough that its reciprocal exceeds \( P(C \mid U) / P(C) \). Therefore, Morris and Larrick’s statement that discounting is implied “[w]henever \( \phi \leq 0 \), or when \( P(E) \) is low relative to \( \phi \)” (p. 347) is correct, albeit not particularly illuminating.
Figure A.2: Contradiction for non-sufficient, mutually independent causes. Shaded gray circles represent the cause $X$.

Now suppose that sufficiency isn’t assumed, but that independence is. Morris and Larrick state that discounting is normative, provided that all additional causes are also independent. This statement turns out not to be correct, as shown by the counterexample in Figure A.2. Though an effect whose causes are all independent is a mythical beast, it is not correct to say that discounting is always implied for such a creature.

Finally, suppose that the causes are not sufficient and not independent. This leaves the criterion as it was originally stated. Morris and Larrick say that discounting is implied “[w]hen $P(E)$ is low, and $P(E \mid A)$ is high, relative to $\phi$” (in this paper, $A$ has been written as $U$). Thus, their criteria refer to co-occurance of $C$ and $U$ and the causal strength of $U$, but miss the role of the possible causal interaction between $C$ and $U$. 
Appendix B

Culture Study Materials

B.1 Scenario Introduction

This experiment concerns how people determine cause and effect relationships. The questions in the experiment concern the people in a hypothetical college class. The hypothetical class is part of the general requirements that all students at the college have to take. The goal of the class is to teach college-level writing, and also oral presentation skills. As part of the work for this class, each student has to give an in-class presentation in front of members of their discussion section. There are about 20 people in each section of the class, but the overall class is much larger.

B.2 Circumstances

1. traffic is bad on the way to the presentation
2. the person is busy making last-minute preparations
3. the person has to take an urgent phone call
4. the person spends a long time getting dressed
5. the person’s bus was running late
6. the person misread their clock
7. the person hit the snooze button when their alarm went off
8. the person got very lost on the way there
9. the person did not set their alarm
10. the person was in a minor car accident
11. the class the person had before the presentation ran long
12. the person’s computer’s printer breaks down
13. the person cannot find the slides for the presentation
14. the person was talking with a friend
15. the person forgot what time the presentation was
16. the person was feeling too nervous to go on
17. the person had to do something more important
18. there was a mishap on the way to class
19. the person forgot something at home and had to go back to get it
20. the person’s roommate needs help with something
21. the person has just started to get sick
22. the person doesn’t care about the class or presentation

B.3 Probability Judgment Phrasing

\[ P(T | E) \] What percentage of the people in this class who are late for their presentations do you suppose are irresponsible or absent-minded?

\[ P(T) \] What percentage of people in the overall class do you suppose are irresponsible or absent-minded?

\[ P(T | E, C) \] Suppose that you knew that everyone who was late for their presentation also faced the listed circumstance that day. What percentage of these people do you suppose are irresponsible or absent-minded?

\[ P(C) \] What percentage of people in the overall class do you suppose faced the listed circumstance on the day of their presentation?

\[ P(C | T) \] What percentage of the irresponsible or absent-minded members of this class do you suppose faced the listed circumstance the day of their presentation?

\[ P(E) \] What percentage of people in the overall class do you think was late for their presentation?
Section B.3. Probability Judgment Phrasing

\[ P(E \mid T) \] What percentage of the irresponsible or absent-minded members of the class do you think was late for their presentation? This group includes some people who faced other circumstances, as well as some who didn’t.

\[ P(E \mid C) \] Of all of the people in the class who faced the listed circumstance, what percentage do you suppose was late to their presentation?

\[ P(E \mid C, T) \] What percentage of people who both (a) are irresponsible or absent-minded, and (b) faced the given circumstance do you suppose was late to their presentation?