Delay and Knowledge Mediation in Human Causal Reasoning

Thesis submitted in February 2002

For the degree of Doctor of Philosophy

by

Marc Buehner

Department of Psychology, The University of Sheffield

Acknowledgement

I would like to thank my supervisor Jon May for his generous support over the last three years, and for giving me the freedom to pursue the questions that interested me most. An especially warm thank you goes to my partner Merideth Gattis, for discussing experimental paradigms with me, commenting on related papers, and for encouraging me in those times when I felt my work was worthless. I am also grateful to Patricia Cheng, for insightful comments on Knowledge Mediation, and for assuring me that I have an entire career in front of me, when she pointed out that the work presented in this thesis is only the beginning of a much larger project. I would also like to express my gratitude to York Hagmayer for engaging in a very stimulating discussion some years ago, which led to a fruitful collaboration in a related project, that inspired my thinking on the work presented here. Thanks also to Andres Haye, Sabine Pahl, Mat White, and Tom Webb, for interesting discussions on related topics, and for reading and commenting related manuscripts, to Rob Goldstone for providing me with references on categorization, and to Michael Waldmann for providing me with hints on utility and causation. Several anonymous reviewers from the journals *Psychological Science* and *Thinking & Reasoning* have provided helpful comments on the work presented in chapters 2, 4, and 6.

This work was funded by the EU T&MR network TACIT. Parts of chapter 1 are taken from my chapter "Inducing Causation: Covariation Assessment and the Assumption of Causal Power" which appeared in M. May & U. Oestermeier (Eds.) (2001). *Interdisciplinary Perspectives on Causation.* Norderstedt, Germany:Libri. Parts from chapters 2, 4, and 6 are currently being revised for publication in *Thinking & Reasoning*, and the work presented in chapter 5 is currently being considered for publication in the *Quarterly Journal of Experimental Psychology*. An overview of the work presented in this thesis was presented at the 2001 International Conference on Causal Inference, Snowbird, Utah.

Table of Contents

Abstract

Contemporary theories of causal induction have focussed largely on the question of how evidence in the form of covariations between causes and effects is used to compute measures of causal strength. A very important precursor enabling such computations is that the reasoner notices that a cause and effect have co-occurred. Standard laboratory experiments have usually bypassed this problem by presenting participants directly with covariational information. As a result, relatively little is known about how humans identify causal relations in real time. What evidence exists, however, paints a rather unflattering picture of human causal induction and converges to the conclusion that humans cannot identify causal relations if cause and effect are separated by more than a few seconds. Associative learning theory has interpreted these findings to indicate that temporal contiguity is essential to causal inference. I argue instead that contiguity is not essential, but that the influence of time in causal inference is crucially dependent on people's beliefs and expectations about the timeframe of the causal relation in question.

First I demonstrate that humans are capable of dissociating temporal contiguity from causal strength; more specifically, they can learn that a given event exerts a stronger causal influence when it is temporally separated from the effect than when it is contiguous with it. Then I re-investigate a paradigm commonly used to study the effects of delay on human causal induction. My experiments employed one crucial additional manipulation regarding participants' awareness of potential delays. This manipulation was sufficient to reduce the detrimental effects of delay. Three other experiments employed a similar strategy, but relied on implicit instructions about the timeframe of the causal relation in question. Overall, results support the hypothesis that knowledge mediates the timeframe of covariation assessment in human causal induction. Implications for associative learning and causal power theories are discussed

1. From Covariation to Causation

1.1. Hume and His Heritage

The world as given to us is a flux of sensations. Yet humans and other intelligent species have evolved to partially predict and even control their environment. How do people structure the world of sensations? What enables them to manipulate their surroundings based on their forecasts? In other words, how do people learn about the causes of events? David Hume (1777/1902) pointed out the fundamental problem for the acquisition of causal structures: the human sensory system is not receptive to causality per se: we cannot "see", "hear", "feel", "taste", or "smell" causal relations. Since causal relations are not explicitly represented in the input, the Humean argument goes, causation must be inferred by some process of induction based on evidence available to our senses. Some researchers have argued that certain physical events give rise to direct causal perception (e.g., the launching effect of collision events, see Michotte, 1946/1963), but a critical analysis of the argument (Cheng, 1993) revealed that simple perceptual mechanisms alone cannot explain people's complex causal attribution patterns. Instead, most researchers in the area now agree that humans use covariational information about the presence and absence of candidate causes and effects to infer causal relations. For a binary candidate cause *c* and effect *e*, the former perceived as occurring before the latter, covariation information can be represented in a 2x2 *contingency table* (Figure 1-1) where cell *a* contains the frequency of the joint presence of a candidate cause and the effect, cell *b* the frequency of events in which the candidate is present but the effect absent, cell *c* the frequency of events in which the candidate is absent but the effect present, and cell *d* the frequency of events in which both candidate and effect are absent.

Figure 1-1. A 2x2 contingency table and proposed strategies for contingency judgments (adapted from Shimazaki, Tsuda, & Imada, 1991)1].

Candidate cause c	Effect e: present	Effect e: absent
present	a	b
absent	c	d

Strategy	Calculation
Cell-a	Compare cell a with other three cells
ΔF	Compare cell a with cell c
ΔD	Compare (a+d) with (b+c)
ΔP	Compare (a/a+b) with c/(c+d)

Philosophers, social psychologists, cognitive psychologists, learning theorists, and computer scientists have vigorously debated how causality can be inferred from a contingency table. Several simple decision rules (see Figure 1-1) were proposed to suggest how causation can be inferred from observable frequencies. A few decades ago the ΔP rule was identified as the normative measure of causality extracted from contingency information (e.g. see Jenkins & Ward, 1965). ΔP is often referred to as the *contingency* between *e* and *c* and can also be expressed in terms of conditional probabilities:

Equation 1. $$\Delta P = P(e|c) - P(e|\neg c)$$

with P(e|c) being the probability of *e* given the presence of *c*, and P(e|¬c) the probability of *e* given the absence of *c*. If ΔP is noticeably positive, *c* would be inferred to produce *e* (be a *generative* cause of *e)*, and if ΔP is noticeably negative, *c* would be inferred to inhibit *e* (be a *preventive* cause of e). If ΔP is neither positive nor negative, *c* does not influence *e* and there is no causal relationship between the two. More recent attempts to describe human causal induction (e.g. Anderson & Sheu, 1995) proposed to add weights into the ΔP rule. The rationale behind this endeavour is that humans do not appear to be equally sensitive to the four cells of a contingency table. Such a modified ΔP rule

Equation 2. $$w_0 + w_1 P(e|c) - w_2 P(e|\neg c)$$

or a weighted linear model (Schustack & Sternberg, 1981; as cited in Anderson & Sheu, 1995)

Equation 3. $$w_0 + w_1 a + w_2 b + w_3 c + w_4 d.$$

of course allow the weights to be set post-hoc and thus give greater degrees of freedom to provide a better fit to actual causal judgment data obtained in experiments than the standard ΔP rule. Typically, it has been claimed that the *a* and *b* cells of a contingency table as displayed in Figure 1-1 are deemed to be most informative for the discovery of a causal relation (e.g. Anderson & Sheu, 1995), and the weights for such parameterised models were adjusted

accordingly. A more recent analysis of the argument (Over & Green, 2001) has shown, however, that this is only the case for scenarios where both the cause and the effect are rare. In situations where causes and effects are very common, the *c* and *d* cells carry the crucial information, and weights would need to be ordered differently.

1.2. The Associationist Approach

Another prominent account of human causal induction also claims to address the problem posed by Hume (1739/1888) and likewise takes information about temporal order and the presence and absence of candidate causes as its starting point. But the way in which this information is parsed so as to come to an understanding of cause is radically different from the previously outlined approaches, in that it does not entail any rules which "read off" information from an episodic memory with a figurative contingency table. Rather, "causal judgment is seen as reflecting no more than the strength of the relevant association between the mental representations of the cause and effect, with the principles governing such attributions being those of associative learning." (Shanks & Dickinson, 1987 p. 230). In an associationist framework, mental representations of causal strength are not the product of a retrospective reasoning process. Instead, causal strength is accounted for in the continually updated association between candidate causes and effects. The principle underlying all associationist theories is strikingly simple: if on a given occasion (learning trial) a cue and an outcome co-occur together, the association between them will increase, if the cue occurs by itself without the outcome, or if the outcome happens on its own without the cue's presence, the association will decrease. The most prominent associative learning model has been the Rescorla-Wagner model (RWM) (1972), originally proposed to explain the processes underlying Pavlovian conditioning:

Equation 4. $$\Delta V_{CS} = \alpha_{CS} \cdot \beta_{US} \cdot (\lambda - \Sigma V)$$

with ΔV_{CS} being the change in associative strength between a conditioned stimulus CS (the cue, e.g. a flash of light) and an unconditioned stimulus US (the outcome, e.g. a foot shock) on a given learning trial and α_{CS} and β_{US} as learning parameters that respectively represent the salience of the CS (e.g. the light's brightness) and the US (e.g. the shock's intensity). λ is the outcome of a given trial and is usually 1 if the US is present and 0 otherwise. Finally, ΣV is the sum of all associative strengths of all present CSs and is therefore interpreted as the "expected outcome" of a given trial. In the RWM and related associationist theories learning consists of reducing the discrepancy between the expected and the actual outcome. When there is a difference between λ and ΣV on a given learning trial, the associative weights of all cues present on that trial are updated according to the rule specified in Equation 4. Eventually, after many trials, the discrepancy between the expected outcome ΣV and the actual outcome λ will approximate zero – learning has reached asymptote. In other words, the cues can fully predict (or explain) the outcome. Attempting to use the RWM to account for human causal induction of course reduces reasoning to associative learning: the candidate cause *c* is mapped onto the CS, the effect *e* onto the US, and the causal power of *c* is mapped onto *c*'s associative strength.

Researchers applying the RWM typically distinguish between two versions of the model, depending on assumptions about the learning parameter β. If β is assumed to be constant between trials on which the US is respectively present and absent ($\beta_{US} = \beta_{\overline{US}}$), the model is referred to as the *restricted* RWM. If one allows the value of β to vary between these trials, the *unrestricted* RWM applies (Miller, Barnet, & Grahame, 1995; Lober & Shanks, 2000 introduced the terms restricted vs. unrestricted RWM). Usually the presence of the US is assumed to be more salient than its absence, because the vast majority of empirical results requires the parameter ordering that

follows from this assumption ($\beta_{US} > \beta_{\overline{US}}$) to fit the data (Shanks, 1991; cf. Wagner, Logan, Haberlandt, & Price, 1968; Rescorla & Wagner, 1972). Chapman and Robbins (1990) demonstrated that the restricted RWM converges in its asymptotic predictions to a simple ΔP rule for the special case in which there is only one stimulus cue. The unrestricted RWM, however, does not compute ΔP. Wasserman, Chatlosh, Elek, & Baker (1993) offered a formula that allows one to calculate asymptotic predictions of the RWM with unequal values of β.

Equation 5. $$V_{asymp} = \frac{\beta_{US} a}{\beta_{US} a + \beta_{\overline{US}} b} - \frac{\beta_{US} c}{\beta_{US} c + \beta_{\overline{US}} d}$$

where a, b, c, and d refer to the four cells of the contingency table displayed in Figure 1-1. Regardless of the exact values one chooses the parameters to take, with $\beta_{US} > \beta_{\overline{US}}$, the (absolute) magnitudes of the judged causal strengths will be smaller as P(e|¬c) = $c/(c+d)$, the *base rate* of *e*, increases for any fixed positive or negative ΔP. But if $\beta_{US} < \beta_{\overline{US}}$, contrary to the typical assumption, the RWM would predict the opposite trend: the (absolute) magnitudes of the judged causal strengths should be larger as P(e|¬c) increases for any fixed positive or negative ΔP. Regardless of assumptions about β, the RWM predicts that P(e|¬c) influences the absolute causal strengths for candidates with equal positive ΔP in the same direction as those with equal negative ΔP. Section 1.5 will review some empirical results relevant to these predictions.

It should be pointed out here that the focus of research on human associative learning (i.e. attempting to explain human causal induction with associative learning theory) has undergone a change in the last two decades. Early studies (e.g. Shanks, 1985, 1987, 1991) were concerned with the actual learning process and investigated acquisition functions. Contemporary

investigations (e.g. Shanks & Lopez, 1996; Shanks, Lopez, Darby, & Dickinson, 1996; Lober & Shanks, 2000; Perales & Shanks, 2000) seldom do this anymore and focus on asymptotic predictions and final judgments instead.

One of the motivations that led to the development of the RWM was to account for the well documented phenomenon of blocking in animal learning (Kamin, 1969; as cited in Rescorla & Wagner, 1972): When one cue is established as a strong predictor for an outcome, subsequent exposure to a compound of the perfect predictor and a novel cue produces very little conditioning to the new element. In terms of the RWM the outcome is then already fully explained by the established predictor and updating of any associative weights, including the one from the novel stimulus, has become unnecessary.

In the 1980s associationism rose in its popularity as an explanation of human causal learning. Attracted by the similarity between *cues* and *outcomes* and *causes* and *effects*, researchers aimed to explain human causal learning through associative principles. The replication of a blocking effect in human causal learning led Dickinson and his associates (see Shanks & Dickinson, 1987; and Shanks, 1993b for an overview) to propose that causal learning can be reduced to associative learning. A substantial amount of subsequent experimental evidence supported this general proposal, and in particular the RWM (Shanks, 1985, 1987, 1993a; Shanks & Dickinson, 1987; Wasserman et al., 1993).

However, the glory of the reductionist approach proposed by the associationists did not live up to the expectations. Sparked by the growing interdisciplinary interest in theories of causation, and by vigorous debate between associationists and other theorists (Melz, Cheng, Holyoak, & Waldmann, 1993; Shanks, 1991; Shanks & Lopez, 1996; Waldmann & Holyoak, 1992, 1997), the "Psychology of Learning and Motivation" series saw the publication of a volume on causal learning in 1996 (Shanks, Holyoak, & Medin, 1996). The opening chapter of this volume, written by the

prominent associationist researchers A.G. Baker, Robin Murphy, and Frédéric Vallée-Tourrangeau, sums up one of the main shortcomings of an associative perspective on causal reasoning: "... experience is stored as a small number of associative strengths. ... information about past events is lost in the computation. In other words, these models do not have episodic memory." (Baker, Murphy, & Vallée-Tourangeau, 1996 p.1). It seems hard to justify how any insightful reasoning process, including causal inference, could be come by without the assumption of an episodic memory.

Another important prerequisite for the intelligent judgment of cause is the insight into causal directionality, as Waldmann, in the same volume (1996), sums up concisely:

> One of the most important examples of abstract causal knowledge that may affect the processing of the learning input is knowledge about causal directionality. We know that the causal arrow is directed from causes to their effects and not the other way around. This fundamental property of causal relations is of the utmost pragmatic importance as it provides the basis for our abilities to reach goals. Effects can be achieved by manipulating causes, but causes cannot be accomplished by manipulating their effects. Thus it is extremely important to be able to distinguish between causes and effects. (Waldmann, 1996, p.52)

An associationist model is incapable of making the crucial distinction between cause and effect. Its scope only entails cues and outcomes. The standard adoption of associative models to causal learning maps *cues* to *causes* and *effects* to *outcomes*. By the same rationale a physician who reasons that the presence of certain bacteria in the stomach is the cause for an ulcer does so not by virtue of a learning mechanism specifically devoted to discover causal structures. Instead, the task is hypothesized to be performed by a simple associative learning algorithm, equivalent to mechanisms assumed to drive a rat's behaviour in a Skinner box.

Such associative learning algorithms fail to encompass causal directionality. However, human causal reasoners are capable of eventually realizing that the ulcer is actually the *cause* of bacterial presence, rather than being produced by the bacteria, if the data structure supports this inference (Hagmayer & Waldmann, 2001). In other words, humans can distinguish between predictive reasoning (bacteria causes ulcer) and diagnostic reasoning (bacteria are an effect of ulcer), even when they are first presented with the effect (i.e. a symptom) and subsequently with the cause, as is usually the case in medical decision making.

Ironically the blocking paradigm which was the boon for associationism in the 1980s also proved to be its bane in the 1990s. In an associative framework cues compete for associative strength, just like causes do in causal reasoning. However, by the fundamental fact of causal asymmetry, effects collaborate in explanatory strength rather than compete against each other. If, for example, the presence of a specific kind of bacteria is established as an effect of an ulcer, and subsequent learning experience reveals that tissue swelling is also a result of ulcer, acquisition of diagnostic explanatory strength in the second effect will not be significantly impaired by the already established explanatory strength of the first effect. Michael Waldmann demonstrated this distinction between predictive and diagnostic causal learning very impressively using a standard blocking paradigm (Waldmann & Holyoak, 1992, 1997; Waldmann, 2000). The same input (cue) was introduced as cause in one condition, and as an effect in the other; consequently the outcome was labeled "effect" in the first and "cause" in the latter group. After having learnt the perfect pairing between cue and outcome, participants in Waldmann's experiments were presented with a second cue that was always paired with the already established cue; this compound of cues was also always followed by the outcome. The results indicated uniformly that people are sensitive to causal asymmetry. When participants thought that the cues were causes, the second cue competed with the first cue, which was

already established as a perfect predictor. Since participants had no information about the effectiveness of the second cue on its own (it was always paired with the established cue), ratings of association between the cue and the outcome were weaker in the second, blocked cue than in the first. However, when participants believed that the cues were effects, and the outcome a common cause of the effects, the cues did not compete in associative strength and participants readily gave equally high ratings to both the established and the "blocked" cue. The RWM of course, failing to capture the directionality of the causal arrow, erroneously predicts blocking between effects in a diagnostic tasks, just as it (correctly) predicts blocking between causes in a predictive task.

1.3. Discovering Causes from Covariation Alone?- The Power View

Apart from all the problems resulting from its reductionist perspective, the associationist approach suffers from a problem that cripples all purely covariation based models: as any introductory statistics textbook admonishes, covariation does not necessarily imply causation. Many events follow each other regularly, yet we are unwilling to infer a causal relation between them. The famous rooster on the farm crows each morning just before sunrise (and the sun does not rise on other times of day when the rooster does not crow), yet we do not infer that the rooster's crowing causes the sun to rise. One prominent answer to the problem of how to decide when a covariation warrants causal inference goes back to Kant (1781/1965) and is called the *power* view. Advocates of this power view claim that prior knowledge (often referred to as *knowledge of mechanism*) about a plausible connection between cause and effect helps humans to interpret covariation information. Specifically, power theorists would argue that unless one knows of or perceives a causal link or mechanism between a candidate cause and an effect,

one cannot infer a causal relation between them (cf. Ahn, Kalish, Medin, & Gelman, 1995; Bullock, Gelman, & Baillargeon, 1982; Michotte, 1946/1963). One event thus causes another by virtue of exerting its *causal power*, by transmitting energy. A prime example is the understanding that butter melts in a heated pan: the stove emits heat (energy), raising the temperature of the pan, which in turn makes the butter melt. In a power framework, effects do not simply follow their causes, rather they are produced, or generated by them. The rooster on the farm may exhibit statistical regularity in its crowing just before sunrise (just like heating precedes melting), but it lacks the critical mechanism or power present in the melting butter example.

The power view has intuitive appeal, but suffers from circularity: according to this view one cannot infer that a relation is causal unless one knows of a mechanism which causally explains the relation, in other words, to identify a relation as causal, one needs to first know that it is causal. Prior knowledge certainly guides the interpretation of covariation input towards causality, as is evident from Waldmann's studies on predictive and diagnostic reasoning. Nonetheless, such knowledge does not come out of the blue and unless one is willing to claim that it is innate, it must be learnt somehow. The power view thus pushes the problem pointed out by Hume one step back, but ultimately fails to solve it. Also, it only makes predictions about the *circumstances under which* a covariation is interpreted as causal but not *how so*, and therefore lacks the computational description of causality present in the contingency and associationist models.

1.4. The power PC theory

A third alternative to causal induction put forward by Cheng (1997) adopts the Humean notion that causality must be computed from observable evidence (i.e. covariation), but also entails a Kantian framework in that it proposes that humans innately postulate that there exist causes in the world

that have the power to produce events and causes that have the power to prevent events and that the goal of causal induction is to infer these powers from observable evidence. A primary motivation for Cheng's power PC theory was to account for previously unexplained empirical phenomena of causal induction. One such phenomenon is the non equivalence of covariation and causation as, for example, in the problem of a "ceiling effect": when experimentally testing whether a manipulation *c* produces an effect *e* one cannot draw a valid conclusion regarding *c*'s power to produce *e* if *e* happens all the time, irrespective of *c*'s presence. Expressed probabilistically, a ceiling effect occurs when $P(e|c) = P(e|\neg c)=1.0$ All purely covariation based models, including all associationist models, cannot represent causal strength as a variable separate from its value. As a result, neither can represent a state of knowledge where causal strength is unknown (i.e. the variable has no value), but covariation has a definite value. In the ceiling effect example, $\Delta P=0$ and all models yield a definite value of causal strength (the contingency models and both the restricted and unrestricted RWM yield 0). Yet, it becomes immediately obvious that causal inference is not warranted in such a situation. Wu and Cheng (1999) demonstrated that humans are sensitive to ceiling effects and their analogs in preventive scenarios, when $P(e|c) =P(e|\neg c)=1.0$ participants in these situations reliably concluded that an outcome from such an experiment is uninformative; purely covariation based models were thus refuted.

Cheng's power PC theory (1997) can account for the uninformative nature of results embodying a ceiling effect by postulating causal power as a variable that is distinct from its value. It follows from the theory that when causes alternative to the candidate cause *c* both occur and influence *e* independently of *c*, and ΔP is non-negative, the generative power of *c* to produce *e* is

Equation 6. $$q_c = \frac{\Delta P_c}{1 - P(e \mid \neg c)}$$

For a non-positive ΔP, the same set of assumptions except that c may now potentially *prevent* instead of produce e yields

Equation 7. $$p_c = \frac{-\Delta P_c}{P(e \mid \neg c)}$$

The unwillingness of people to estimate causality in a ceiling effect situation logically follows from Equation 6.: it is undefined when P(e|¬c)=1.0 due to division of 0 by 0. The two equations illustrate some important empirical consequences predicted from the power PC theory. One implication is that scenarios involving equal levels of ΔP but different values for P(e|¬c) should yield different causal judgments. When Equation 6. applies, candidate causes in scenarios with equal nonnegative ΔPs should be judged to have increasingly *large* generative power as P(e|¬c) increases, but does not equal 1. In contrast, when Equation 7. applies, candidates with equal nonpositive ΔPs should be evaluated to have increasingly small preventive power as P(e|¬c) increases towards 1. When P(e|¬c)=0, Equation 7. is undefined: a reasoner cannot draw any conclusions about the power of c to prevent e, if e never happens in the first place in c's absence. When ΔP=0, both equations predict judgments of causality to be 0, as long as the denominator of the relevant equation is not also 0. The predictions derived from the power PC theory reflect descriptions of human causal judgments on an ordinal level only. Note that (a) P(e|¬c) influences the (absolute) magnitude of estimated causal strength in opposite directions for preventive and generative causes, and (b) the direction of these influences is not dependent on any parameter settings.

Let me illustrate the intuitive nature of these predictions with an example. Suppose you are a researcher trying to evaluate the effectiveness of several new headache relieving drugs. In one study you administer the new drug to a group of eight participants and a placebo to a control group of eight participants. You find that all eight participants in the control group complain about headaches, whereas only six of the eight participants in the experimental group suffer from headaches. $\Delta P = P(e|c) - P(e|\neg c) = .75 - 1.00 = -.25$. Assuming that all alternative causes of headaches are constant across the two groups, you will assume that if not for the drug, all eight participants in the drug group would have had headaches, just as in the control group. The drug, therefore, has a small preventive power, preventing headaches with a probability of .25. In another study, four out of eight participants in the control group and two of eight in the treatment group reported headaches; $\Delta P = P(e|c) - P(e|\neg c) = = .25 - .50 = -.25$. Again, assuming that alternative causes remain constant between groups, these causes would have produced headaches in four of the eight participants in the drug group, just as in the control group. The drug therefore prevents headaches in two of these four participants, yielding a probability of .50 to prevent headaches. Thus, although $\Delta P = -.25$ here, as in the preceding study, one would attribute a higher preventive power to the latter candidate.

1.5. Evaluation of the different Approaches

The above approaches differ fundamentally in what they postulate the reasoner assumes to remain invariant across contexts. According to the computational causal power approach (power PC theory, Cheng, 1997) that which stays constant are (unobservable) causal relations in the environment; according to purely covariational models (i.e. decision rules like ΔP) and associationist models (e.g. RWM), it is (observable) covariations between entities. It follows that the goal of a computational causal power approach is

to infer (distal) causal relations between entities in the world, whereas purely covariational and associationist models aim to gather features of the proximal stimulus.

Several predictions regarding causal judgments can distinguish the computational causal power approach from purely covariational and associationist accounts. Recall that the power PC theory predicts that, with an increasing base-rate of *e*, the absolute values of causal judgments of candidates with the same ΔP will increase when ΔP is positive, decrease when ΔP is negative, and remain at 0 when ΔP is 0. Table 1-1 illustrates these predictions.

Table 1-1. Predictions from power PC theory for contexts with non-positive and non-negative values of ΔP (preventive and generative contexts, respectively). Undefined values of causal power are indicated with a hyphen (-).

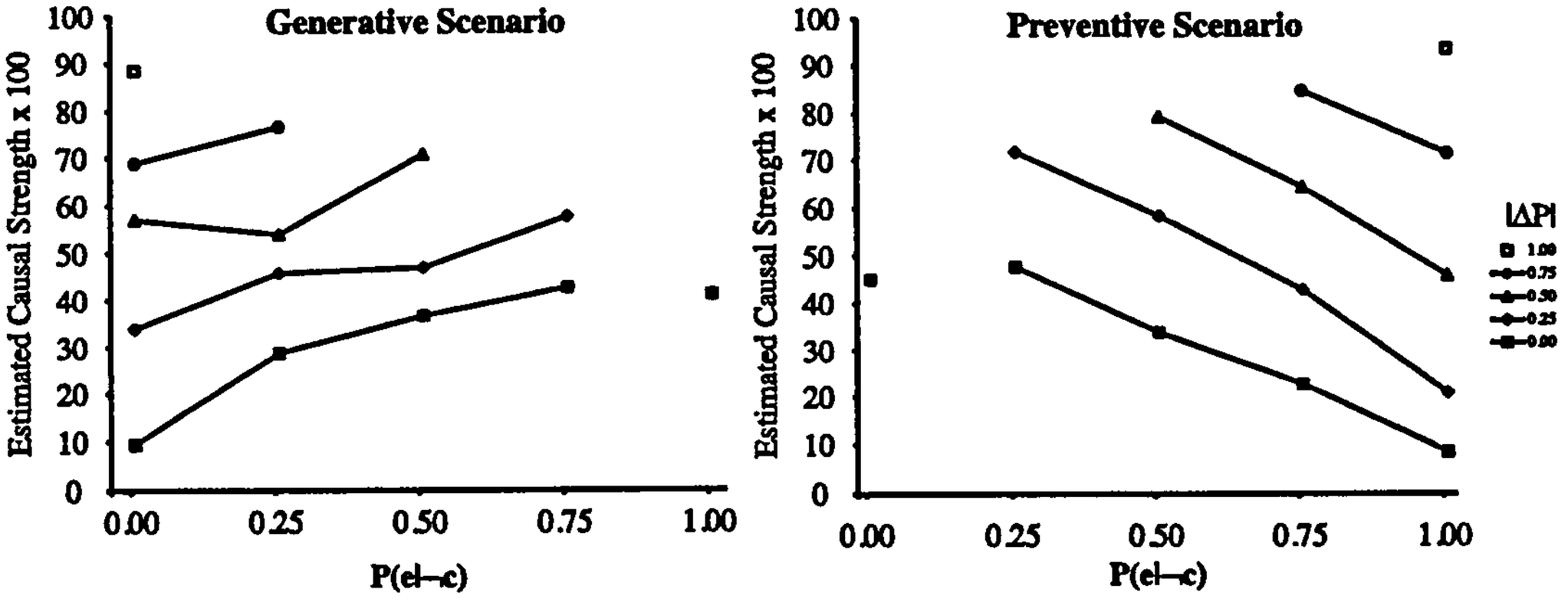
Generative Scenario				Preventive Scenario			
P(e|c)	P(e|~c)	ΔP	Causal Power	P(e|c)	P(e|~c)	ΔP	Causal Power
1.00	0.00	1.00	1.00	0.00	1.00	-1.00	1.00
1.00	0.25	0.75	1.00	0.00	0.75	-0.75	1.00
0.75	0.00	0.75	0.75	0.25	1.00	-0.75	0.75
1.00	0.50	0.50	1.00	0.00	0.50	-0.50	1.00
0.75	0.25	0.50	0.67	0.25	0.75	-0.50	0.67
0.50	0.00	0.50	0.50	0.50	1.00	-0.50	0.50
1.00	0.75	0.25	1.00	0.00	0.25	-0.25	1.00
0.75	0.50	0.25	0.50	0.25	0.50	-0.25	0.50
0.50	0.25	0.25	0.33	0.50	0.75	-0.25	0.33
0.25	0.00	0.25	0.25	0.75	1.00	-0.25	0.25
1.00	1.00	0.00	-	1.00	1.00	0.00	0.00
0.75	0.75	0.00	0.00	0.75	0.75	0.00	0.00
0.50	0.50	0.00	0.00	0.50	0.50	0.00	0.00
0.25	0.25	0.00	0.00	0.25	0.25	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	-

Also recall that, whereas the power PC theory is parameter-free, no consistent parameter settings of the RWM can predict such a pattern of results. Furthermore, the ΔP rule postulates ΔP as the sole indicator of causal strength and therefore cannot accommodate any kind of base-rate influence at all.

A first empirical test of these predictions (Buehner & Cheng, 1997) demonstrated a clear base-rate influence for situations with constant ΔP. More importantly, it also produced an interaction between the sign of ΔP (positive vs. negative) and the direction of the base-rate influence. Figure 1-2 illustrates Buehner & Cheng's results. Although these findings unequivocally refuted simple decision rules like ΔP and both variants of the RWM, the debate between computational causal power theorists and associationists is not yet resolved. Buehner & Cheng's data, while supporting the power PC theory's predictions regarding base-rate influence, actually also produced results problematic for the theory. Recall that causal judgments in conditions with ΔP=0 should be zero, irrespective of the value of the base-rate P(e|¬c). This was clearly not the case, as the bottom lines in both panels of Figure 1-2 appear to be significantly influenced by the base-rate. Also, there was a substantial influence of ΔP on the causal ratings of candidates with the same causal power (e.g. the topmost data points in both panels of Figure 1-2 all share identical causal power of 1, yet these conditions elicited different causal ratings, which appear to be influenced by changes in ΔP).

Lober and Shanks (2000) replicated Buehner & Cheng's results, and interpreted them as evidence against the power PC theory. Although they admitted that the interaction between the sign of ΔP and the direction of the base-rate influence effectively rejected the RWM (Perales & Shanks, 2000), they still defended an associationist position by pointing out that Pearce's (1987) model of stimulus generalization can in principle predict this interaction. It actually turns out that two of the more complex decision rules, the weighted ΔP model (Anderson & Sheu, 1995, Equation 2), and the weighted linear model (Schustack & Sternberg, 1981, Equation 3) can also be

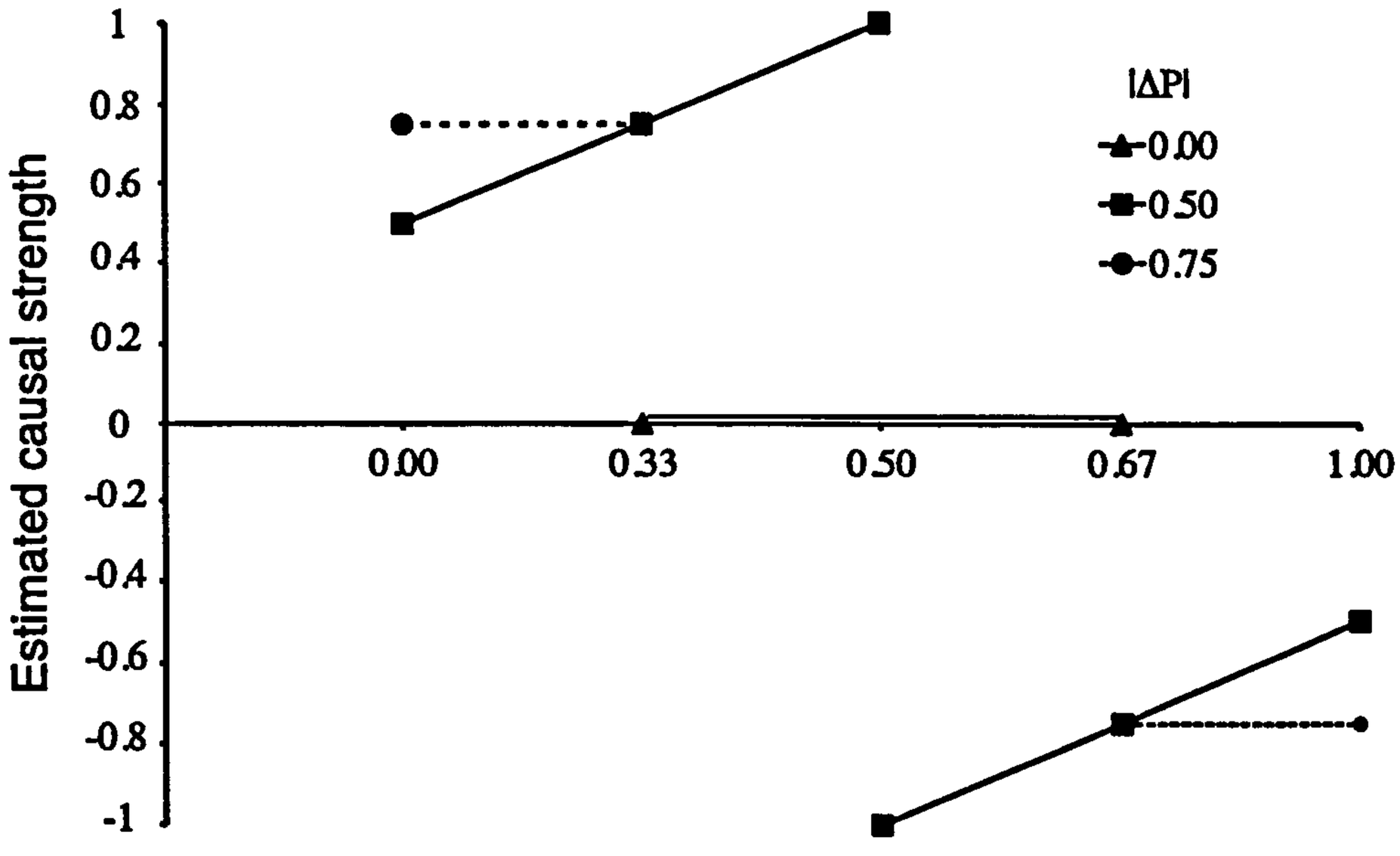
Figure 1-2. Results from Buehner & Cheng (1997). Lines connect conditions with identical values of ΔP.



fitted to the data patterns reported by Buehner & Cheng, Lober & Shanks, and Perales & Shanks.

However, Buehner and Cheng (Buehner, Cheng, & Clifford, 2001; Cheng & Buehner, 2000) have argued that the deviations from the predictions of the power PC theory are actually a result of ambiguities in the experimental materials and are not rooted in fundamental properties of the reasoning process. They presented a follow-up experiment with clearer materials (Buehner et al., 2001). While the results (see Figure 1-3) still preserved the main feature – the sign of ΔP and the base-rate interact to influence causal judgments – other factors no longer influenced causal judgments: causal power as predicted by the power PC theory proofed to be the sole determinant of judged causal strength. In particular, conditions with identical causal powers of .75 were not longer sensitive to variations in ΔP, and conditions with ΔP (and causal power) of 0 were no longer influenced by variations in base-rate.

Figure 1-3. Results from Buehner et al. (2001). Base-rate is represented on the abscissa. Solid lines connect conditions with identical levels of ΔP, dashed lines connect conditions with identical causal powers but varying levels of ΔP. Because judgments of preventive causal strength are plotted as negative values, an interaction between base-rate influence and the sign of ΔP is reflected by identical slopes for preventive and generative judgments.



The debate between computational causal power theorists and associationists is still going on, but the focus is shifting more and more towards technical and methodological details. Regardless of what future data on human causal judgment will show, one can already make some definitive statements about the mental leaps from covariation to causation. Cheng's (1997) analysis has made an enormous contribution to the field of causal reasoning research. Her derivations have shown that the ΔP rule, which was until then held as the normative measure of causal strength, in fact often offers

a poor estimate of causal strength. The causal power equations (Equation 6 and Equation 7) she proposed have replaced ΔP as the normative benchmark against which human reasoning ought to be compared. Even contemporary associationists (e.g. Lober & Shanks, 2000; Perales & Shanks, 2000) agree that Cheng's power PC theory is a normative theory of causal induction. The remaining disagreement is whether it also is a descriptive model of the judgment processes.

2. With (or Without?) Temporal Contiguity to Causation

The preceding chapter reviewed recent research aimed at shedding light onto the question of how humans infer causal relations from sensory information. Because this research has focused primarily on the question of what exactly is inferred from available evidence (i.e. how to obtain measures of causal strength from covariational information), a very important aspect of causality has largely been overlooked: the temporal relation between causes and effects. In order to derive causal knowledge from covariation, an organism must first be able to successfully identify (in real-time) that two events have co-occurred.

The work I have analyzed in chapter 1 always bypassed this problem by presenting participants with information that was already processed in some way. The vast majority of experiments either supplied participants with covariational information presented in contingency tables, or employed pre-defined, discrete, learning trials. Buehner and Cheng (1997), for instance, used fictitious lab reports in their experiments. Each lab report represented a particular rat and informed the participant whether or not this rat had been vaccinated against a certain virus, and whether or not this rat developed the disease associated with the virus in question. In other words, in these experiments it was always perfectly clear whether or not a candidate cause and effect co-occurred or not.

Unfortunately, real-life outside experimental psychology laboratories is not as neatly organized as the average causal reasoning experiment. Some events have immediate outcomes, others do not reveal their consequences until later. While it is an understood necessity that every cause must precede its effect to some extent (although these time spans sometimes may well be below our perceptual threshold, e.g. between flicking a switch and a bulb lighting up), cognitive psychologists know relatively little about people's sensitivity to

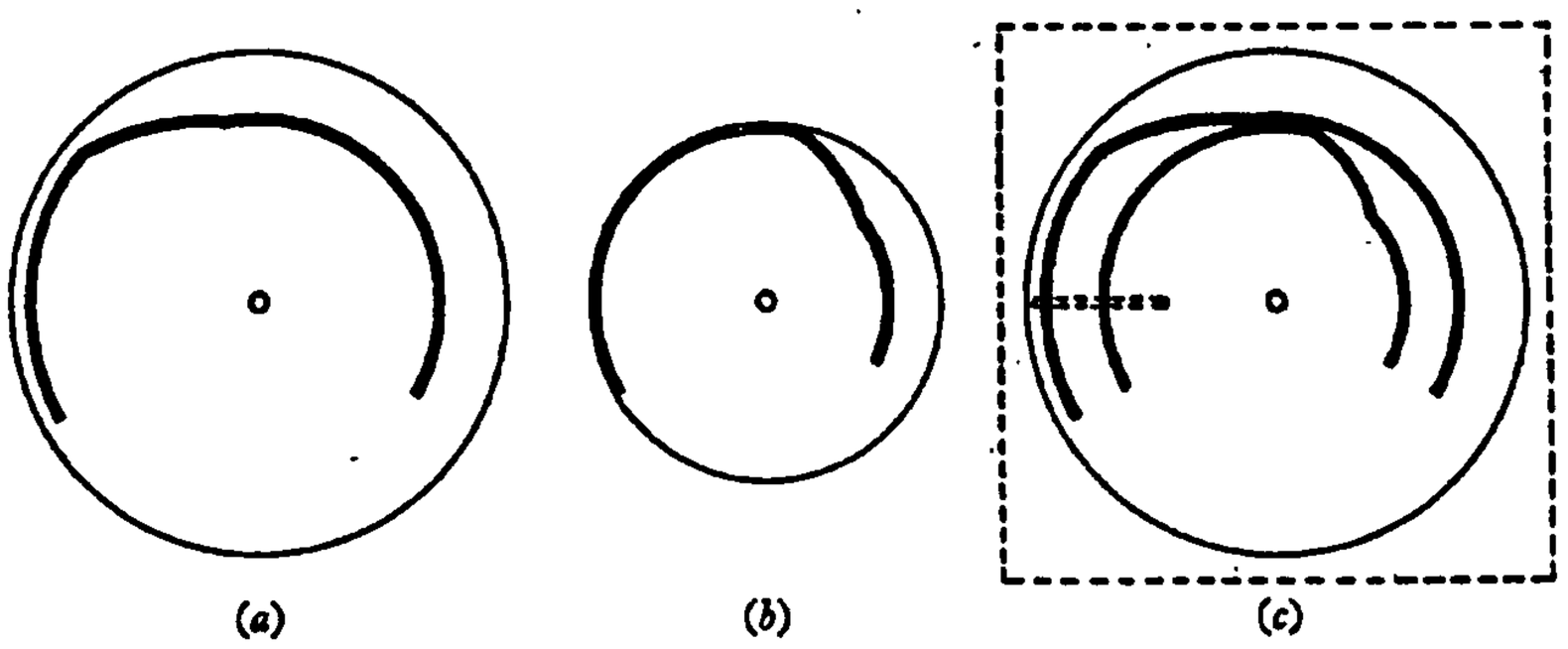
the *timeframe* of causal relations. Experimental work most closely relevant to the problem has investigated the importance of temporal contiguity in causal induction.

2.1. Previous Investigations into the Role of Temporal Contiguity in Causal Induction

2.1.1. Michotte's (1946/1963) Launching Paradigm

One of the earliest investigations related to the question of how temporal contiguity influences causal induction was reported by Michotte (1946/1963). His experimental apparatus consisted of a disk covered by a mask except for a small slit (the view-port). Painted on the disk were circular lines centred around the midpoint of the disk. The lines were not perfect (concentric) circles, however, but instead had some curvature (i.e. the distance from the midpoint changed with the angular distance from the slit). Figure 2-1 depicts typical disks as Michotte used them. If the disks were rotated behind the mask, the viewer perceived the lines as approaching and receding objects. Michotte employed various disks with different line parameters in his experiments in order to vary the spatial and temporal parameters of the objects perceived by the participant. For my purposes here, one configuration is of particular importance: the one that is commonly referred to as the "launching paradigm". In this setup, one object (A) is perceived as moving towards a stationary object (B). As soon as A collides with B, B moves away on the same trajectory, while A remains stationary. This usually creates a powerful impression of causality: object A "launched" object B, in other words A caused B's movement. If, however, the disks were arranged in such a way that object A appeared to collide with B, and then both A and B remained stationary for a couple of seconds, followed by B moving away, the causal

Figure 2-1. Typical disks as used by Michotte (1946/1963). The horizontal line in (c) represents the view-port. Taken from Michotte, A. E. (1946/1963). *The perception of causality* (T. R. Miles, Trans.). London, England: Methuen & Co.



illusion disappeared: A no longer appeared to cause B's motion; instead B appeared to move on its own.

Based on his findings, Michotte put forward his theory of "perceptual causality"; his claim was that certain (physical) events give rise to direct causal perception. He cited participants' subjective reports of the powerful causal illusions, more particularly statements indicating that participants could not "help" but see A cause the motion of B, even though they knew A and B were simply drawings on paper disks and were in no way causally linked. Michotte's claims were, of course, in stark contrast to Hume's (1739/1888) notion of causality as an idea, a construct resulting from a mental process. Michotte's claims, however, also shared one important aspect with Hume's theories about causality: they emphasised the role of temporal contiguity in causal induction. Just as Hume pointed out 200 years earlier: Michotte's findings appeared to experimentally demonstrate "that whatever objects are consider'd as causes or effects are *contiguous*" (Hume, 1739/1888, p.75).

2.1.2. Shanks, Pearson, & Dickinson's (1989) Instrumental Paradigm

Although Michotte's (1946/1963) demonstrations of the importance of temporal contiguity for causal judgment influenced subsequent work on human causal judgment, they at the same time had only limited generalizability. One restriction was that the launching paradigm was based on one-trial observations, whereas the bulk of causal learning literature emerging between the 1960s and 1980s was concerned with causal judgments derived from covariational information, i.e. multiple pairings of causes and effects (e.g. Jenkins & Ward, 1965; Allan & Jenkins, 1980; Mendelson & Shultz, 1976; Siegler & Liebert, 1974; Shanks, 1985, 1987; Wasserman & Neunaber, 1986). Another mostly methodological concern was that the majority of this work employed instrumental learning paradigms, i.e. experimental situations in which participants interacted with the apparatus, rather than passively watching it as in Michotte's studies. Typically, participants had to find out how strongly pressing a button or a key caused a certain event (usually a bulb lighting up). The apparatus would be constructed in such a way that there was a probabilistic relationship between the participants' actions and the effect. Participants in such studies were left to interact with the apparatus either for a fixed amount of time (e.g. 2 minutes), or until they emitted a certain number of responses (e.g. 25 key presses). After this period of evidence sampling, they had to indicate the extent to which their actions caused the effect.

Inspired by Michotte's (1946/1963) earlier work, it seemed only logical for Shanks, Pearson, and Dickinson (1989) to systematically investigate the role of temporal contiguity in human causal induction from an instrumental paradigm. Their task involved judging how strongly pressing the SPACE bar made a triangle flash on a computer screen. Participants could interact with the computer for a fixed amount of time and sampled evidence by

repeatedly pressing the SPACE bar and observing whether or not the outcome occurred. The apparatus was programmed to let the triangle flash with .75 probability each time participants pressed the SPACE bar (and never when the bar was not pressed). Shanks et al. varied the temporal interval between (reinforced) presses and the occurrence of the effect from 0 seconds to 16 seconds. Participants' estimates of causal effectiveness decreased systematically as the delay increased. In fact, if causal actions and observed effects were separated by more than two seconds, participants in Shanks et al.'s study evaluated a .75 contingency schedule to be just as ineffective as non-contingent control conditions. In other words, participants could no longer distinguish between causal and non-causal relations after a delay of more than two seconds. Several subsequent studies reported by Reed (e.g. 1992; 1999) used a similar paradigm and reported the same results: delays always impaired causal judgments. Reed's experiments focussed on "signalling" as a way to alleviate the detrimental effects of delay. He adapted Shanks et al.'s paradigm, but added a signal (usually a row of Xs) to fill the gap between participants' (reinforced) key presses and the triangle lighting up. Reed could show that such signals alone significantly improved the assessment of delayed contingencies, even though the signals were never mentioned in the instructions. One could argue, however, that contingencies are no longer delayed, if participants receive immediate feedback (via the signal) about the effectiveness of their actions, even though the instrumental action and the reinforcer are, strictly speaking, still separated in time. Be that as it may, Reed reproduced Shanks et al.'s finding regarding detrimental effects of delay in the control conditions that involved the absence of immediate feedback.

2.1.3. Developmental Studies in the Piagetian Tradition

Developmental psychologists in the 1970s were interested to find out which of the two Humean (1739/1888) cues to causality – contiguity and

regularity (covariation) – are more fundamental to children's conceptions of causality. Piaget (1969; as cited in Siegler & Liebert, 1974), had argued that causal reasoning in children under the age of seven or eight years was basically immature, in that young children always rely on temporal contiguity as a cue for causality, irrespective of the degree of regularity. I will review two experiments that aimed to contrast the two Humean cues against each other. The basic strategy employed in both of these studies was to construct a situation where an effect could be explained by either of two candidate causes. One candidate consistently covaried with the effect, but was not contiguous with it; the other candidate did not consistently covary with the effect, but was contiguous with it.

2.1.3.1. Mendelson & Shultz's (1976) Bell Box Study

Mendelson and Shultz (1976) employed an apparatus consisting of two wooden boxes. One (bottom) box housed a bell inside it, and the ringing of this bell constituted the effect; the task of the participants (children between 4.5 and 7.5 years of age) was to explain what made the bell ring. The other (top) box had two holes on its top, into which marbles could be dropped by the experimenter. The holes were marked by different colours; dropping a marble in an individual hole (A or B) constituted a candidate cause. An additional manipulation consisted in whether or not the two boxes were connected by a rubber tube ("physical model present vs. absent"), positioned in such a way to elicit the imagination that balls dropped in one box could travel down the tube to the bell box (the paper is frustratingly vague about the details of this manipulation). Regardless of this manipulation, the two boxes never were actually connected, and marbles dropped in the top box never travelled down the tube to the bell box; marbles always remained in the top box, and the bell in the bottom box was always controlled by a hidden footswitch operated by the experimenter.

Children were presented with three different observations: 1.) a ball was dropped in hole A, five seconds later a ball was dropped in hole B, followed by an immediate ringing of the bell (A—BX). 2.) a ball was dropped in hole B, but nothing happened (B), 3.) a ball was dropped in hole A, and five seconds later the bell rang (A—X). Each child had the opportunity to make each observation twice, resulting in six observations altogether. It is evident that A consistently covaried with the effect, but was separated from it by a five second delay, while B did not consistently covary with the effect, but on trials where B and X co-occurred, they were contiguous. After completing the observation phase, the experimenter asked the child four questions: a) if A is present, will X happen? b) if B is present, will X happen? c) what makes X happen? d) Make X happen, using either A or B. Mendelson and Shultz assigned values of 0 and 1 to each answer in such a way that 1 always indicated a preference of causal attribution to A, and 0 reflected a preference for B.

The results from Mendelson & Shultz (1976) indicate that children in the model:present condition preferred the consistent, non-contiguous cause A, whereas children in the model:absent condition (the boxes apparently were spatially separated and unconnected, although the article does not mention this) mostly attributed causality to the contiguous, inconsistent cause B. The authors' attempt to determine which of the two Humean cues to causality is more fundamental failed:

> It is somewhat hazardous to conclude anything about the relative importance of covariation and temporal contiguity from the results of this experiment. Either principle may be applied, depending on the existing conditions. Such a conclusion leaves unanswered the question of the fundamental basis of causal attribution. How does a child "know" when to apply covariation and when to apply temporal contiguity? Is there yet another, more essential, principle of causal

inference which is used to construct this knowledge? (Mendelson & Shultz, 1976, p.412)

There is one aspect of Mendelson & Shultz' (1976) paradigm that is hugely problematic: it rested on deceiving the participants. Regardless of which condition the children were in, the effect never really had anything to do with either one of the two candidate causes A or B. Even though the children could not see the foot switch through which the experimenter controlled the bell, one cannot know whether the children might not have considered the possibility that something other than A or B was responsible for ringing the bell. Because the procedure did not give the participants the possibility to express such a belief, but instead forced them to pick either A or B, they may have just selected one of them according to some preference, or even randomly. Mendelson & Shultz performed an analysis that ruled out that children responded randomly, but that still does not guarantee that participants responded according to their causal beliefs. Be that as it may, there is a contiguity bias inherent in the data, suggesting that temporal contiguity is important for causal induction. The principal finding of a main effect of model present vs. absent will be discussed in section 2.2.3.

2.1.3.2. Siegler & Liebert's (1974) Light Bulb Study

Siegler & Liebert's (1974) procedure involved an electrical apparatus consisting of three boxes; one was a computer on which, when turned on, lights would flash according a to pre-programmed quasi random sequence, one was a "card programmer" into which differently coloured IBM cards could be inserted, and one was a stand on which a light bulb was mounted. The children (5 to 6 and 8 to 9 year olds) were instructed that their task was to find out what makes the bulb light up. The procedure was always as follows. The experimenter turned the computer on, which, presumably (not explicitly mentioned, however), made the lights on the computer flash, and emit clicking noises. During the next 80 seconds the (effect) light bulb on the stand

illuminated eight times at pre-programmed random points for one second each. Presumably (although, again, not explicitly specified), the computer continued to emit clicking noises and flash its lights for the whole 80 seconds. After this period, the experimenter explained to the children, "Now we've seen that the computer can make the light go on". He next mentioned that there would also be another way of making the light go, and that would be by inserting cards into the programmer. He explained that he would insert 6 (or 12) different cards into the programmer. The instructions explicitly emphasised that both the computer and the card programmer could make the light go, but every time that the bulb would light up, the child would have to decide whether the light was turned on by the computer or the card programmer.

The timing between inserting the card and the bulb lighting up was determined according to four experimental conditions: 1) the light flashed immediately each time the card was inserted (100% contiguous, 6 trials), 2) the light flashed 5 seconds after the card was inserted (100% delay, 6 trials), 3) the light flashed 6 out of 12 times a card was inserted, and when it did, it did so immediately, and the light never flashed on any of the other trials (50% contiguous, 12 trials), 4) the light flashed 6 out of 12 times a card was inserted, but only 5 seconds after the card was inserted, and at no other time (50% delay, 12 trials). Thus, every participant observed the effect 6 times, and every time they had to indicate whether they thought the computer or the card caused the bulb to light up on that trial. After the last trial, children were asked whether they thought "overall" it was the card or the computer that made the light flash.

Analyses on both the trial-by-trial and the overall attribution data revealed significant effects of contiguity only, but no main effects of regularity: participants were more likely to identify the programmer as the causal agent, if the light flashed immediately after a card was inserted, irrespective of whether the light flashed 100% or 50% of the time a card was inserted. A more detailed analysis of the trial-by-trial data revealed that

towards the end of the experiment the older age group (8 to 9 year olds) showed some sign of recognizing the importance of regularity.

The most problematic aspect of Siegler and Liebert's (1974) paradigm is the way they set the two competing causes (computer and programmer) against each other. The computer was always presented as a good plausible cause on its own before the actual experiment begun. Furthermore, the pre-experimental 80 seconds exposure to it demonstrated that the computer controls the light completely randomly. The card programmer, in contrast, was never presented in isolation, but always in conjunction with the computer also being activated. This constitutes a classic blocking paradigm (see section 1.2): if a novel candidate cause (here: card programmer) is only ever presented in the presence of an already established predictor (here: computer), and the presence of both candidates produces the effect, one cannot learn anything definitive about the causal status of the novel cue. Instead, one has to be uncertain, and rely on other sources for causal judgments. The contiguity bias in the data thus may reflect a primitive form of the representativeness heuristic (Tversky & Kahneman, 1974). In the contiguous conditions children observed an instantaneous pairing between inserting the card and the bulb lighting up. This evidence might have violated their ideas about the randomness with which the computer caused the bulb to light up during the pre-experimental exposure. This violation of the idea of randomness might then have led them to attribute causality to the card instead of to the computer. In the delayed conditions, in contrast, the feedback was not in such direct opposition to the randomness experienced before. Even with 100% regularity, the pairing between card and light may have been hard to notice, particularly with a host of clicking and flashing happening on the computer in the intervening five second interval. Regardless of the methodological and theoretical criticism one can cite against both Siegler & Liebert's (1974) and Mendelson and Shultz' (1976) studies, they both suggest that temporal contiguity is a very important principle in children's causal induction.

2.1.4. Human Sensitivity to the Timeframe of Causal Relations in Real World vs. Laboratory Tasks: A Paradox?

The preceding sections reviewed experimental evidence relevant to the question how people can identify causal relations in real time, in situations where the co-occurrence between cause and effect is not as self-evident as in the kind of experiments discussed in chapter 1. This evidence paints a rather unflattering picture of human causal induction, however. Across very different experimental paradigms the results converged to the same conclusion: if cause and effect are separated by more than a few seconds, people fail to correctly identify the causal relation between them.

This finding clashes with everyday causal cognition, where people seem to be able to identify with relative ease causal relations where cause and effect are separated in time such as those between infection and outbreak of a disease, sexual intercourse and pregnancy, or sowing seeds and plants growing. Before I proceed to try and resolve this paradox, I will review in sections 2.2.1 and 2.2.2 how two major theoretical frameworks – associationism and causal power as introduced already in chapter 1 – explain the importance of temporal contiguity demonstrated in laboratory studies. The third major approach to human causal induction reviewed in chapter 1, computational causal power (Cheng, 1997) remains silent about issues of delay and temporal contiguity, and will therefore not be discussed in section 2.2. Unlike associationism which offers process models of causal induction, the power PC theory is a computational level description (Marr, 1982) of the inference process which takes covariational information as its input. Variations in temporal contiguity therefore fall outside its scope. I will, however, come back to this point in section 6.3.

2.2. Explanations for the Importance of Temporal Contiguity

Section 2.1 reviewed experimental results from studies inspired very different theoretical backgrounds. Michotte's (1946/1963) studies on mechanical causality were influenced by the causal power view (Kant, 1781/1965) as outlined in section 1.3. Shanks et al.'s (1989) paper, on the other hand, followed an associationist tradition. Although experiments carried out in both the causal power and the associationist tradition revealed converging results regarding the importance of temporal contiguity, the reasons given as to why participants either failed to report a causal relation, or reported a substantially degraded relation when cause and effect were separated by only a few seconds, vary between accounts.

The developmental studies described in section 2.1.3 were inspired by a Piagetian framework, and were mostly aimed to illustrate how causal induction changes as the child reaches Piaget's various developmental stages. These theories are not relevant for my purposes here, but some aspects of the findings bear direct relevance to the explanations reviewed here, and I will point those out in section 2.2.3.

2.2.1. Associationism

According to an associationist interpretation of causal inference, causal learning is identical to associative learning. Judged causal strength reflects no more than the associative strength between candidate cause (equivalent to a conditioned stimulus or response) and effect (unconditioned stimulus) (Shanks & Dickinson, 1987). Every time a cause and an effect co-occur together, the association between them is strengthened until it reaches the maximum strength the effect can support (i.e. asymptote), and every time the cause fails to produce the effect, the cause-effect association weakens (cf. section 1.2).

Associationist theorists were very much inspired by David Hume's (1739/1888) treatise on causality (in fact, the name "associationism" probably goes back to Hume's philosophical enquiries about how "associations" can give rise to "complex ideas"). Hume had postulated two necessary cornerstones for causal relations: contingency and contiguity. Section 1.2 already discussed how associationism accommodates contingency as a cue to causality. But claiming heritage to Hume also entails holding contiguity as an essential cue to causality.

Shanks and Dickinson (1987, p. 231), in a paper outlining the principles of an associationist account of human causal learning explained the importance of temporal contiguity as follows: "Contemporary accounts are usually silent about the actual interevent interval over which an association can be formed, but all argue that the size of the increment in associative strength accruing from a pairing decreases as the contiguity is degraded." The message is clear: compared to immediate cause-effect pairings, delayed ones will always deliver weaker evidence for the causal relation in question. This is not to say that associationism would claim that humans can only learn causal relations when cause and effect follow each other immediately, even though it is certainly tempting to jump to this conclusion. In a more recent theoretical paper David Shanks wrote about his earlier experiments: "It should be emphasized (...) that much longer delays can certainly be tolerated in other situations. The slope of the contiguity function is likely to be highly task-specific." (Shanks, 1993b, p.323). This statement was clearly intended to account for such rare findings as the Garcia effect (Garcia, Ervin, & Koelling, 1966; Garcia & Koelling, 1966), which demonstrated that animals can bridge considerable time-spans in taste-aversion learning paradigms. In such experiments, the animal typically has free access to a substance with a novel flavour (e.g. saccharin flavoured water). Some time after having ingested the novel tasting substance, the animal is made to feel ill (either by being injected with apomorphine hydrochloride, e.g. Garcia et al., 1966; Schafe, Sollars, &

Bernstein, 1995; or by being exposed to X-rays, Garcia & Koelling, 1966). Garcia could show that the animals attributed the experienced sickness to the new food (rather than a competing predictor, say, a flashing light), and that they could learn the relation even when CS (flavour) and US (illness) were separated by a 75 minute delay. The critical aspect of this finding for my purposes is that such taste-aversion learning experiments have shown animals to be capable of forming cue-to-consequence associations even when the stimuli are considerably separated in time. Such long delays can only be tolerated in this specific paradigm, however, which led theorists to propose that gustatory and olfactory cues are "biased" to be associated with internal malaise, "even when these stimuli are separated by long time periods" (Garcia et al., 1966, p.122).

Associationism does not provide any clues about which parameters determine how much delay participants can tolerate in certain situations, or how one could explain "biases" or "task-specificity" other than by attributing them to hard-wired preferences (the same as prior knowledge?) shaped by natural selection (as suggested by Garcia et al., 1966; Garcia & Koelling, 1966). Be that as it may, an associationist account of human causal learning implies that temporal contiguity between cause and effect is an essential component for successful causal induction. The experimental results I have reviewed in section 2.1.2 suggested that humans fail to identify causal agency if their own causal actions are separated from the effects by more than two seconds. Even though one can neither interpret these results to reflect absolute boundary conditions constraining the inference process, nor get any hints as to how such boundary conditions might be determined, associationism is clear in postulating that contiguity is an important principle in causal induction. Everything else being equal, contiguous cause-effect sequences should always result in stronger impressions of causality than non-contiguous sequences. This postulate received a serious blow in 1995 when Glenn Schafe and colleagues (Schafe et al., 1995) published a study which re-investigated the

boundary conditions of taste-aversion learning in rats. Their experiments showed that rats failed to associate a novel taste with illness, if the CS-US interval was very short (10 seconds), and only learned the connection if CS and US were separated by delays of at least 15 minutes. In other words, a delayed pairing of CS and US resulted in higher increments of associative strength then a contiguous pairing of the same stimuli. This finding of course is in stark contrast to the principles of associative learning, which postulate the importance of contiguity (cf. Dickinson, 2001). To my surprise, Schafe et al.'s paper has received very little attention in the scientific community, and there are no suggestions as to how associationism could be modified to explain their result.

2.2.2. Causal Power

According to the causal power view (Ahn, Kalish, Medin, & Gelman, 1995; Bullock, Gelman, & Baillargeon, 1982) the crucial component of causal inference is knowledge about some causal power or mechanism linking cause and effect. The regular co-occurrence between a cause and an effect can only be rendered to reflect a causal relation, if the reasoner knows of a causal mechanism that explains how the cause could bring about the effect. In Michotte's (1946/1963) launching paradigm the stimuli created a visual illusion of two moving objects. Even though the observer knew that he or she was not viewing real physical objects, but only drawings on paper disks, the illusion was strong enough to get the observer to apply his or her knowledge about the physical properties of impact to the stimuli. Our world experience tells us that if a moving object collides with another, stationary, object, and the force of impact is sufficiently large to set the stationary object in motion, it does so at the instant of impact. Because participants in Michotte's and similar studies presumably applied their naïve understanding of physics to the task, they evaluated contiguous launching events as reflecting a causal relation

(A caused B to move away), but refrained from doing so if contiguity was disrupted by a delay (B moved on its own).

The causal power explanation for why previous experiments revealed contiguity as an important prerequisite in causal induction differs from the associationist explanation in one very important respect. Whereas associationism holds contiguity as an essential cue to causality, causal power theory does not make any such claims. Whether or not a particular action sequence gives rise to a causal impression depends on the observer's assumption about the potential mechanism linking cause and effect. If the assumed mechanism implies immediacy (as in the launching paradigm), only contiguous sequences should be interpreted as causal, whereas delayed sequences should fail to create such impressions. One could, however, easily imagine a causal mechanism that does not imply immediate cause-effect pairings, for instance as between sexual intercourse and giving birth. If the reasoner assumes such a delayed mechanism, delayed sequences should readily be judged as causal, whereas immediate sequences should fail to create a causal impression. According to causal power theory, there is nothing special about temporal contiguity. What matters are the assumptions about the causal mechanism that the reasoner brings to the task.

2.2.3. Einhorn and Hogarth's (1986) Knowledge Mediation Hypothesis

As I have already mentioned in chaper 1, associationism was inspired by Hume's (1739/1888) philosophy of causality, while causal power was endorsing a Kantian (1781/1965) understanding of cause. A decade before Cheng's (1997) formal unification of these two seemingly mutually exclusive frameworks, Einhorn and Hogarth (1986) already had tried to combine useful aspects from both philosophies. Cheng's analysis focused exclusively on the question how human reasoners can take the mental leap from covariation to

causation, and did not address the necessary precursor: how intelligent organisms notice regularities (covariations) in the first place. Einhorn and Hogarth's review did not offer a formal analysis or computational model the way Cheng's article did, but addressed the problem of temporal contiguity.

Einhorn and Hogarth (1986) postulated, in line with Hume's (1739/1888) treatise, that contingency (i.e. covariation) and contiguity both are very important cues to causality. The earlier developmental work I reviewed in section 2.1.3 had tried to identify which of the two principles places stronger constraints on the inductive process, but failed to come to any definitive conclusions: "It is somewhat haphazard to conclude anything about the relative importance of covariation and temporal contiguity (....) Either principle may be applied, depending on the existing conditions." (Mendelson & Shultz, 1976, p.412). Einhorn and Hogarth argued that the main function of contiguity is to be an "important cue for directing attention to contingencies between variables, and such contingencies may then be considered as to their causal significance." (Einhorn & Hogarth, 1986 p.10). In other words, contingency or covariation is the more important one of the two cues; contiguity merely enables the reasoner to notice a covariation, but the crucial information relevant for causal assessment lies in the covariation itself. A subsequent formal analysis of a range of developmental data (including the experiments reviewed in section 2.1.3) by Cheng (1993) confirmed Einhorn and Hogarth's point by showing that covariation is a necessary component of all causal relations, even in those paradigms that aimed to contrast contiguity and covariation against each other.

So far, Einhorn and Hogarth's (1986) analysis fits the paradoxical findings from the laboratory studies that uniformly demonstrated how people fail to identify causal relations when cause and effect are separated by a delay. However, Einhorn and Hogarth also accounted for our intuitions about everyday causal inference, where people can recognize causal relations involving delays as follows:

When temporal and/or spatial contiguity is low (or temporal contiguity is erratic), inferring causality becomes more difficult. That is, in the absence of contiguity, relations are hard to develop, unless one uses intermediate causal models to link the events (...). For instance, the temporal gap between intercourse and birth requires some knowledge of human biology and chemistry to maintain links between those events (Einhorn & Hogarth, 1986, p.10).

This *Knowledge Mediation Hypothesis* of course borrows heavily from the causal power account described in section 2.2.2. In fact, Einhorn and Hogarth (1986) go on to illustrate that in certain scenarios high temporal contiguity may conflict with other cues to causality, most notably covariation. Consider the example of a non-smoking man who takes up smoking on a particular day and is diagnosed with lung cancer the following day. The high temporal contiguity and perfect contingency between smoking and lung cancer conflict with each other. Even though we hold smoking as an established cause for lung cancer, the timeframe in this example is much too narrow to render the causal connection plausible. Some other alternative causes (perhaps having worked as a coal miner for 20 years) must have produced lung cancer in this example.

Einhorn and Hogarth's (1986) knowledge mediation hypothesis somehow fuses the explanations offered by associationism and causal power. Although it acknowledges temporal contiguity as a useful cue that helps to identify contingencies between causes and effects, it does not bestow an especially privileged role to it. People can overcome the need for temporal contiguity, if they know of some causal link, mechanism, or chain that takes time to unfold. In other words, the influence of time is mediated by prior knowledge. Whether or not a particular (contiguous or delayed) covariation will be judged causal depends on the assumptions about the causal mechanism (in particular the assumptions about the timeframe of this mechanism) that a reasoner brings to the task. The Knowledge Mediation hypothesis nicely

accounts for Mendelson and Shultz' (1976) findings: children were more likely to attribute causality to a delayed candidate cause, when they were given a rationale for this delay (a rubber tube through which the marble might have travelled); in the absence of such a rationale (no connection between the boxes), they could not bridge the temporal gap and attributed causality to a contiguous alternative instead.

It is important to distinguish the Knowledge Mediation hypothesis from the signalling effect (cf. section 2.1.2). The typical finding from signalling experiments (e.g. Reed, 1992, 1999) is that a neutral novel signal stimulus delivered immediately after the occurrence of a reinforced response can bridge the temporal gap between response and outcome, relative to control conditions involving the same temporal gap but no signal stimulus. The crucial difference between Signalling and Knowledge Mediation is that the former, but not the latter relies on noticeable changes regarding the perceived evidence to alleviate detrimental effects of delay. A visible signal, such as a row of Xs, is perceivable evidence; assumptions about causal mechanisms, in contrast, are not.

2.3. Distinguishing Knowledge Mediation from Associationism

2.3.1. Generating Predictions

For the purposes of generating predictions, the knowledge-mediation and causal power accounts do not differ. Both frameworks state that whether or not a particular covariation will be identified as causal depends on the reasoners' prior assumptions about the causal mechanism in question. They are, therefore, top-down approaches to causal reasoning: pre-existing mental concepts or structures determine how subsequent sensory experience will be parsed and processed. It follows that identical sensory experiences could

potentially be parsed differently, if the reasoner applies different assumptions to the task. Imagine that a person presses a button, and a minute later a green light illuminates. If the person assumes the connection between the button and the light to be an ordinary electric circuit, this delayed course of events would not qualify as a causal sequence. Our prior knowledge about electricity entails that the connection is very fast, below our perceptual threshold; the light should turn on instantly, but it did not. Probably the electric circuit is broken (or button and light never were connected in the first place), and the light lit up for some other reason (e.g. someone else might have turned it on with a different switch). If, however, the person assumes the button triggers a timer-relay, as on a pedestrian crossing, a delay of one minute between pressing the button and the signal changing is within the expected range, and the causal connection is obvious. Because the assumptions about the causal mechanism (and the implications about the timeframe of the relation) were different in the two situations, the very same sensory experience of a delayed covariation will give rise to a causal interpretation in the latter, but not the former scenario. Whether delay is detrimental to causal induction or not therefore depends on the temporal assumptions the reasoner brings to the task. A logical consequence of the above argument is that Knowledge Mediation under certain circumstances would also predict a detrimental effect of contiguity. If a reasoner expects a delayed relation, immediate contingencies should not be attributed to the causal relation in question. If the person pressed the button and an instant later the traffic light turns green, one would most likely think that either the light would have changed anyway, or that someone else had already pressed the button earlier.

The bottom-up nature of associationism stands in stark contrast to the top-down ideas of Knowledge-Mediation. In associationism, sensory experiences give rise to mental connections (associations) according to the laws of a specific learning mechanism, and every abstract idea can be reduced to be the result of associations. Prior knowledge has no place in such an

account, unless it is the result of previous associations itself (i.e. pre-existing associative strengths between a particular cue and outcome). What is not, possible, however, is that abstract ideas influence how sensory experience will be parsed. The same covariational evidence should always give rise to the same causal impression. While associationism acknowledges that different stimuli and effects may vary as to how far they can be separated by a delay for reasoners to still learn the connection successfully (Shanks, 1993b), it cannot account for different interpretations of exactly the same evidence, as outlined in the previous paragraph. Moreover, it also predicts that if two situations are identical in all respects, except that in one the relation between cause and effect is more contiguous than in the other, the contiguous pairing should consistently give rise to stronger impressions of causal strength. In associationism it is not possible that a delayed relation elicits a stronger association than an immediate relation. Schafe et al.'s (1995) results that I reviewed in section 2.2.1 have already contradicted this principle.

2.3.2. Current Evidence does not Favour One Account over the Other

Knowledge-mediation in causal inference is well studied, but research has mainly focused on how prior beliefs interact with covariation assessment: how prior knowledge determines whether a given covariation will give rise to causal impressions or not (e.g. White, 1995; Ahn et al., 1995; but see also Lien & Cheng, 2000), and how previously acquired category structures (Waldmann & Hagmayer, 1999) or assumptions about the direction of learning (predictive vs. diagnostic, see Waldmann & Holyoak, 1992; Waldmann, 2000) influence covariation assessment in causal induction. However, whether and how beliefs about the timeframe of causal relations influence adults' assessment of delayed or contiguous contingencies is not yet clear.

A careful analysis of the predictions derived from the knowledge mediation and associationist accounts and the experimental materials employed in previous investigations reveals that current findings as reviewed in section 2.1 cannot distinguish between these two competing explanations. The reason for this lack of discrimination is that all these experiments employed scenarios where participants either expected contiguous cause-effect pairings, or used paradigms that lend themselves to such interpretations.

Michotte's (1946/1963) and related perceptual causality experiments presented participants with images of colliding objects. Participants brought to the task their world knowledge about the physical properties of impact and therefore expected immediate motion after impact. If this expectation was violated, the sequence was judged as non-causal. The same principle applies to Shanks et al.'s (1989) instrumental paradigm, although the similarity is not immediately obvious. The experiment was conducted on a computer, and even though computers were not as widely used then as they are now, undergraduate students probably already had basic ideas about, if not prior experience with, interacting with computers. Among these ideas certainly is that a computer is an electronic device, and therefore, like a light bulb connected to a light switch, should – in principle – react to input immediately. Again, causality was underestimated or even negated, if this expectation of immediacy was violated. In fact, Shanks et al. were already aware that this particular problem could potentially limit the generalizability of their results: "subjects in judgment studies such as ours assume that the word 'causes' in the experimental instructions means 'causes immediately'. After all, they presumably have considerable experience of the immediacy of cause-effect relations in such electrical devices as computers" (Shanks et al., 1989, p.155).

The pattern of findings reported in the experimental literature – delays impair causal reasoning performance – is thus compatible both with the Knowledge Mediation hypothesis and the associative learning approach, which denies knowledge mediation. It is impossible to know whether participants in

earlier experiments refrained from attributing causality to delayed event sequences because temporal contiguity is essential to causal learning (as associationism would argue), or because an expectation of immediacy was contradicted by an experienced delay (as predicted by the Knowledge Mediation account). What is needed to disentangle the predictions from both accounts and to allow a systematic investigation into the role of temporal contiguity in causal induction is a paradigm that explicitly addresses and manipulates participants' expectations about the timeframes of the causal relations in question. In addition, participants would have to be exposed to immediate as well as delayed cause-effect pairings. In other words, the design would comprise the factors (experienced) Time and Knowledge/Assumptions (about the timeframe). Either temporal contiguity is as essential for causal learning as associationism suggests, in which case one would only expect a main effect of time, but no effects or interactions associated with Knowledge. Alternatively, contiguity may not be essential to causal induction, but meaningful interpretations of event streams crucially hinge on knowledge about the timeframe of the causal relation in question, in which case different temporal assumptions should lead to diverging interpretations of identical evidence, i.e. one would expect main effects and/or interactions associated with Knowledge. Such studies would effectively constitute an experimental test of Einhorn & Hogarth's (1986) knowledge mediation hypothesis, which is, surprisingly, still lacking in the literature. Mendelson and Shultz' (1976) study, which preceded Einhorn & Hogarth's review by a decade, comes closest to such a design. The physical model (rubber tube) present vs. absent manipulation effectively manipulated whether or not a delay between cause and effect was plausible. Unfortunately their paper lacked overall methodological rigor and the analysis of the data left many questions unanswered. This may well be a reason for why it was not even cited in Einhorn & Hogarth's review.

2.4. A New Paradigm

Anne Schlotmann (1999) reported a developmental study that involved immediate and delayed causal relations, and also explicitly manipulated participants' knowledge about the timeframe of the causal mechanism in question. The paradigm relied on the basic idea used by Mendelson and Shultz (1976), but Schlottmann improved the method considerably. Her apparatus consisted of a "mystery box" which could contain one of two toys. The box had two holes on its top, and the toys inside the box could be placed in such a way that balls dropped into the holes fell onto the toy. One of the toys made a bell ring immediately if a ball was dropped onto it (the weight of the ball made a see-saw swing and hit the bell), while the other toy made the bell ring a few seconds after the ball was dropped onto it (the ball rolled down a sloped runway to crash into the bell). There was always only one toy inside the box, and it was positioned in such a way that only balls dropped from one of the two holes fell onto it and subsequently made the bell ring. Participants (five to ten year old children and a control group of adults) were left to explore the box and the two mechanisms, and learned that one mechanism made the bell ring immediately, while the other one involved a delay. The children then had to predict how long each mechanism would take to make the bell ring. They were also asked to make diagnostic inferences ("Which mechanism is inside the box, the slow one or the fast one?") after observing one ball being dropped in a hole, followed by the bell ringing either immediately or after a delay. Children of all age groups performed very accurately on these tasks, both when the fast and the slow mechanism were inside the box. This result indicates that they could use their knowledge of the different timeframes of the mechanism to bridge temporal gaps.

The critical test in Schlottmann's study (1999), however, pitted knowledge of mechanism and contiguity directly against each other and involved a forced choice between a delayed and a contiguous candidate cause.

First, the experimenter put the slow mechanism inside the box and confirmed that the child understood that the slow mechanism was inside the box. The child did not know, however, under which hole the mechanism was placed. The experimenter then dropped a ball in one hole (unbeknownst to the child: the one with the slow mechanism under it), waited for a couple of seconds, and then dropped a second ball in the other (ineffective) hole. The dropping of the balls was carefully timed so that the bell would ring immediately after the second ball was dropped. Children then had to indicate which ball they thought had made the bell ring. Results showed that contiguity was necessary at least for 5 and 7-year olds: they (erroneously) preferred to attribute causality to the second ball (contiguous cause), even though they explicitly knew that the operating causal mechanism involved a delay. Older children and the control group of adults correctly attributed causality to the first ball (noncontiguous cause). Schlottmann's results from the forced-choice task show that contiguity is a very important cue to causality at least for young children, but her results from the exploration phase also show that even young children could learn that the slow mechanism involved a delay, and could successfully use this knowledge to draw simple causal inferences of both predictive and diagnostic nature. In the complex forced-choice task they failed, however, presumably because they could not integrate knowledge of a delayed mechanism with the immediate perceptual feedback of contiguity they received through the second (noncausal) ball. Adults and older children learned that mechanism was superordinate to contiguity, and could thus successfully integrate the contiguous feedback with their knowledge of mechanism to come to the correct conclusion.

Schlottmann's (1999) study and results from the exploration phase provides very specific support for the knowledge mediation hypothesis (Einhorn & Hogarth, 1986), but her paradigm has some limitations that make it hard to generalize from her results. In particular, her experiments fall outside the scope of associative learning theory. The exploration phase, during

which participants learned about the two possible configurations of the apparatus was mostly driven by acquiring physical concepts and understanding how each of the two mechanisms worked. Although the children experienced balls being dropped into the mystery box several times, one could hardly refer to these encounters as learning trials in an associative learning sense. Participants' feedback was mostly in the form of a Socratic dialogue with the experimenter (e.g. "The slow runway ball is dropped first, and when it is almost there, then the fast seesaw ball is dropped second", Schlottmann, 1999, p. 307). Associative-learning algorithms operate via error-correction, driven by whether the effect occurs or not, given the presence of a particular candidate cause, but cannot represent feedback of such complex quality as in Schlottmann's study. Another limiting factor that is true for all causal learning paradigms in the mechanistic tradition is that they usually employ *deterministic* causal mechanisms. Both mechanisms in Schlottmann's study always made the bell ring, the only difference being how quickly they did so. Most causal relations, however, are *probabilistic*. One of the key advantages of covariation-based theories (including associative learning, but also the computational causal power approach, Cheng, 1997) is that they explicitly account for knowledge gained from probabilistic feedback. Furthermore, Schlottmann's paradigm involved observations of and interactions with real physical mechanisms (e.g. see-saws and runways). While this constitutes an important overall strength, it precludes conjectures about what may have contributed to participants' poor performance in typical laboratory tasks as reviewed in section 2.1, which often employed artificial stimuli (e.g. presentations on a computer screen). Finally, the dependent measures in Schlottmann's study were based on a forced-choice between two options. Systematic studies on the impact of delay (e.g. Shanks et al., 1989) used rating scales to probe causal judgments. Only the latter procedure, but not the former, is sensitive to gradual decreases in judged causal strength due to delays.

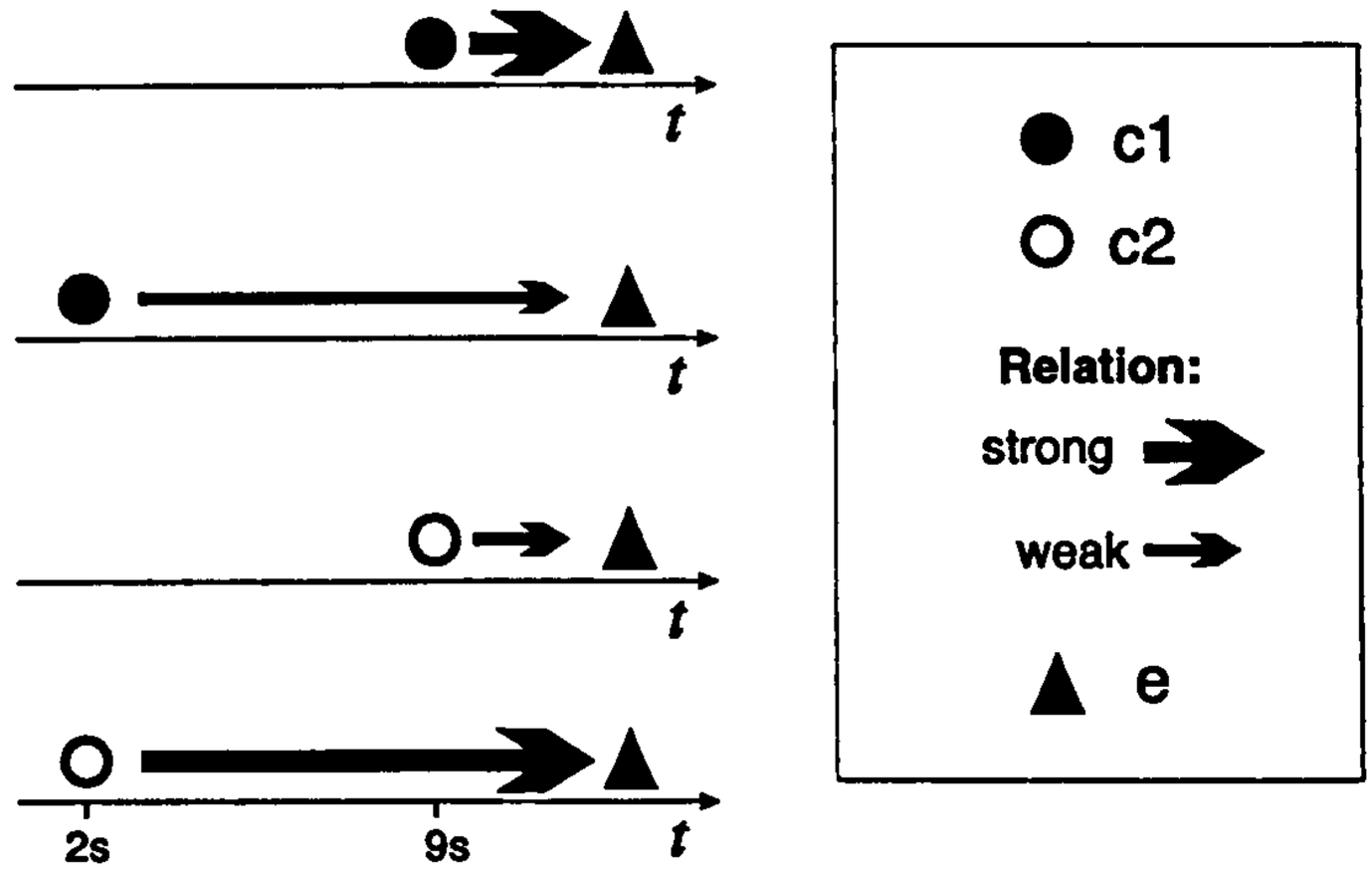
In the remainder of this thesis, I will examine whether Schlottmann's (1999) findings extend to learning derived from situations more commonly used in causal learning tasks with adult participants. The tasks I employed involved instrumental learning with probabilistic reinforcement. Participants were instructed to learn how strongly a particular action causes an outcome on a computer screen, and had to indicate their causal beliefs on a rating scale. As in Schlottmann's study, I employed immediate and delayed causal relations, but I manipulated participants' expectations about the timeframe of the causal relations in the instructions they received. First, however, I will investigate whether temporal contiguity by itself plays a privileged role in causal induction independent from how it may or may not interact with knowledge about the timeframe of causal relations. The experiment that I will report in chapter 3 employed a completely novel context, where (adult) participants could not apply any notions of mechanism. Instead, causal knowledge could only by learned from observing contingencies. In contrast to Schlottmann's design, this study will pit contiguity against contingency rather than against knowledge of mechanism. In particular, it will address whether adults can integrate these two Humean cues and come to an analogous conclusion that contingency is superordinate to contiguity. One could therefore also say, that this first study follows the tradition of Mendelson & Shultz (1976) and Siegler & Liebert (1974).

3. Experiment I: Contiguity vs. Contingency as Cues to Causal Strength

Experiment I explored people's sensitivity to time as an indicator of causal strength. More specifically, this experiment investigated whether people are capable of learning that the causal effectiveness of a certain event can vary depending on the temporal relation between the causal event and the target effect. According to previous results, people do attach various causal powers to the same events, depending on the degree of cause-effect contiguity. In Shanks et al.'s study (1989), for instance, participants reported that the same physical action (pressing SPACE) became less causal the further in time it was separated from the effect. Such results show that people degrade estimates of causal effectiveness for an event, if it is temporally separated from the effect. In other words, temporal distance served as a cue that indicated a decay of causal effectiveness. Shanks et al.'s findings have been interpreted to mean that temporal contiguity is essential for human causal induction.

An alternative hypothesis is that people use time as a more general indicator of causal strength. Under such an interpretation, temporal contiguity would not be essential and would not play a specially privileged role; it would merely be one of many values that the variable "time" can take, and it would not necessarily always facilitate the discovery of causal relations. If reasoners are sensitive to time as a variable that can carry causal information, then one should not only be able to demonstrate decays of judged causal effectiveness as temporal contiguity is degraded, but also the reverse finding. In particular, people should be able to learn that temporal distance sometimes may indicate an increase in causal effectiveness.

Figure 3-1. Causal structures employed in Experiment I. The causal effectiveness of candidate causes c1 and c2 changes dependent on the temporal position with respect to the effect *e*.



Experiment I sought to investigate whether people's sensitivity to time as a variable that carries information about causal effectiveness is confined to one direction only (i.e. degradation in temporal contiguity always indicates decays in causal effectiveness), or whether people can likewise learn the reverse relation (i.e. degradation in temporal contiguity sometimes may indicate increases in causal effectiveness). The paradigm adopted in Experiment I used events completely novel and unfamiliar to participants, in order to avoid any effects of prior knowledge or experience. Over a series of learning trials participants sampled information about the causal effectiveness of two candidate causes to produce an effect. Information was presented in such a way that only one cause occurred on any given learning trial, and the temporal position of that cause could be early or late in the observation period. At the end of each observation period participants were informed whether or not the effect occurred. To test whether people are sensitive to time as a dimension in general, a causal structure as illustrated in Figure 3-1 was employed: one cause (c1) had a strong causal effectiveness if it occurred late in the episode (i.e. contiguous with the effect), but a weak effectiveness if it

occurred early in the episode (i.e. non-contiguous with the effect); the causal power of the other cause (c2) had the opposite relation with respect to time: c2 had a weak causal effectiveness if it occurred late (i.e. contiguous with the effect), but a strong effectiveness if it occurred early (i.e. non-contiguous with the effect).

If people are able to make full use of time as a predictor of causal strength, their evaluations of causal strength for the relevant episodes should reflect the true underlying causal structures of both c1 and c2. Specifically, they should report that c1 is a strong cause if it occurs late, but a weak cause if it occurs early in the episode, and that c2 is a weak cause if it occurs late, but a strong cause if it occurs early in the episode. Alternatively, if contiguity is essential to causal induction, participants would be expected to correctly identify the causal structure of c1 only, because its structure is aligned in such a way that causal effectiveness decays as contiguity is degraded. Because temporal position changed the causal effectiveness of c2 in the opposite direction of what would be expected from such an account, one would expect causal ratings not to reflect the change in causal effectiveness in the same systematic way as for c1.

3.1. Method

3.1.1. Participants

Twenty-four students enrolled in an undergraduate psychology course at the University of Sheffield participated to fulfil part of a course requirement or to receive a small nominal payment.

3.1.2. Materials, Design, and Procedure

Instructions asked participants to pretend they were space troopers on a mission at a foreign planet, where they had to monitor an alien rocket launch site and a crater near the site. They were told that aliens sometimes pop out of the crater, and that their mission was to find out how strongly the aliens popping out of the crater caused the rocket to launch. The instructions specified that participants would watch the crater and site during several observation periods, and that all observation episodes had the same fixed duration. At the end of each episode participants had to predict whether they thought the rocket would launch or not. The instructions stressed that every episode constituted an independent observation, and whether or not the rocket launched in a particular episode was solely determined by what happened during that episode.

The causal structure implemented in this experiment involved one binary effect (the rocket launches vs. doesn't launch) and two candidate causes (big vs. small alien popping out of the crater). The temporal position (i.e. occurrence) of a candidate cause within a 12s fixed length observation period was either early or late (2s or 9s after start of an observation period, respectively). The experiment included two levels of causal power (1.0 and .3); causal powers were assigned such that one cause (c1) had a causal power of 1.0 if it occurred late in the episode, and a power of .3 if it occurred early in the episode. The other cause, c2, had the reverse assignment; c2's power was .3 if it occurred late in the episode, and 1.0 if it occurred early in the episode (see Figure 3-1). The assignment of the two types of stimuli (big or small alien) to the causal roles (c1 vs. c2) was counterbalanced. In one group (C1-Big), c1 was the big and c2 the small alien, in the other group (C1-Small) c1 was the small and c2 the big alien. For example, in the C1-Big group, episodes containing a late popping of the big alien or an early popping of the small alien were assigned a 100% probability of rocket launch; episodes

containing an early popping of the big alien or a late popping of the small alien were assigned a 30% probability of rocket launch (see Table 3-1). There were also base rate episodes, which involved neither c1 nor c2 and which were never followed by a rocket launch.

Table 3-1. Causal structures in Experiment I for the two counterbalancing groups. Temporal Position is defined relative to start of episode. Probability denotes the probability that a given episode was followed by a rocket launch. Base rate trials involving neither C1 or C2 are represented by a hyphen (-). Each combination of candidate cause and temporal position occurred 10 times.

Group	Candidate Cause	Temporal Position	Probability
C1-Big	C1 (big)	2s	0.33
		9s	1.00
	C2 (small)	2s	1.00
		9s	0.33
	-	n/a	0.00
C1-Small	C1 (small)	2s	0.33
		9s	1.00
	C2 (big)	2s	1.00
		9s	0.33
	-	n/a	0.00

The experiment comprised five categories of observation periods: four types of periods produced by factorial combination of candidate cause and temporal position (C1-early, C1-late, C2-early, C2-late), plus a base rate trial involving neither c1 nor c2. Each type of episode occurred 10 times throughout the experiment. Whether or not a particular episode was followed

by the effect was determined by the underlying causal structure. For instance, three out of the ten episodes with c1 occurring early, and all ten episodes with c1 occurring late, were followed by a rocket launch. Trials were scheduled randomly. The experiment was carried out on a Macintosh computer programmed with PsyScope (Cohen, MacWhinney, Flatt, & Provost, 1993).

A learning trial consisted of a 12s observation period followed by a 3s countdown period and the outcome event. The observation periods consisted of a background picture of the launching site accompanied by a 12s outer space sound effect. Candidate causes occurred during the observation period according to the scheduled trial type. The candidate causes were 2s long animations of a green alien popping out of the crater and making a beeping sound. The big alien measured 9 cm, the small alien 4 cm. Figure 3-2 and Figure 3-3 are screenshots from these animations and illustrate the proportions between aliens and rocket. Immediately after each observation phase, a countdown from 5 to 0 was displayed in the top left corner of the screen with 3cm large yellow numbers. Each number flashed for 500ms, and a siren-like sound was played during the 3s countdown phase. During the countdown phase participants had to indicate whether they thought the rocket would launch or not by pressing Y or N on the keyboard, respectively. If participants did not make a prediction within the 3s countdown phase, they were prompted to do so after the countdown reached 0. Once participants made their prediction and the countdown phase had reached its end, the outcome event was displayed. The outcome consisted either of a 3s animation showing the rocket's lift-off superimposed on the background picture or an unaltered 3s display of the background picture with the rocket stationary on the launching pad. The lift-off sequence was accompanied by a launching sound, and the no-launch display was accompanied by a sad sounding beep.

Figure 3-2. Experiment I: Small alien popping out of the crater.

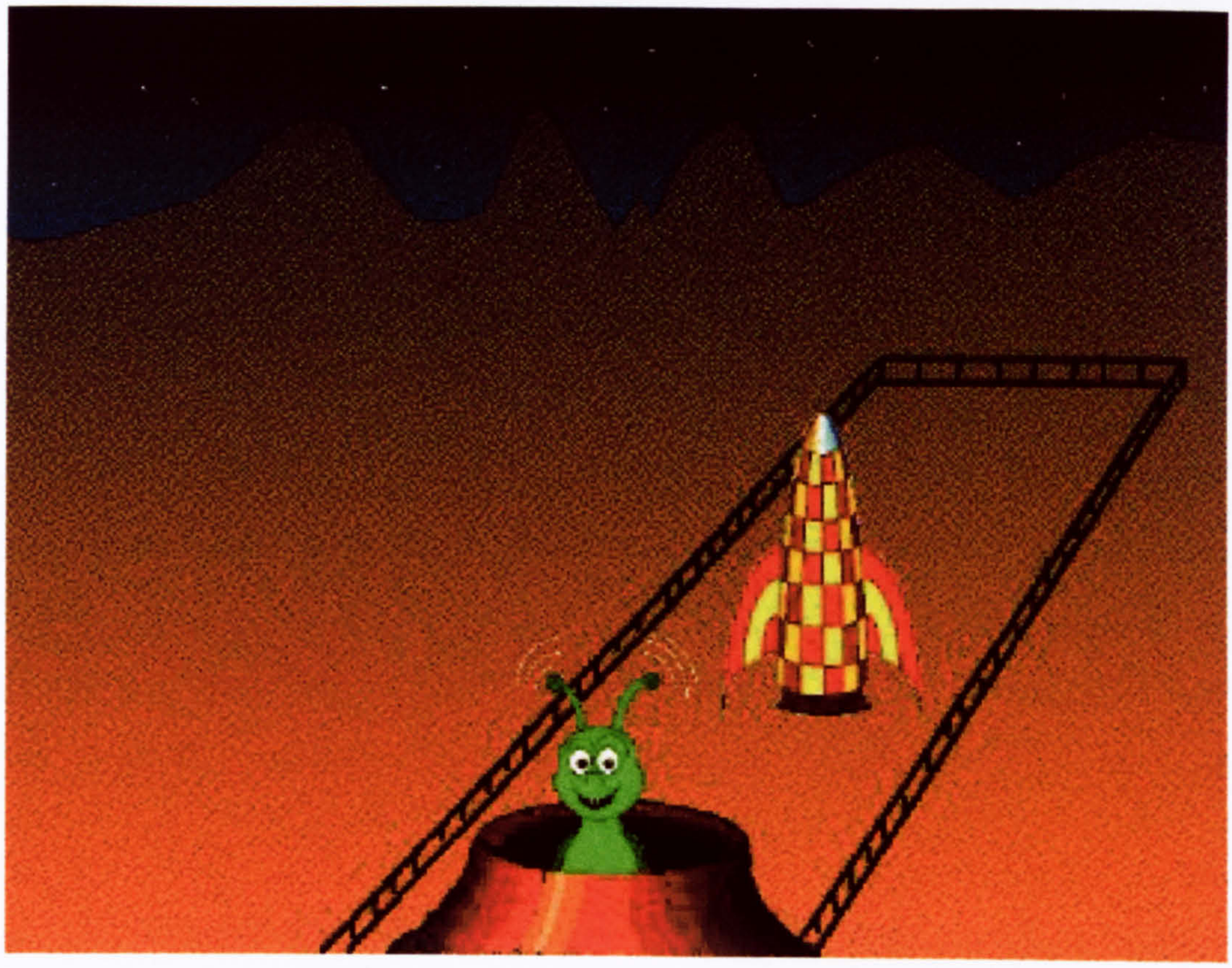


Figure 3-3. Experiment I: Big alien popping out of the crater.

The 50 trial learning-phase was followed by a test phase. Participants were instructed that they would observe the launching site again but that now they would not be able to observe whether the rocket launches or not. Instead, they were asked to predict for each observation period, how likely it was that the rocket launches. There were a total of 4 test trials, one for each combination of candidate cause (c1, c2) and time (early, late). The test trials consisted of an observation period as described above, followed by a judgment prompt. For each test trial, participants were asked to imagine seeing 100 episodes like the one they've just seen and to indicate how many of these 100 episodes they thought would be followed by a rocket launch.

3.2. Results

Participants evaluated c1 as a strong cause if it occurred late in the episode (M=76.42, STD=22.60), but as a weak cause if it occurred early (M=47.46, STD=31.47); participants evaluated c2 in exactly the opposite direction with respect to time, c2 received low causal rating when it occurred late (M=49.79, STD= 32.73), but high ratings when it occurred early in the episode (M=72.21, STD=25.27). Mean causal ratings for c1 and c2 were subjected to a 2x2 repeated measures ANOVA with the factors *candidate cause* (c1, c2) and *time* (early, late); the significance level was set to .05. A preliminary analysis revealed that the counterbalanced assignment of the two candidate causes to the two alien sizes (big, small) did not produce any significant effects or interactions, so the main analysis collapsed over this factor. As expected, neither *candidate cause* nor *time* produced any significant main effects on their own, but there was a highly significant *candidate cause x time* interaction, $F(1,23)$=10.40.

Figure 3-4. Mean ratings of causal effectiveness of c1 and c2 in Experiment I, depending on their temporal position with respect to the effect. Error bars denote standard errors.

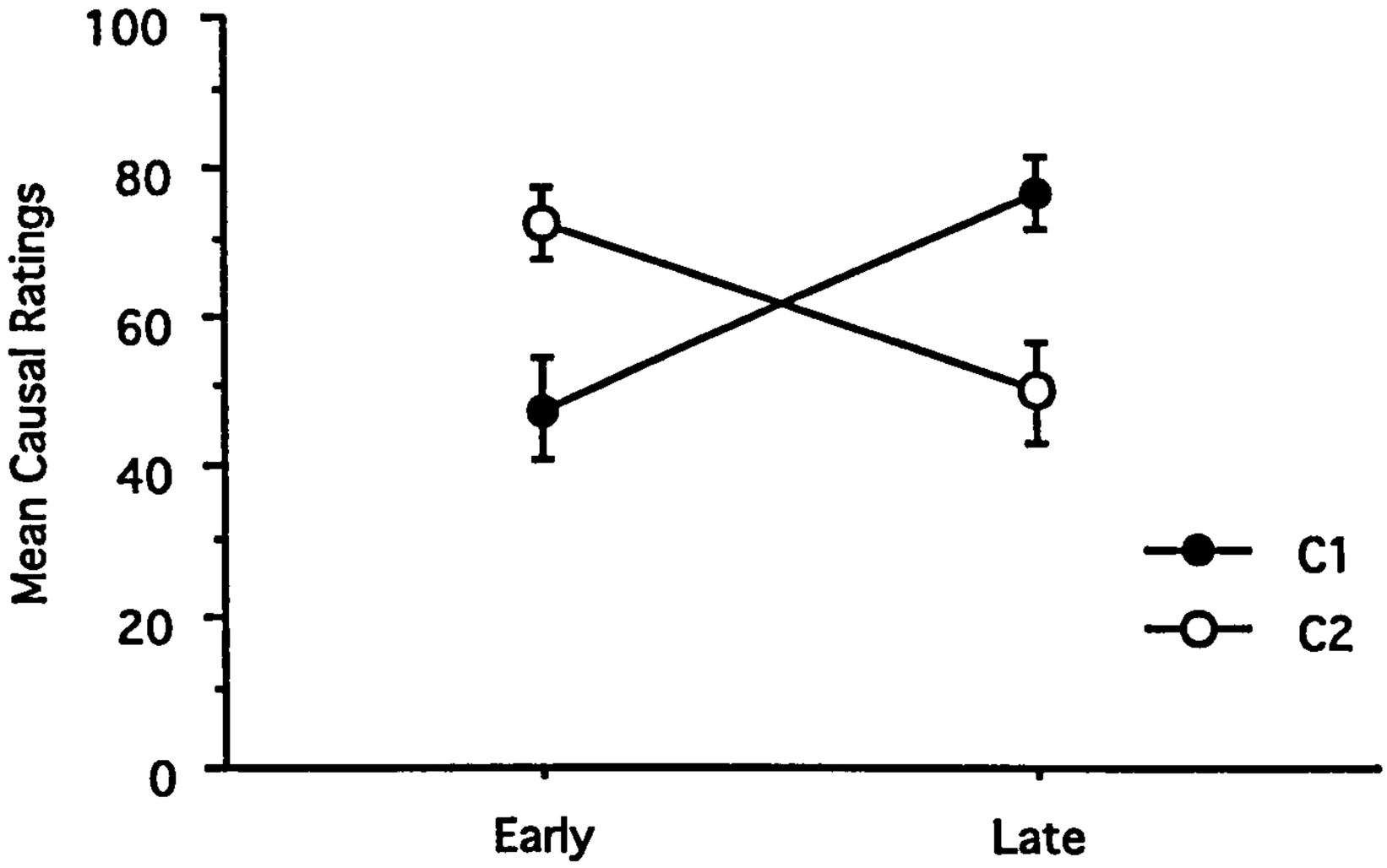


Figure 3-4 illustrates the nature of the interaction. These results indicate that participants correctly identified the underlying causal structure implemented in this experiment. In particular, they were sensitive to changes in each candidate's causal strength brought about by the temporal position. Participants could not only learn that the causal effectiveness of c1 decays the further in time it moved away from the effect, but they also learnt that the effectiveness of c2 increased as temporal contiguity with the effect was degraded.

3.3. Discussion

Experiment I demonstrated that people are capable of learning to use temporal information to distinguish between strong and weak causal relations. Information about the causal strength of two candidates to produce an effect

was presented. Causal strengths for each candidate varied as the temporal position of the candidate changed within an observational period. One candidate, c1, was a perfect predictor if it occurred late in the episode, in which case it was closely followed by the effect. If c1 occurred early in the episode (thus temporally separated from the potential occurrence of the outcome), the causal effectiveness dropped to .3. The other candidate, c2, was only a weak predictor if it occurred late in the episode (thus contiguous with the potential delivery of the outcome). If c2 occurred early, however, it always produced the effect later on, so that c2 and the effect were non-contiguous to each other. The true underlying causal structure for c1 thus was congruent with the idea that temporal contiguity is essential for causal induction, while the causal structure for c2 contradicted it. Participants correctly learned the causal structures of both c1 and c2, as indicated by the *candidate cause x time* interaction, and they learnt them both equally well, as indicated by the absence of a main effect of *candidate cause*.

This finding indicates that temporal contiguity is not essential for human causal induction. Participants did not exclusively focus on the temporal relation between candidate causes and effect as a direct indicator of causal strength. Instead, they attended to the true changes in underlying causal strength, and used temporal position flexibly to identify those changes. In other words, they integrated the two Humean (1739/1888) cues to causality, contingency and contiguity, to come to a rational understanding of cause. Experiment I also demonstrated that people understand the importance of time in causal induction. The explanatory (causal) power of the same physical event with respect to a particular outcome can vary depending on when in time (relative to the outcome) the event occurs. As pointed out by Einhorn & Hogarth (1986) temporal contiguity is often necessary to render a particular event a plausible candidate cause for an outcome (as in the launching paradigm, see Michotte, 1946/1963). Sometimes, however, certain events cannot be plausible causes of an outcome if they occurred contiguously with it,

but can only be rendered causal by an intervening delay (as, for example, sexual intercourse as a cause of giving birth, c.f. Einhorn & Hogarth, 1986). In these two preceding examples, reasoners presumably draw on prior knowledge about the timeframe of the relevant causal mechanisms involved, when they evaluate the evidence. Experiment I used a completely novel scenario, however, disallowing the recruitment of prior knowledge to solve the task. Participants nonetheless successfully unravelled the complexities of the employed causal structures, showing that humans can learn the timeframe of causal relations from probabilistic feedback.

3.3.1. Causal Reasoning or Categorization?

An alternative interpretation of these results is to recast the reasoning task in Experiment I as one of *categorization* rather than causal learning. Participants were presented with learning episodes of a fixed duration. Whether or not the effect occurred after the end of each episode depended on what happened when within each episode. A categorization task thus would be to decide whether a particular episode is or is not likely to result in a rocket launch. *What* and *when* of course reflect the two dimensions relevant for solving the task: stimulus type (c1 vs c2, for each counterbalancing group associated with a particular alien size, big or small) and time of occurrence (early, late). Because each dimension took on only two possible values, the dimensions were dichotomous. Furthermore, the two dimensions were completely independent from each other: having a particular value on one dimension, say time:early, did not constrain in any way what the value on the other stimulus dimension (size) would be. Time and Size thus were the two *separable* (Garner, 1974) dimensions relevant for stimulus classification.

Stimuli were structured in such a way that both of the two dimensions were necessary for the classification task, because the classification rule followed a principle of exclusive disjunction. In order to be followed by a

rocket launch every time, the episode needed to include either c1 late in the episode or c2 early in the episode; if c1 occurred early, or c2 late, the episode would be a weak causal episode. Garner (1974) called this a "condensation" task, and showed that people understand that they need to process both dimensions to carry out the classification. Such condensation tasks are usually a little bit harder than tasks that can be solved by attending to one dimension only. In so called "filter" tasks, for example, only one dimension (say, time) would be relevant for the classification, the other dimension (say, size) would be irrelevant; one could decide the causal status of the episode merely by attending to time, with the size of the stimulus having no impact on the classification at all. In "correlation" tasks, the two dimensions would correlate in determining the category structure, i.e. "Small&Early" would need to be discriminated from "Big&Late". "Small&Late" and "Big&Early" would not enter such a correlation task, as they never include the full set of combinations (in which case it would be impossible to form a correlational structure). Participants can solve correlation tasks by attending to one dimension only, but which one they attend to does not matter, either size or time both carry the distinctive information.

The *candidate cause x time* interaction shows, however, that participants separated the two dimensions, and realized that they need to process them both in order to make correct classifications. The particular nature of the interaction and the absence of main effects for *time* or *candidate cause* furthermore show that participants grasped the exclusive disjunction between the two dimensions. Participants correctly classified the four possible event types (Big&Early, Big&Late, Small&Early, Small&Late) into strong and weak causal episodes.

3.3.2. Going beyond categorization

One limiting factor of Experiment I is that it was based on discrete learning trials. Within each episode it was perfectly clear whether, when, and what stimulus occurred. This discrete trial structure is of course what made it possible to recast Experiment I as a categorization task: discrete events had to be classified according to an underlying causal structure into those that were highly likely, and those that were less likely to be followed by the effect. Causal induction problems outside experimental psychology laboratories rarely comprise such conveniently marked event boundaries, however. Instead, reasoners have to apply their notions about the timeframe of the causal relation in question to a continuous stream of events: first, to be able to decide whether a given event coincided with an outcome at all, or whether the temporal distance was so great that assuming co-occurrence would not be warranted (e.g. having eaten undercooked chicken three weeks ago does not covary with the symptoms of a food poisoning today); second, to decide what the implications of the temporal relation between event and outcome are (having eaten undercooked chicken is more likely to be the cause of food poisoning, if one has eaten the chicken 12 hours ago than if one has eaten it 1 hour ago, because symptoms of Salmonella infection usually develop between 12-24 hours after ingestion).

It thus appears sensible to shift the focus of investigation away from studies based on discrete learning trials, and employ a continuous paradigm instead. In such a paradigm the participant is exposed to a continuous stream of events. Causal episodes take place within the continuous flow of time, and are not specially marked. It is thus the task of the observer/reasoner to decide how to segment the event stream, and, more importantly for my purposes here, to apply assumptions about the timeframe of the causal relation in question onto the continuous stream of events. In other words the reasoner has to decide how much time can plausibly pass between an occurrence of a

candidate cause and an effect, so that the co-occurrence is interpreted as evidence for the causal relation in question. Beyond that limit, an occurrence of the effect would be attributed to alternative causes, and thus interpreted as evidence against the causal relation in question.

Several studies employing such a continuous paradigm have been reported, most notably from David Shanks and colleagues (e.g. Shanks & Dickinson, 1987; Shanks et al., 1989). Apart from employing a continuous paradigm, these studies differed from Experiment I in another important way. All these previous studies involved constant identical causal structures, and only varied the temporal distance between cause and effect. Contrary to the paradigm I employed in Experiment I, temporal position in these studies thus did not signal a change in true causal effectiveness. Rather, the causal action (a button press) had the same programmed effectiveness in all conditions, and the only thing that changed was the delay after which the delivery of the outcome took place. Shanks et al. and several replications uniformly reported that cause-effect delays of more than two seconds induced a failure in participants to recognize a causal relation. The causal relations Shanks et al. employed were based on moderately strong contingencies (.75), but if cause and effect were separated by more than two seconds, participants reported them to be equivalent to non-contingent control groups.

Experiment I demonstrated that people are able to learn that the extent of temporal separation between cause and effect may indicate decreases (and increases) in causal effectiveness. The true underlying causal (or probabilistic) structure did in fact change in Experiment I, and the pattern of change was determined by both stimulus dimensions (time and size), according to a principle of exclusive disjunction. Participants' sensitivity to the temporal dimension therefore was only rational. In Shanks et al.'s experiments (1989), however, changes in the extent of temporal separation between cause and effect were not accompanied by changes in the underlying causal structure; the contingency between cause and effect remained constant at .75, irrespective of

the degree of cause-effect contiguity. Shanks et al.'s participants nonetheless reported weaker causal relations in the delayed than in the immediate conditions. This finding suggests two interpretations.

1.) Particpants in Shanks et al.'s study were irrational. Although Experiment I has shown that contiguity was not essential for successful causal induction, this finding was based on a discrete trial structure. Shanks et al.'s study, on the other hand employed a continuous paradigm. It may well be that the identification of a delayed causal relationship is much harder in such a continuous paradigm than it was in the trial-based study in Experiment I. If temporal contiguity is indeed essential for causal induction (from continuous event streams), then reasoners should always attend to the temporal dimension, even in situations where it has no relevance for the causal relation in question. Postulating a rigid fixation to a particular stimulus dimension (time), regardless of its informative value is clearly at odds with a rational perspective of human reasoning.

2.) Particpants in Shanks et al.'s study have brought to the task their own prior beliefs about the causal relation in question. Shanks et al. openly admitted that "subjects in judgment studies such as ours assume that the word 'causes' in the experimental instructions means 'causes immediately'. After all, they presumably have considerable experience of the immediacy of cause-effect relations in such electrical devices as computers" (Shanks et al., 1989 p. 155). Participants' failure to correctly evaluate delayed causal relations thus could simply reflect a mismatch between their expectations about the task, and the feedback they received during the task. The available data do not allow one to favor one explanation over the other.

The remainder of this thesis will be dedicated to empirical investigations that do allow a contrast between these two interpretations. The strategy I employed was to replicate Shanks et al.'s experiment as closely as possible, with one important modification: I manipulated participants' assumptions about the timeframe of the causal relation in question.

Experiments II and III explicitly controlled those assumptions by informing one group of participants that the causal relation may involve a delay, while the other group of participants was not informed about delays. Experiments IV through VI employed implicit manipulations; cover stories induced participants to either assume an immediate or delayed relation.

4. Experiments II and III: Explicit Manipulation of Timeframe Assumptions

4.1. Experiment II

Experiment II was closely modelled after Shanks et al.'s (1989) original paradigm. Participants were instructed that their task was to find out how strongly pressing the SPACE bar caused a triangle to light up on the computer screen. As in the original experiment, I programmed the apparatus to produce an effect with a probability of .75 if the participant pressed SPACE. In earlier studies (e.g. Shanks et al., 1989, Experiments 1 and 2) such probabilities were defined relative to experimenter-defined learning trials. This usually means that only the first response within a specified time-bin (e.g. 1s) is recorded and subjected to the reinforcement schedule. Employing such arbitrary learning trials unnecessarily reduces the ecological validity of the procedure, as the participants are usually not informed about the restrictions of the underlying trial structure. It potentially also results in a discrepancy between the programmed and the objectively achieved cause-effect contingency. If participants ever respond at a rate that is higher than the frequency of trial-spaces (e.g. more than once a second), a substantial proportion of their responses would not be subjected to the reinforcement schedule and thus be effectively unreinforced. Experiment II therefore did not involve any pre-defined learning trials, but employed a truly continuous paradigm instead.

I employed two experimental (master) conditions within-subjects, one involving immediate cause-effect pairings (contiguous condition), and one involving cause-effect pairings separated by a 4s delay (delay condition). To check whether participants could distinguish between causal and non-causal conditions, I furthermore employed two yoked control conditions. In these, the apparatus played back the outcome pattern participants had generated on

the previous experimental condition, and participants' actions had no consequences at all. The crucial modification from Shanks et al.'s (1989) earlier procedure was that I manipulated participants' expectations about the timeframe of the causal relation in question between subjects. One group of participants was told that sometimes the triangle would light up only after a certain delay (Instruction group), while the other group of participants received no such instructions (No instructions group).

4.1.1. Method

4.1.1.1. Participants

Fifty-one volunteers (40 females, 11 males) were recruited via a departmental notice board and through the University of Sheffield's volunteer email distribution list. Some participated to fulfil part of a course requirement, others received a small nominal payment. One participant failed to comply with the instructions and was dropped from the analyses.

4.1.1.2. Design

In a mixed design the factor *instruction* about delay (Instruction/No Instruction) was manipulated between subjects, and the factor *time* (contiguous/delay) within subjects. Each participant worked on two blocks, each consisting of one of the two experimental (master) conditions (contiguous or delay) followed by a corresponding control (yoked) condition. In both experimental conditions the probability that a key press would produce the effect P(e|c) was set to .75. This probability was not defined per unit of time (i.e. relative to a specific time-bin); in other words the schedule did not employ learning trials of any kind. If a response resulted in an effect, it followed it instantly in the contiguous condition, and after a 4 second delay in the delay condition. Responses made during the delay period and presentation of the

effect were recorded but had no programmed consequences. More specifically, if participants pressed the SPACE bar several times in a row, each response triggered the probability generator until it first scheduled an outcome. Any subsequent responses between the first successful key press and the occurrence of the outcome had no chance to produce further outcomes, but were still recorded as unreinforced responses.

In the experimental conditions the effect never occurred in the absence of a response, i.e. $P(e|\neg c)$ was set to 0. In the yoked control conditions the effect occurred totally independently of the participants' behaviour: the apparatus played back the outcome pattern that participants had produced in the preceding experimental condition. Responses made during the yoked conditions were recorded but had no programmed consequences. The order of conditions was constrained by the blocks so that yoked control conditions immediately followed their corresponding experimental master conditions. Whether a participant worked on the contiguous or delay block first was counterbalanced between subjects. The two blocks were only used as a means to control the administration of the master/yoke pairs. Participants saw and worked on four conditions, but were unaware that the conditions were organized in two blocks.

4.1.1.3. Materials, Procedure, and Apparatus.

The experiment was carried out on Apple Macintosh computers located in separate cubicles and programmed in Macromedia Director 7.0. Participants used the keyboard to make responses and enter causal ratings.

After the experimenter started the computer program, participants read the following instructions on the screen (modified from Shanks et al., 1989)

> Please read the following instructions very carefully. Take as much time as you like. Your task in this experiment is to judge the extent to

which you can make something happen on the computer screen. There will be a triangle on the screen like this:

The outline of an equilateral triangle with sides 10 cm long was presented on the screen below the text. The instructions continued:

Now press the SPACE bar and see what happens…

When the participant pressed the SPACE bar, the triangle lit up in red for 500ms, accompanied by a beep. The red triangle and beep constituted the effect. After the effect had occurred, the triangle reverted to its original outline state and the instructions continued:

… and press it a few more times …

The participant had to press the SPACE bar and observe the triangle flash for another four times. The instructions then continued:

… the triangle lights up.

Tell the Experimenter when you are ready to proceed

When the participant indicated that she was ready, the experimenter pressed the RETURN key and the instructions continued:

Sometimes the triangle might light up of its own accord, like this:

The triangle flashed four times at 2s intervals, irrespective of any responses. The instructions then continued:

Your task in this experiment is to find out whether pressing the SPACE BAR has any effect on whether or not the triangle lights up. At any time you may choose whether or not to press the SPACE BAR. You can press it as often or as little as you like. However, because of the nature of the task it is to your advantage to press it some of the time and not to press it some of the time.

Tell the Experimenter when you are ready to proceed

When the participant indicated that she was ready, the experimenter pressed the RETURN key and the instructions continued:

> Sometimes the triangle will flash when you press the SPACE BAR, and sometimes it will not.

Participants in the *Instructions present* group then read:

> Sometimes pressing the SPACE BAR will cause the triangle to light up immediately and sometimes it will cause the triangle to light up after a certain delay like this...
>
> (Press the SPACE BAR)

The triangle then flashed 4s after the participant's first response. The instructions continued:

> (... and press it a few more times)

Participants then had to press the SPACE bar to produce a delayed effect four more times. On each occasion, the first recorded keypress launched the delay and produced the effect. Any key presses participants made during the delay had no consequences.

Both groups of participants then read:

> You must judge the extent to which pressing the SPACE BAR is the cause of the triangle lighting up
>
> Tell the Experimenter when you are ready to proceed

When the participant indicated that she was ready, the experimenter pressed the RETURN key and the instructions concluded:

> You will be given four different problems, each lasting two minutes.
>
> The relationship between pressing the SPACE BAR and whether or not the triangle lights up will be constant within each problem but may well differ from one problem to the next.

At the end of each problem you will be asked to give an estimate on a rating scale of the extent to which you think pressing the SPACE BAR caused the triangle to light up during that problem.

If you are ready, then the Experimenter will start the first problem for you

If the participant had no further questions, the experimenter then started the experiment and left the room. Participants then worked on four problems for two minutes each. After each problem, the screen blanked out and the following text appeared:

Type in a number to indicate your judgment of the extent to which pressing the SPACE BAR caused the triangle to light up.

Use a scale from 0 to 100.

100 indicates that pressing the SPACE BAR *always* caused the triangle to light up,

0 indicates that pressing the SPACE BAR *had no effect* on whether or not the triangle lit up.

Click the OK button after typing in your judgment and the experiment will continue.

Your judgment:

After participants had entered their answer on the keyboard and clicked the OK button on the bottom of the screen, they proceeded to the next problem. The experiment lasted about 15 minutes.

4.1.2. Results

4.1.2.1. Behavioural Data

All statistical analyses adopted a significance level of .05, except where otherwise noted. Table 4-1 displays the mean number of responses produced per minute, the number of outcomes per minute, and the obtained probabilities that a response triggered an outcome, P(e|c). In contrast to earlier, learning-trial based, studies (e.g. Shanks et al., 1989; Reed, 1992) I calculated P(e|c) as the actual probability that *any response* – even those made during a delay period – produced an effect. Outcome presentation during the two yoked control conditions was entirely independent of participants' behaviour, so Table 4-1 lists the P(e|c) values for these conditions as 0. However, participants' responses will sometimes by chance be followed by outcomes in yoked conditions, so the actual value of P(e|c) may not accurately reflect participants' subjective perceptions. Table 4-2 therefore additionally reports the average number of responses made before each outcome (counted from the delivery of the preceding outcome). Note that the average value of 1.3 responses before an outcome in the contiguous master condition reflects that on average 3 out of 4 responses (i.e. 75%) were followed by an outcome.

Overall participants responded less in the delay (M=17.98, STD=16.39) compared to the contiguous experimental (master) conditions (M=24.9, STD=18.49); they also responded less when they expected a delayed relation (Instruction group, M=16.62, STD=15.57) than when they did not expect a delay (No Instructions group, M=26.26, STD=18.58). An ANOVA on the response rates revealed a main effect of *Instruction*, $F(1,46)=7.301$, a *Time x Master/Yoke*, $F(1,46)=6.414$, and a *Instruction x Master/Yoke x Order of Problems* $F(1,46)=6.346$ interaction, plus a marginal *Time x Instruction* interaction, $F(1,46)=3.401$, $p=.072$.

Table 4-1. Experiment II. Mean Response Rates[a], Outcome Rates[a], and P(e|c) in the Contiguous and Delay Master and Yoked Conditions for participants informed and uninformed about potential delays who worked on the delayed problems in the first or second block.

		Time			
		Contiguous		Delay	
	Group	Master	Yoked	Master	Yoked
Response Rate	Informed				
	Delay 1st	25.42	8.23	11.65	12.08
	Delay 2nd	18.00	20.25	11.08	14.88
	Uninformed				
	Delay 1st	21.13	23.71	17.42	25.13
	Delay 2nd	34.23	27.81	31.19	37.15
Outcome Rate	Informed				
	Delay 1st	19.27		5.04	
	Delay 2nd	12.63		4.13	
	Uninformed				
	Delay 1st	15.58		5.38	
	Delay 2nd	25.96		6.81	
P(e|c)	Informed				
	Delay 1st	0.78		0.65	
	Delay 2nd	0.72		0.59	
			0.00		0.00
	Uninformed				
	Delay 1st	0.73		0.51	
	Delay 2nd	0.76		0.32	

Note. Outcome Rates are identical for Master and Yoked conditions because the latter played back the outcome pattern generated in the former. P(e|c)=0 for all Yoked conditions because responses made during Yoked conditions were never subjected to the reinforcement schedule. See text for discussion.

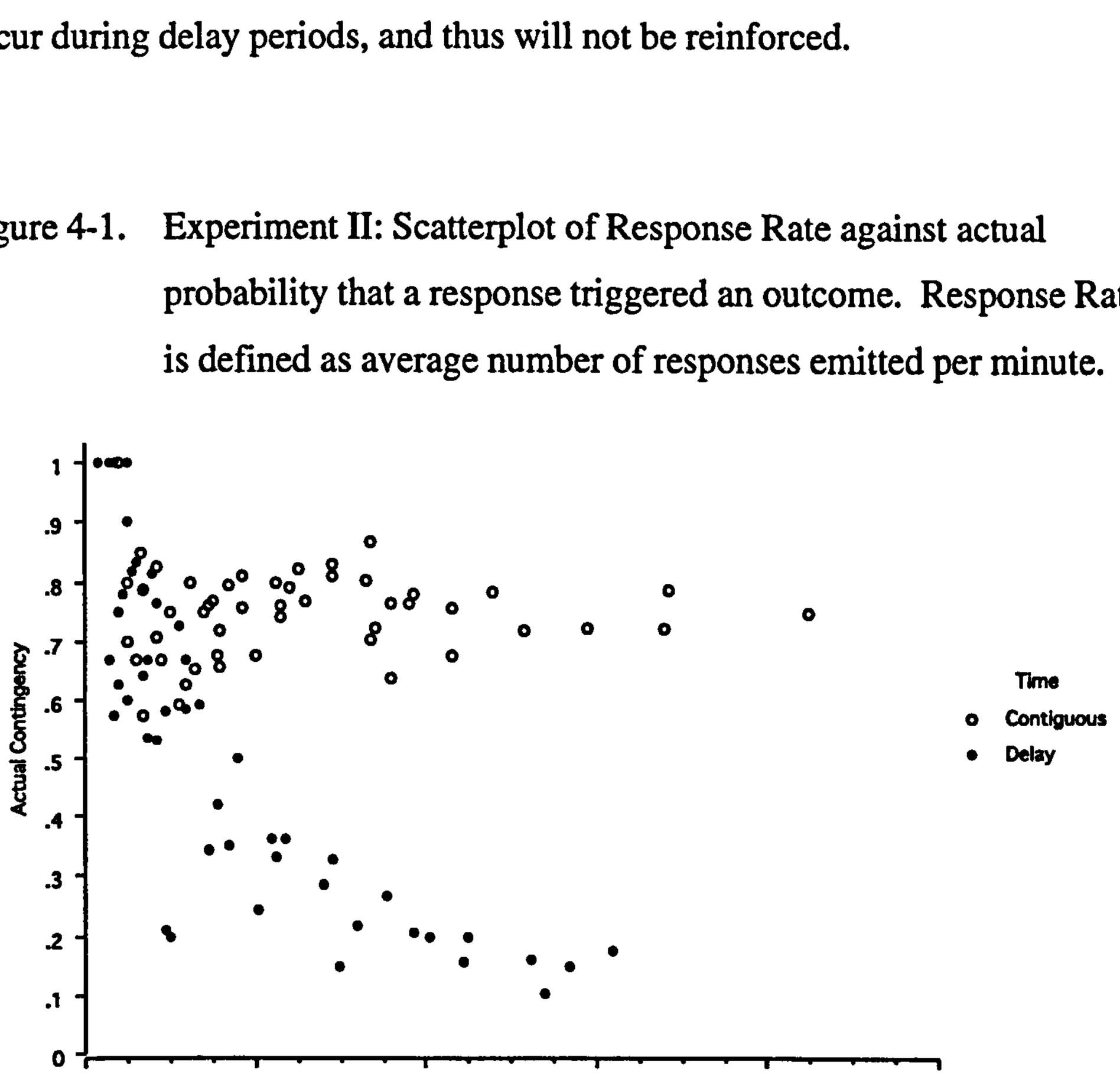
[a]per minute

Table 4-2. Experiment II. Mean Number of Responses emitted before Outcomes, and Mean Proportion of Outcomes which were not preceeded by any response in the Contiguous and Delay Master and Yoked Conditions for participants informed and uninformed about potential delays who worked on the delayed problems in the first or second block

		Time			
		Contiguous		Delay	
	Group	Master	Yoked	Master	Yoked
Responses before Outcome	Informed				
	Delay 1st	1.26	0.53	1.77	1.57
	Delay 2nd	1.39	1.39	2.12	3.07
	Uninformed				
	Delay 1st	1.39	2.55	2.95	4.30
	Delay 2nd	1.32	1.11	4.36	4.85
Proportion of Outcomes not preceded by Responses	Informed				
	Delay 1st	0.00	0.69	0.00	0.34
	Delay 2nd	0.00	0.31	0.00	0.26
	Uninformed				
	Delay 1st	0.00	0.57	0.00	0.24
	Delay 2nd	0.00	0.59	0.00	0.25

An ANOVA of the average number of Outcomes per minute revealed that the delay conditions produced significantly fewer outcomes per minute (M=5.36, STD=2.41) than the contiguous conditions (M=18.53, STD=14.01), $F(1,46)$=54.606 (see Reed, 1992 for a similar finding). As Table 4-1 shows, actual P(e|c) values in the master conditions were lower in all delay conditions (M=.51, STD=.27) compared to contiguous (M=.75, STD=. 08) conditions (cf. Reed, 1992). This finding was particularly strong for the group who did not receive instructions about a potential delay (M=.41, STD=.27), and can be attributed to their generally higher response rate. Higher response rates lowered the actual P(e|c) in the delay but not the contiguous conditions, because a high response rate implies that proportionally more responses will occur during delay periods, and thus will not be reinforced.

Figure 4-1. Experiment II: Scatterplot of Response Rate against actual probability that a response triggered an outcome. Response Rate is defined as average number of responses emitted per minute.



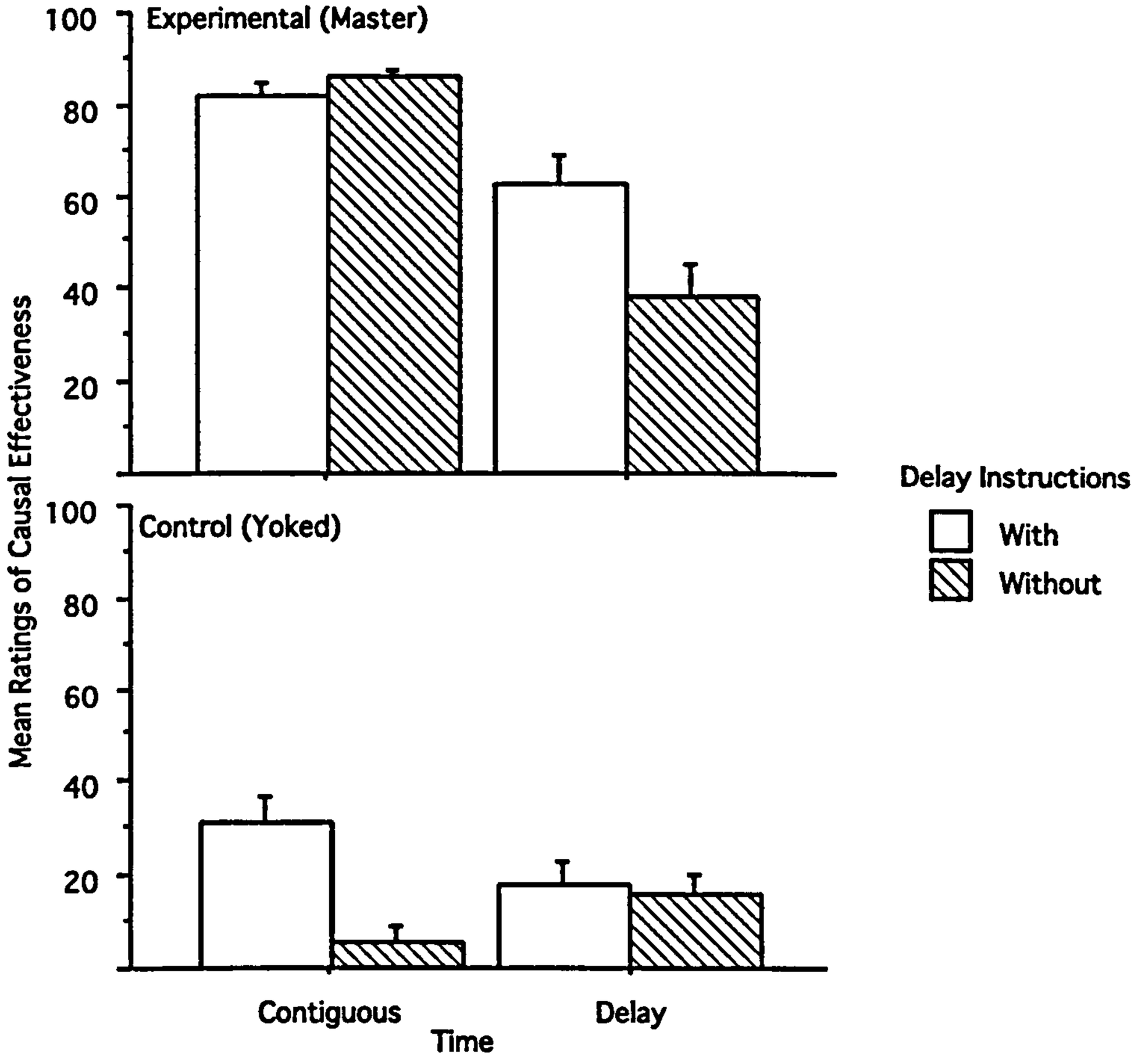
An ANOVA on the P(e|c) values revealed significant main effects of *Time*, $F(1,46)=42.705$, *Instructions*, $F(1,46)= 8.195$, and a *Time x Instructions* interaction, $F(1,46)=7.950$. Figure 4-1 plots actual values of P(e|c) against response rate for each participant in the contiguous and delay experimental conditions. Visual inspection reveals that participants in the contiguous condition consistently experienced contingencies within a relatively narrow range around the programmed value of .75 (+/- .15); in the delay condition, however, participants' experienced contingencies were inversely proportional to their response rate. In other words, the evidence for a causal relationship became weaker as participants attempted to sample more of this evidence.

To summarize, the behavioural data show that a delay of reinforcement resulted in considerably lower actual (experienced) contingencies than the value programmed in the reinforcement schedule. The objective evidence for a causal relation thus was noticeably weaker in the delay than the contiguous conditions. Note that this factual difference is completely independent of any detrimental effects outcome delays may have on the subjective evaluation of causal relations. In addition, if participants were ignorant of the possibility of a delay, they responded at a high rate compared to participants informed about delays, which lowered the contingency even more.

4.1.2.2. Causal Judgments

Figure 4-2 displays mean ratings of causal effectiveness for experimental (Master) and control (Yoked) conditions for participants who did and did not receive delay instructions. Visual inspection reveals that participants could distinguish between contingent (i.e. experimental) and non-contingent (i.e. control) conditions. Paired t-tests between the four Master/Yoke pairs corroborated this observation, all $ts(24)>3.4$. Because my focus is on contingency and delay as determinants of judged causal strength, subsequent analyses will concentrate on the experimental conditions.

Figure 4-2. Experiment II: Mean ratings of causal effectiveness in experimental master (top) and yoked control (bottom) conditions. Error bars indicate standard errors.



In the experimental conditions, a delay in the response-outcome relation generally resulted in lower causal ratings, but to a lesser extent if instructions informed participants about potential delays. An ANOVA of causal ratings in the experimental conditions revealed a highly significant effect of *Time*, $F(1,46)=53.774$, significant effects of *Instructions*, $F(1, 46)=3.900$, and a *Time x Instructions* interaction, $F(1, 46)=8.755$, confirming the qualitative pattern described above. The counterbalancing factor *Order of Conditions* – whether a participant worked on the block containing the contiguous condition first and on the delayed one second, or vice versa – also

produced a main effect, $F(1, 46)=16.536$, and a marginal *Time x Order of Conditions* interaction, $F(1, 46)=3.876$, $p=.055$.

4.1.3. Discussion

Experiment II revealed several interesting new facts about the impact of reinforcement delays in human instrumental causal learning. On the purely behavioural side, three main findings stood out: a) instructing participants about potential outcome delays induces them to respond less often compared to participants ignorant of that possibility, b) irrespective of instructions, outcome delays lead to fewer responses and thus result in fewer outcomes produced per minute (cf. Reed, 1992 who found the same result), and c) employing a reinforcement delay while maintaining the programmed contingency schedule leads to lower actual (experienced) contingencies relative to immediate reinforcement, and hence to weaker objective evidence for the causal relation in question.

Three key aspects characterize participants' ratings of causal effectiveness: a) participants distinguished between contingent and non-contingent conditions, thus demonstrating a genuine capacity to identify when their actions were and were not causally related to the effects; b) when a delay separated actions from outcomes, participants generally rated their actions to be less effective, compared to conditions employing the same programmed contingency with immediate reinforcement; c) this detrimental effect of delay on participants' ratings of effectiveness was less pronounced if participants were instructed beforehand about the possibility of outcome delays. Knowledge thus mediates the influence of delay in human causal induction.

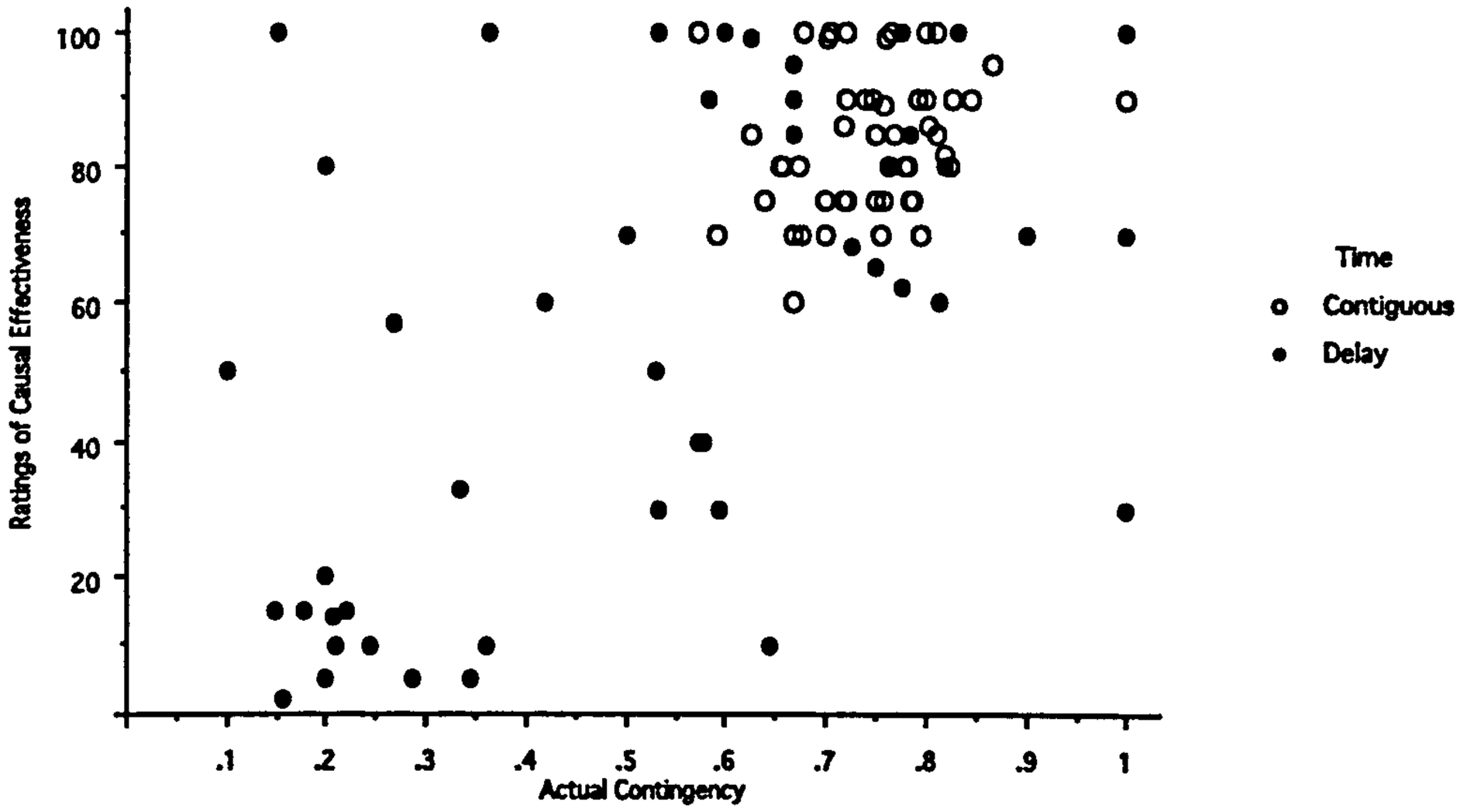
4.1.3.1. Problems with the Free-Operant Procedure

Analysis of the behavioural data has shown that introducing a 4s delay between response and outcome significantly lowered the experienced cause-

effect contingency. It is next to impossible to tell whether participants generally provided lower ratings of causal strength in the delay condition simply because the objective evidence – the cause-effect contingency – was weaker, or because of detrimental effects outcome delays may have on the formation of causal beliefs (cf. Shanks & Dickinson, 1987; Shanks et al., 1989), or a combination of the two. Visual inspection of Figure 4-3 suggests however, that, in general, causal ratings in both the contiguous and delay conditions corresponded well to the actual contingencies sampled by each participant. I used regression to predict causal ratings from actual contingencies, with the intercept forced to be zero. A regression coefficient β of 100 in this analysis would indicate perfect correspondence between contingency and causal judgment, while higher and lower values would indicate over- and under-estimation, respectively. The regression results suggest that actual contingency in fact is a very good predictor of causal ratings in both the contiguous (R^2=.98, β=111.93) and delay (R^2=.72, β=90.47) conditions. Lower average ratings of causal effectiveness in the delay condition thus may indeed reflect participants' sensitivity to the weaker evidence for the causal relation in question.

In order to analyse the influence of delay on causal ratings in addition to, or independent from the difference in contingencies between the immediate and delay group, one may be tempted to perform an Analysis of Covariance on the judgment data, with contingency as the covariate. Shanks et al. (1989) performed such an analysis and reported that delay still had a detrimental effect on causal judgments, even when contingency was taken into account. Unfortunately, this is not a valid analysis, at least for the data from the current experiment, as the covariate – contingency – is completely confounded with one of the independent variables – time. Another possibility is to compare judgments of only those participants who experienced similar contingencies in both the immediate and delay conditions.

Figure 4-3. Experiment II: Ratings of causal effectiveness in the Contiguous and Delay conditions plotted against the actual contingencies experienced by each participant.



A further ANOVA on participants' causal ratings included only participants who experienced a contingency of .57 or higher in both the contiguous and delay condition. I chose .57 as the cut-off point because it was the lowest actual contingency sampled from the contiguous condition. This criterion left me with 36 participants who received delay instructions and only 12 participants who received no delay instructions. No main effects or interactions were significant, most notably no effect of *Time*, $F(1,20) = 0.721$. This finding supports the idea that action-outcome delays may have no direct detrimental effects on human causal judgments other than lowering the objective evidence (contingency) for the causal relation in question.

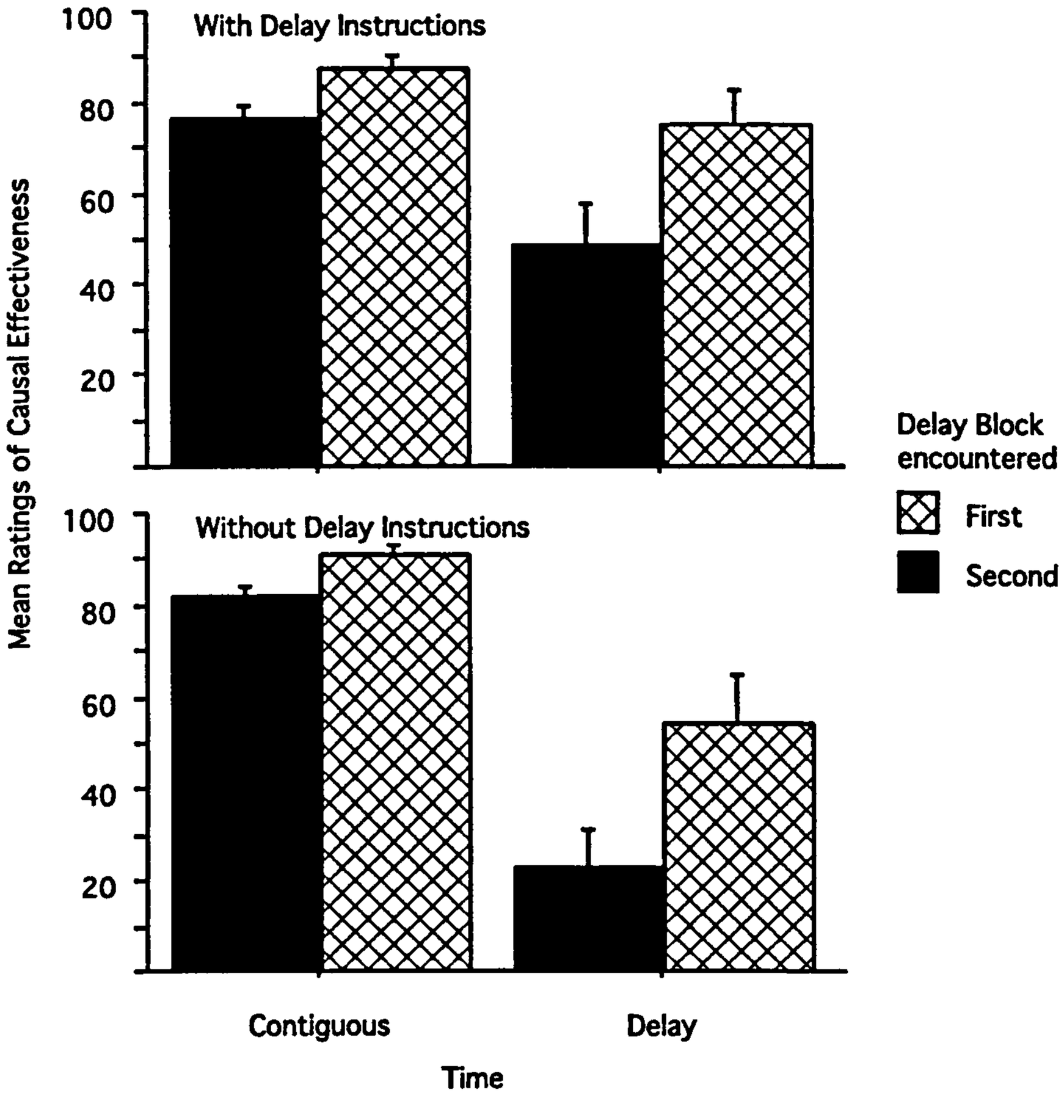
4.1.3.2. Order effects: The impact of previously experienced contiguity

One finding from the ANOVA of causal ratings in the contingent condition came as a surprise: the counterbalancing factor *Order* – whether participants received the contiguous problem first and the delayed one last, or vice versa – produced a significant main effect. This finding warrants closer examination. Comparable previous studies (Shanks et al., 1989; Reed, 1992) either did not control for the order of conditions, or reported analyses collapsed over this factor. Figure 4-4 illustrates the order effect on causal ratings for the contiguous and delayed experimental conditions. Visual inspection reveals that the counterbalancing order had a particularly strong effect on participants' evaluation of the delayed contingencies: participants gave much higher ratings to delayed contingencies if they had not already experienced the contiguous problems than if they had. A linear contrast comparing causal ratings in the delay conditions from participants who first worked on the immediate conditions to those who worked on it last revealed a significant difference between the two groups, $F(1,46)=19.029$. More importantly, a comparison of the two hatched bars in the top panel of Figure 4-4 shows that causal ratings from participants in the instruction group did not seem to be significantly reduced, by a 4s delay, if the delayed problem was experienced as the first problem (*M* for *Time:0s* =87.77, *M* for *Time:4s* = 74.70), $t(24)=1.515$, *n.s.* This non-significant finding of course means a failure to reject the null hypothesis, and one has to be careful when interpreting such a result. A good way to measure the interpretability of the result is to compute the power of the test. I performed a power analysis as described in Howell (1997) on this particular comparison. I set the hypothesized difference between the two means for *Time:0s* and *Time:4s* under $H_1$to be 30. This value was based on an overall difference of 34 between causal ratings derived from the 0s and 4s conditions, collapsed over the factors *Instructions* and *Order*. The effect size *d* is then computed

according to the formula $d=\mu_1-\mu_2/\sigma_{X1-X2}$ where $\mu_1-\mu_2$ is the hypothesized difference and σ_{X1-X2} is the standard deviation of the difference scores. For this test, the effect size was d=.96, which in turn gives a power value of .94 for N=13 (the number of participants who received delay instructions and first worked on the delayed problems). In other words, the probability to correctly reject the null hypothesis (i.e. to discover a difference, if there really is one) is 94%. With such a high power, one is in a reasonable position to interpret a null finding to indicate the absence of an effect.

This order effect means that a traditionally held (single) determinant of judged causal strength – associative strength – (c.f., e.g.Shanks & Dickinson, 1987) falls short for two reasons. Causal judgments derived from identical, delayed, cause-effect contingencies were more accurate when participants a) thought a cause-effect delay plausible compared to when they had no reason for such expectations and b) when they were first confronted with the delayed problem compared to first working on a contiguous problem and seeing the delayed one next. This means that both prior knowledge and experience influence the way people reason about causal relations. Other related work (Buehner & Hagmayer, 2001) has shown that people's inferences from delayed causal episodes depend dramatically on whether participants have previously encountered delayed or contiguous action sequences. In Buehner & Hagmayer's experiment participants interpreted delayed event sequences to indicate a generative causal relation if they previously experienced similarly delayed episodes; however, they interpreted the same delayed event sequence to indicate a preventive causal relation, or no relation at all, if they previously experienced contiguous episodes. Although associative learning theory in general (see Shanks & Dickinson, 1987) predicts that delays result in weaker judgments of causal strength, it cannot capture the mediating effects that prior knowledge and experience have on the impact of delay, unless one could re-interpret them as pre-existing weights or associative strengths. I will return to this question in the General Discussion.

Figure 4-4. Experiment II: The effect of experienced contiguity. Bars represent mean ratings of causal effectiveness in Contiguous and Delay conditions for participants who did (top) and did not receive instructions about potential delays (bottom), split by whether they experienced the Delay block first and the Contiguous block second, or the Delay block second and the Contiguous block first. Only experimental (master) conditions are displayed. Error bars indicate standard errors.



4.2. Experiment III

Experiment II revealed that whether and how strongly cause-effect delays impair causal judgments crucially depends on both prior knowledge and experience. Experiment II also demonstrated that the detrimental effect of a 4s delay may disappear completely under optimal conditions. Analysis of the contingency structures in Experiment II revealed that a 4 second delay significantly lowered the actual contingencies experienced by the participants, even though the delay and contiguous conditions shared the same programmed contingency (.75). The standard Free-Operant procedure employed in Experiment II and earlier studies (e.g. Shanks et al., 1989, Experiment 1; Reed, 1992, Experiments 1 and 2) thus precludes an unambiguous assessment of direct effects of action-outcome delays, unconfounded by weaker objective evidence. The aim of Experiment III is to allow such an assessment. I modified the Free-Operant procedure so that *every* response, including those made during delay periods, was subjected to the reinforcement schedule. As in Experiment II, there were no arbitrarily defined learning-trials.

My hypothesis is that ensuring equally strong evidence for the causal relation in the contiguous and delay conditions will generally reduce the detrimental effect of delay. Furthermore, results from such an optimised paradigm should provide even stronger support for the knowledge-mediation hypothesis. In particular, I predict that participants in the instruction group will no longer judge the causal effectiveness of pressing SPACE to be weaker in the delay than in the contiguous condition. I also expect that all participants, regardless of whether they were instructed about potential delays or not, will evaluate the delay condition in Experiment 3 as more causal than in Experiment 2.

4.2.1. Method

4.2.1.1. Participants

Fifty volunteers (42 female, 8 male) were recruited via a departmental notice board and through the University of Sheffield's volunteer email distribution list. Some participated to fulfil part of a course requirement, others received a small nominal payment. Three participants failed to comply with the instructions and were dropped from the analyses.

4.2.1.2. Materials, Procedure, Apparatus, and Design

The apparatus was programmed to subject *every* response to the reinforcement schedule, including those made during delay periods. All other aspects of materials, procedure, apparatus, and design were identical to Experiment II.

4.2.2. Results

4.2.2.1. Behavioural Data

Table 4-3 displays the mean numbers of responses produced per minute, outcomes per minute, and the mean obtained probabilities that a response triggered an outcome P(e|c), as defined above. Table 4-4 additionally lists the average number of responses made before each outcome.

Table 4-3. Experiment III. Mean Response Rates[a], Outcome Rates[a], and P(e|c) in the Contiguous and Delay Master and Yoked Conditions for Participants informed and uninformed about potential delays who worked on the delayed problems in the first or second block

		Time			
		Contiguous		Delay	
	Group	Master	Yoked	Master	Yoked
Response Rate	Informed				
	Delay 1st	41.05	19.00	12.23	18.64
	Delay 2nd	23.25	11.21	11.67	12.46
	Uninformed				
	Delay 1st	43.42	20.63	17.58	26.33
	Delay 2nd	24.67	27.67	16.00	11.08
Outcome Rate	Informed				
	Delay 1st	30.64		8.82	
	Delay 2nd	17.08		8.63	
	Uninformed				
	Delay 1st	31.79		12.88	
	Delay 2nd	18.13		11.75	
P(e|c)	Informed				
	Delay 1st	0.71		0.73	
	Delay 2nd	0.74	0.00	0.75	0.00
	Uninformed				
	Delay 1st	0.74		0.77	
	Delay 2nd	0.75		0.79	

[a]per minute

Table 4-4. Experiment III. Mean Number of Responses emitted before Outcomes, and Mean Proportion of Outcomes which were not preceeded by any response in the Contiguous and Delay Master and Yoked Conditions for Participants informed and uninformed about potential delays who who worked on the delayed problems in the first or second block

		Time			
		Contiguous		Delay	
Data	Group	Master	Yoked	Master	Yoked
Responses before Outcome	Informed				
	Delay 1st	1.48	1.00	1.42	3.57
	Delay 2nd	1.32	0.93	1.34	1.25
	Uninformed				
	Delay 1st	1.34	0.72	1.30	2.04
	Delay 2nd	1.35	3.11	1.31	1.07
Proportion of Outcomes not preceded by Responses	Informed				
	Delay 1st	0.00	0.61	0.18	0.32
	Delay 2nd	0.00	0.61	0.28	0.40
	Uninformed				
	Delay 1st	0.00	0.64	0.32	0.40
	Delay 2nd	0.00	0.45	0.37	0.57

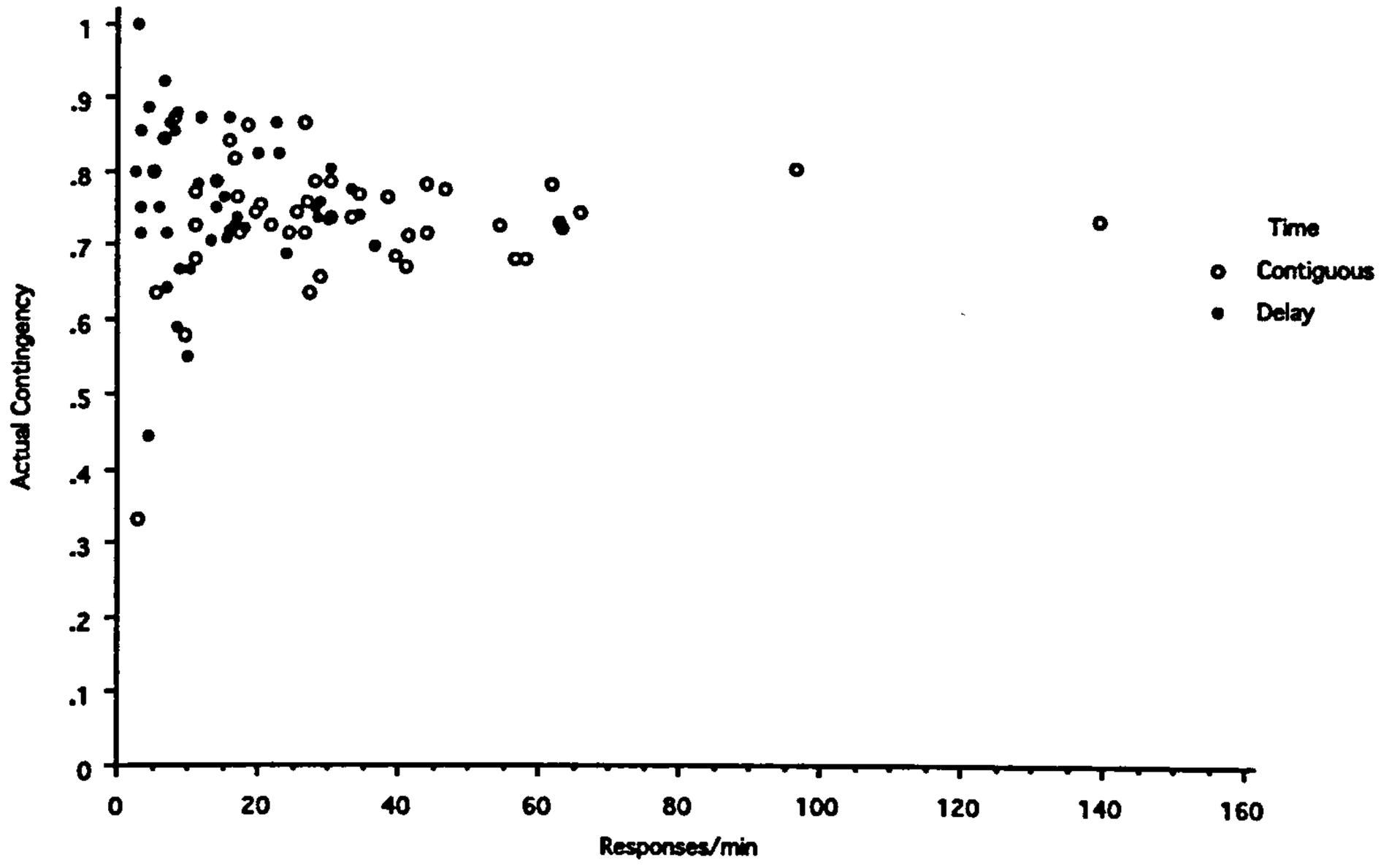
Overall, participants made fewer responses under a delayed (M=14.42, STD=9.50) as compared to a contiguous reinforcement schedule (M=32.93, STD=25.47), and in control (yoked) than experimental (master) conditions. An ANOVA on the response rates revealed significant main effects of *Time*, $F(1,43)=44.09$, *Master/Yoke*, $F(1,43)=9.78$, and of *Order*, $F(1,43)=4.25$, as well as an *Time x Master/Yoke* interaction, $F(1,43)=8.92$, and a *Time x Master/Yoke x Order* interaction, $F(1,43)=6.42$. The latter interaction can be

explained by increased levels of responding in the contiguous experimental problems from the group of participants who first worked on the delayed problems. Unlike Experiment II, whether or not participants were instructed about potential delays did not significantly influence response rates.

As in Experiment II and earlier studies, outcome delays resulted in fewer outcomes per minute. An ANOVA on the average number of outcomes per minute revealed main effects of *Time*, $F(1,43)=29.54$, and *Order*, $F(1,43)=5.12$, as well as a *Time x Order* interaction, $F(1,43)=6.43$. The main effect of *Order* and the *Time x Order* interaction are a consequence of the order effect on response rates described above: higher response rates in the contiguous problems from participants who had already worked on the delayed problems compared to those who had not, resulted in correspondingly more outcomes to be produced per minute.

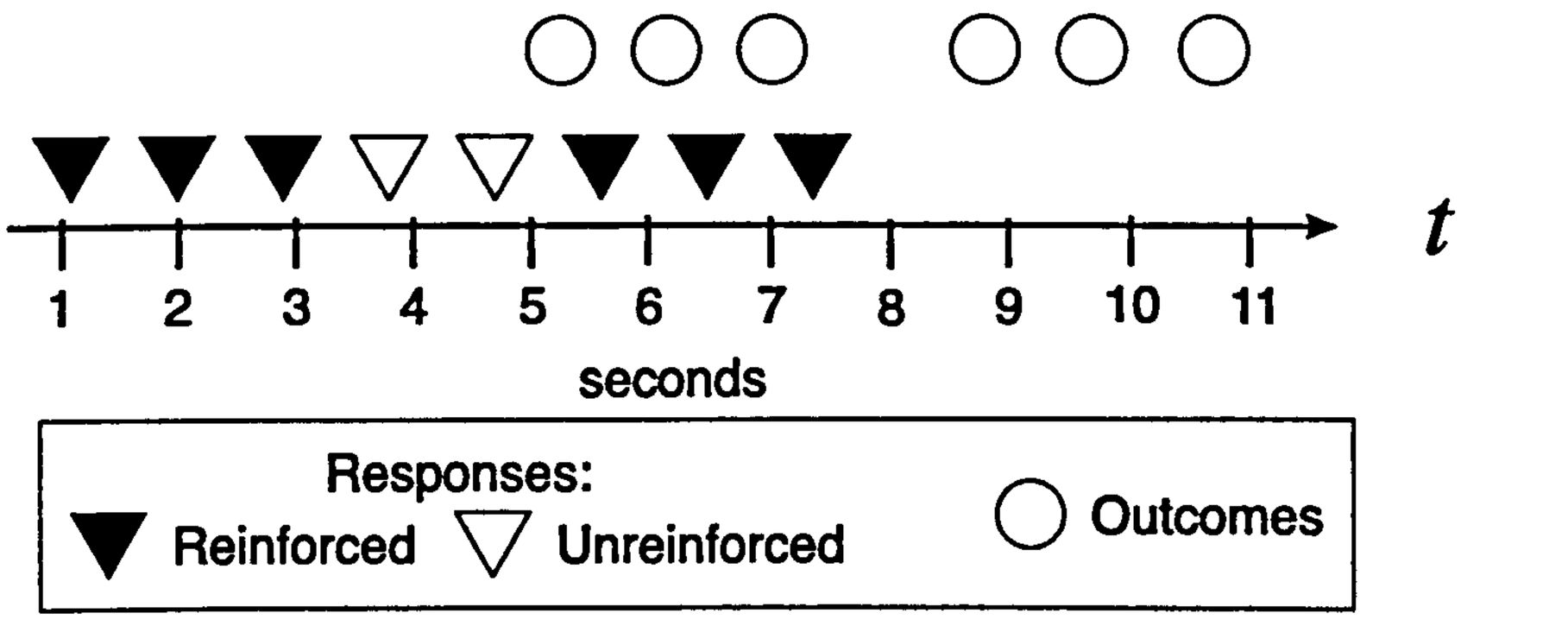
Inspection of the P(e|c) values in Table 4-3 shows that, as intended and contrary to Experiment II, introducing an action-outcome delay did not lower the actual probability that a response produces an outcome. An ANOVA revealed no significant effects, most notably no effect of *Time*, $F(1,43)=2.21$. Visual inspection of Figure 4-5, the analogue of Figure 4-1 from Experiment II, reveals that participants' feedback about the probability that their responses produce outcomes varied in a relatively narrow range around the programmed value .75 in both the contiguous and delay conditions. Unlike Experiment II, this objective evidence for a causal relationship between actions and outcomes did not grow weaker, but instead stabilized as participants sampled more of this evidence.

Figure 4-5. Experiment III: Scatterplot of Response Rate against actual probability that a response triggered an outcome. Response Rate is defined as average number of responses emitted per minute.



To summarize, the behavioural data in Experiment III show that although a delay of reinforcement led participants to respond at a lower rate compared to immediate reinforcement, it did not degrade the experienced probability that responses produce outcomes as it did in Experiment II. This need not mean that participants' experienced *contingency* is unaffected by the action-outcome delay, however. Because every response made in the contingent conditions was subjected to the reinforcement schedule, including those made during delay periods, the resulting action-outcome pattern in the delayed problem may create an (erroneous) perception of $P(e|\neg c) > 0$. Specifically, if participants in the delay condition pressed SPACE several times in a row, each of these responses would have produced with 75% probability an outcome 4 seconds later. The resulting pattern thus could look like the one displayed in Figure 4-6, where some outcomes occur immediately

Figure 4-6. A possible action-outcome pattern from the Delay condition in Experiment III, resulting from the modification employed in the free-operant procedure.



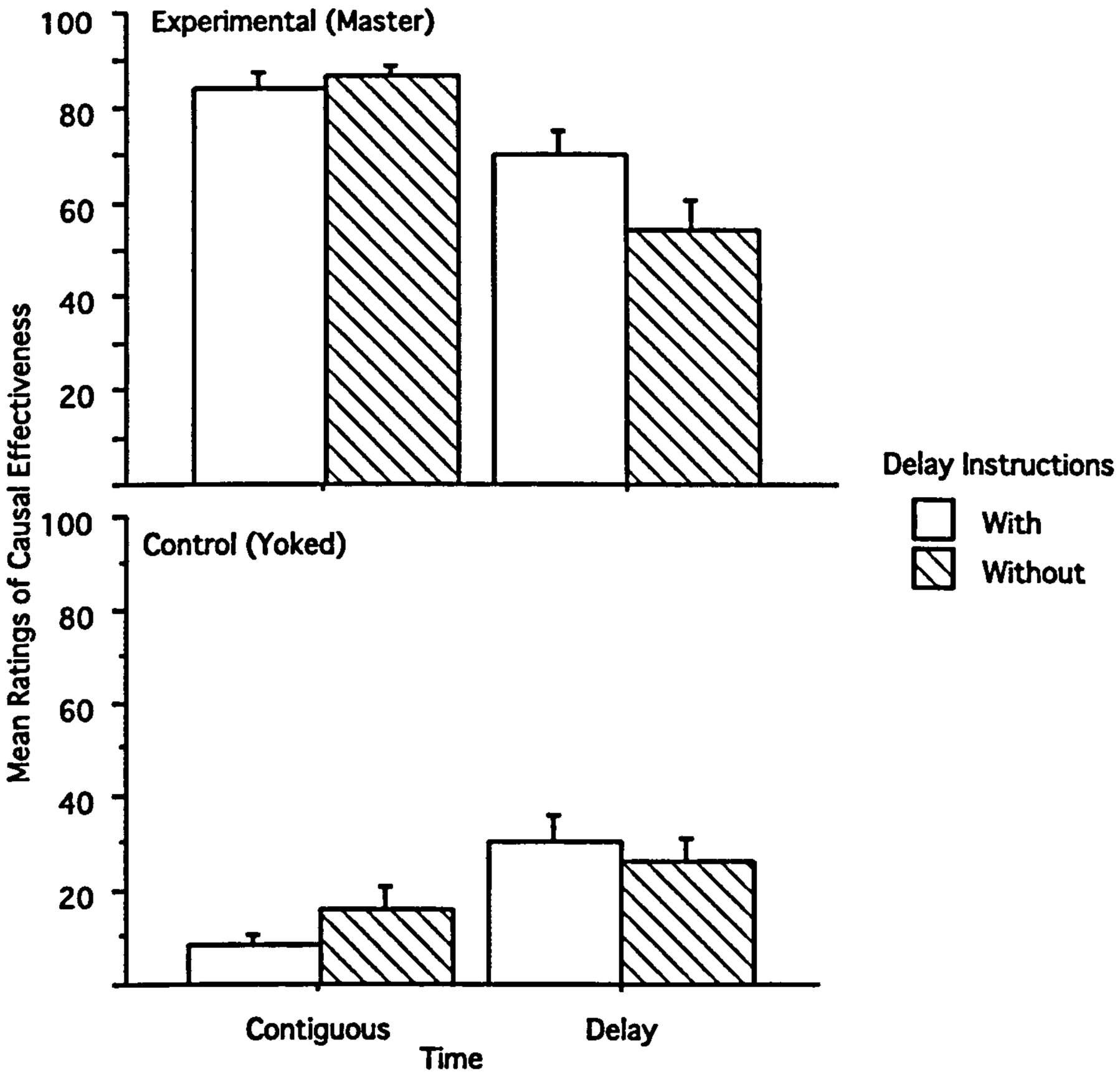
after other outcomes, without any responses apparently preceding them. Consequently, some participants may have had the impression that the effect sometimes occurs "on its own", i.e. $P(e|\neg c) > 0$. Note, however, that delays in Experiment II lowered the *objective* evidence for the causal relation in question.

In Experiment III, in contrast, the objective evidence was constant between immediate and delayed conditions, but delays may still have produced weaker *subjective impressions* of this evidence. It is not possible to measure such apparent values of $P(e|\neg c)$, because the experimental design deliberately avoided pre-defined learning trials. Table 4-2 and Table 4-4 provide, however, the averaged proportion of outcomes that were not directly preceded by any response in Experiments II and III, respectively. While this measure by no means is equivalent to $P(e|\neg c)$, it offers some insight into participants' experienced action-outcome structures. An average of 18% to 37% of outcomes in the experimental delay condition of Experiment III took place without a response occurring since the previous outcome. The strict free-operant procedure employed in Experiment II in contrast, guaranteed that this measure was 0% in both the contiguous and delay conditions.

4.2.2.2. Causal Judgments

Figure 4-7 displays mean ratings of causal effectiveness for experimental (Master) and control (Yoked) conditions for participants who did and did not receive delay instructions. Visual inspection reveals that participants could distinguish between contingent (experimental) and non-contingent (control) conditions. Paired t-tests between the four Master/Yoke pairs corroborated this observation, *t*s(22)>6.2 for the comparisons in the Instruction group, and *t*s(23)>4.6 for the comparisons in the No Instructions

Figure 4-7. Experiment III: Mean ratings of causal effectiveness in experimental master (top) and yoked control (bottom) conditions. Error bars indicate standard errors.



group. Again, subsequent analyses will concentrate on the experimental conditions.

As in Experiment II, an action-outcome delay generally resulted in decreased judgments of causal strength. An ANOVA for causal ratings in the experimental conditions revealed a significant effect of *Time*, $F(1, 43)=26.66$, and a *Time x Instructions* interaction, $F(1, 43)=5.31$. I was mostly interested to see whether *Time* still had a significant effect in the group of participants who received instructions about potential delays. Planned comparisons between ratings in the contiguous and delayed conditions revealed that the effect of *Time* was significant only in the No Instructions group, $t(43)=5.23$; even though causal ratings were likewise higher in the contiguous than in the delayed condition, *Time* failed to produce a significant effect in the Instruction group, $t(43)=2.01$, as I predicted. Again, one has to be cautious about interpreting a null finding, but a power analysis with an effect size $d=.62$ (based on the observed overall difference of 20 between contiguous delayed conditions for participants who received delay instructions) reveals that the t-test has power of .85 with $N=23$ (the number of participants in the Instruction group). This corresponds to the power value of .84 indicated by the statistics package (StatView) for the *Time* x *Instructions* interaction.

As expected, the modification of the Free-Operant procedure overall resulted in a somewhat less pronounced effect of *Time* in Experiment III compared to Experiment II; causal judgments in the delay conditions were higher in Experiment III than in Experiment II (a cross-experimental comparison of ratings in the delay condition fell short of significance, however, $t(96)=1.75$, $p=.08$). Unlike in Experiment II, the order of conditions did not significantly influence causal ratings.

4.2.3. Discussion

Experiment III replicated the most important finding of Experiment II: if participants were aware that outcome delays might occur, they could accommodate them much better than if they were ignorant of potential delays. In order to eliminate the confounding of delay with lower contingencies and thus weaker objective evidence for the causal relations, Experiment III employed a modified reinforcement procedure that ensured that every response was subjected to the reinforcement schedule. As I predicted, this change in procedure improved the accuracy of causal ratings in the delay condition both for participants who were and were not instructed about potential delays. More importantly, causal relations involving a 4s delay were no longer evaluated as significantly less causal than contiguous causal relations, if participants were informed of potential delays.

The order of conditions – whether participants worked on the contiguous or delay block first – had a somewhat different impact on participants' behaviour and judgments in Experiment III than in Experiment II. In Experiment II, order had no systematic influence on participants' patterns of responding and evidence sampling, but had clear effects on participants' causal ratings. When evaluating delayed causal relations, participants in Experiment II were significantly more accurate when they had not already worked on a contiguous problem, than when they had. In fact, the detrimental effect of delay was no longer significant in the group of participants who were aware of potential delays. In Experiment III, order influenced participants' behaviour of responding and evidence sampling, but did not affect the accuracy of causal ratings. Presumably the order effect on causal ratings disappeared because the modified Free-Operant procedure from Experiment III improved accuracy overall, leaving no room for the order effect to manifest itself.

4.3. Discussion and Summary of Experiments II and III

Experiments II and III investigated the roles of temporal contiguity and delay in human causal induction from free operant procedures. In line with earlier studies (e.g. Shanks & Dickinson, 1987; Shanks et al., 1989; Reed, 1992), these experiments showed that cause-effect delays generally lead to a degradation of causal judgments. I investigated the claim that temporal contiguity is essential for human causal induction, and contrasted it with the hypothesis that participants' incapability to correctly identify delayed causal relations may reflect a mismatch between their expectations regarding the timeframe of the causal relations and the actual feedback they received. When interacting with computers, people by default expect their actions to produce immediate results. To control for expectation-experience mismatches when experiencing delays in the computer-controlled paradigm, I manipulated whether participants were made aware of the possibility of delays or not. Instructions either explicitly mentioned that in some conditions the causal action may produce the effect only after a certain delay, or did not mention delays at all. In line with predictions derived from Einhorn & Hogarth (1986), such knowledge about potential delays crucially mediated how participants interpreted delayed causal relations. In both experiments, participants derived higher estimates of causal strength from identical free-operant procedures (implemented with a 4s reinforcement delay), if they were aware of potential delays compared to if they were ignorant of potential delays. Experiment II also showed that Free-operant procedures, often used to study the influence of delay on human causal induction, are actually ill-suited for this purpose, because they confound delay with weaker objective or subjective evidence for the causal relation in question. When I improved the procedure in Experiment III to curtail this problem, participants who were aware of potential delays no longer evaluated delayed causal relations as significantly weaker than contiguous ones. Because this modified Free-Operant procedure was so

successful, the remaining experiments will all employ this improved procedure.

Experiments II and III have shown that reasoners take into account explicit instructions about the timeframe of a causal relation when evaluating evidence sampled from a continuous paradigm. Under optimal conditions these explicit instructions eliminated the detrimental effect of delay in human causal induction, so that causal ratings derived both from immediate as well as delayed cause-effect pairings reflected the underlying contingency alone, unbiased by the degree of contiguity. This result is noteworthy by itself, given the number of reports that have claimed that cause-effect delays always impair causal judgment.

Experiments II and III adapted Shanks et al.'s (1989) method and thus employed very impoverished stimuli. Shanks et al. pointed out that participants in their experiment presumably expected an immediate causal relation. In order to test the hypothesis that expectations about the timeframe guide the parsing of causal episodes it was thus necessary to instruct a sub-group of participants in Experiments II and III *explicitly* about potential delays. However, in everyday causal cognition reasoners evaluate evidence in the absence of explicit instructions about the timeframe. Rather, they recruit existing knowledge about the physical world around them and apply this knowledge when evaluating a particular causal relation. Experiments IV through VI tested whether the findings from Experiments II and III also extend to a more ecologically valid setting. Rather than instructing participants *explicitly* about the timeframe of the causal relation in question, the goal was to create scenarios involving different causal mechanisms that would elicit *implicit* assumptions of either immediacy or delay.

Another limitation on the generalizability of the results from Experiments II and III is that the paradigm they employed only allowed a one-sided test of Einhorn and Hogarth's (1986) knowledge mediation hypothesis. The experimental setup, by default, would have triggered expectations of an

immediate causal relation. Experienced delays between causes and effects, which otherwise resulted in degraded causal judgments, could be tolerated without significant negative effects when participants were aware of the possibility of delays. In other words, knowledge could bridge temporal gaps to the extent that delayed and immediate contingencies gave rise to equivalent impressions of causal strength – the detrimental effect of delay disappeared. Recall, however, that Knowledge Mediation under certain circumstances also predicts a detrimental effect of contiguity (see sections 2.2.3 and 2.3.1.). To test this strong prediction of the Knowledge Mediation hypothesis, it would be essential to create scenarios where reasoners assume that a delay between cause and effect is not only merely *plausible* (as in Experiments II and III), but also *necessary*. The scenarios I picked in Experiments IV through VI were geared to allow such a strong test of Einhorn & Hogarth's Knowledge Mediation hypothesis.

5. Experiments IV through VI: Implicit Manipulation of Timeframe Assumptions

5.1. Questionnaire study preceding Experiments IV through VI

In order to ensure that the scenarios to be employed in the next experiments really do elicit the desired assumptions about the timeframe of the involved causal relation, I decided to first gather data on people's expectancies about the timing of cause and effect in various different scenarios. I created ten scenarios, half of which I thought would imply a contiguous causal relation, and half a delayed causal relation (see Appendix A for full descriptions of each scenario). The ten scenarios were combined in an online-questionnaire that asked participants to simply provide estimates about how much time they thought passes between cause and effect in each scenario.

The scenarios from Appendix A were arranged in five different random orders to create five different questionnaire sheets. Questionnaires were written in the HTML programming language to be viewable with any standard Internet browser. After the description of each scenario (see Appendix A), there was a prompt "Put your answer here. Please specify minutes (m) or seconds (s)" followed by a field where participants could enter their estimates. People were solicited to fill out the questionnaire via a mass-email containing a WWW link sent to staff and postgraduate students in the Department of Psychology, University of Sheffield, and to researchers in the EU T&MR network TACIT. A snowball technique was used, asking each recipient to forward the message to friends and colleagues. The link in the e-mail pointed participants to a page containing an invisible randomiser that determined which of the five different random orders of the ten scenarios was displayed. Participants were asked to imagine each of the ten scenarios and to indicate what length of cause-effect delay they would expect in each scenario. After

they had entered their expectations for all ten scenarios, they had the opportunity to leave their email address and supply comments or queries in case they wished to receive a debriefing message via email.

Table 5-1 lists the mean and median expected delays (in seconds) collected from the first questionnaire (N=54), plus frequency counts for five temporal categories. Analysis of the responses and comments revealed that some participants thought that the "Journalist", "Grenade" "Database", and "Infrared" scenarios were ambiguous with respect to the timeframe they imply. I therefore refined these scenarios (see Appendix B) and collected data from another 38 participants. Table 5-2 lists the outcome of the refined questionnaire.

Table 5-1. Results from first web-based questionnaire, N=54. Time estimates are all displayed in seconds. The top five scenarios were meant to induce expectations of delay, the bottom five expectations of immediacy

	Mean	Median	S.D.	Entries per category				
				<1s	1s	1-5s	5-10s	>10s
Journalist	14.14	2.00	43.92	8	14	12	10	10
Elevator	43.49	30.00	35.34	0	0	0	2	51
Pedestrian	48.06	30.00	33.46	1	0	0	2	51
Grenade	5.36	3.00	5.85	5	6	18	18	6
Database	13.24	6.00	15.02	3	1	10	23	16
Keyboard	0.13	0.00	0.27	50	4	0	0	0
Lightswitch	0.20	0.00	0.39	48	5	1	0	0
Infrared	0.59	0.10	1.44	40	9	3	1	0
Camera	0.32	0.10	0.47	44	6	2	0	0
Doorbell	0.30	0.10	0.39	43	10	0	0	0

Table 5-2. Results from refined web-based questionnaire, N=38. Time estimates are all displayed in seconds.

	Mean	Median	S.D.	Entries per category <1s	1s	1-5s	5-10s	>10s
Journalist	18.69	2.00	53.02	6	10	6	7	7
Elevator	48.14	30.00	44.04	1	0	0	2	33
Pedestrian	48.49	30.00	43.08	0	1	0	0	36
Grenade	15.31	8.00	16.77	2	2	5	11	15
Database	5.32	1.00	11.80	17	3	7	4	6
Keyboard	0.07	0.00	0.20	36	1	0	0	0
Lightswitch	1.88	0.00	9.72	32	4	0	1	1
Infrared	0.53	0.10	0.95	29	5	3	1	0
Camera	0.39	0.10	0.49	28	9	1	0	0
Doorbell	0.36	0.00	0.64	30	6	2	0	0

The questionnaire study confirmed that people do have specific expectations about cause-effect delays and are able to express them. Moreover, the expected distinction between "immediate" and "delayed" scenarios was clearly reflected in participants' response patterns. Expectations in the five contiguous scenarios all clustered around 0 seconds with very little variance; expectations for the delayed scenarios varied considerably, both between and within scenarios. The findings of the questionnaire study were encouraging in that they demonstrated that people are aware of scenario-specific cause-effect delays.

The scenarios to be employed in the next experiment had to fit several constraints. Participants should have the opportunity to gather as much evidence as they wished, so the ideal causal relation would be one without a "refractory period". The "elevator" scenario, for instance, implies a considerable refractory period: once the effect happens (elevator arrives), it would not be possible to observe the effect again for a long time, as the

elevator would stay there until someone else calls it to a different floor. To put it differently, it should be possible and plausible to observe multiple occurrences of the effect in a short time interval. Another obvious constraint for the delayed scenarios was that the expected delay should be noticeably different from zero, but at the same time be not too large, as a very long delay would not be pragmatic in a typical laboratory experiment. After all these considerations, I decided that the (refined) Grenade scenario best fit the requirements for implying a delayed relation. For the contiguous relation I picked the Light switch scenario, because similar scenarios have already been employed in other causal reasoning experiments (e.g.Wasserman et al., 1993).

5.2. Experiment IV

Experiment IV is a more ecologically valid extension from Experiments II and III. Unlike in Experiments II and III, participants in Experiment IV were not explicitly instructed about the timeframe of the causal relation in question. In contrast, two different cover stories served to create two distinct scenarios, aimed at implying expectations about immediate or delayed causal mechanisms. The suitability of the scenarios for this purpose has been determined by the questionnaire study reported in 5. As in Experiments II and III, participants could sample evidence from a free-operant paradigm. There was one constant cause-effect contingency (.75), but the degree of cause-effect contiguity varied from immediate (0s delay) to 2s and 5s delay.

If knowledge and expectations about the timeframe of a causal relation indeed influence how reasoners parse event streams, they should affect the degree of causal beliefs reasoners infer from the evidence. More precisely, identical covariational structures should give rise to different causal beliefs, depending on participants' assumptions about the timeframe of the causal relation in question. If participants think the relation in question implies

immediate cause-effect pairings, episodes containing such immediate pairings should be evaluated causally. If expectations of immediacy are violated by the experience of delayed cause-effect pairings, the same episode should be judged as non-causal, and occurrences of the effect should be attributed to alternative causes other than the candidate in question. On the other hand, if participants think that the relation in question involves a causal mechanism that takes time to unfold, an episode involving delayed cause-effect pairings would be expected to elicit high judgments of causal effectiveness. If, however, such delay assumptions are violated by the experience of contiguous cause-effect pairings, the episode should be evaluated as non-causal, and the covariation should be judged as spurious.

Experiment IV thus allows a two-sided test of the knowledge-mediation hypothesis. Experiments II and III could only test whether the detrimental effect of a cause-effect delay is attenuated once reasoners were instructed that a delay was plausible. The instructions in Experiments II and III merely stated that a delay was possible. It was never mentioned as necessary. The Light bulb and Grenade scenarios employed in Experiment IV, however, are very specific in terms of the time span they imply between cause and effect. Just as immediacy is necessary for the Light bulb scenario to be plausible, a delay is necessary for the Grenade scenario to be believable. A bulb lighting up five seconds after one has flicked a switch, or an explosion several miles away immediately after one has launched a grenade should both fail to create a causal impression. Whereas the time elapsing between cause and effect is too long in the first scenario, it is too short in the second scenario to be deemed plausible.

There are two ways to test knowledge-mediation with respect to the timeframe of causal relations. One way is to compare causal ratings derived from identical covariations manifested in immediate and delayed cause-effect pairings and to check whether evaluations of delayed pairings improve when participants are given a rationale for the delay. This was the strategy followed

in Experiments II and III. Another way is to likewise compare causal ratings in immediate and delayed cause-effect pairings, but this time to check whether evaluations of immediate pairings *decrease* when participants expect a delay. Experiment IV will allow both tests.

Experiment IV adopted a similar methodology as Experiment III. Participants sampled evidence in a Free-Operant procedure, modified as in Experiment III to guarantee stable objective evidence across delays. Expectations about the timeframe of the causal relations were manipulated implicitly via two different cover stories. Unlike in Experiments II and III, however, this manipulation was implemented within participants, i.e. every participant was exposed to each of the two scenarios. Another difference in methodology concerned the control conditions. Experiment IV did not employ yoked control conditions like Experiments II and III. Instead, two control conditions were included where the outcome sometimes occurred on its own, independent of the participant's behaviour. The reason for this change was that Experiment IV included three levels of delay (0s, 2s, and 5s), and including yoked control conditions for each of them would have made the experiment too long.

5.2.1. Method

5.2.1.1. Participants

18 undergraduate students (6 male, 12 female, median age: 19) from the University of Sheffield participated either to fulfill a partial course requirement or to receive a small nominal payment.

5.2.1.2. Design

I combined two thematic scenarios and five levels of causality to produce a 2 x 5 within subject design. *Scenario* had the levels Light bulb and

Grenade. The five levels of causality involved three experimental and two control levels. The three experimental conditions all shared the conditional probability P(e|c)=.75, but had different cause-effect delays (0, 2, and 5 s). In these three experimental levels the outcome never happened unless the participant pressed the button (P(e|¬c)=0). There were no pre-defined learning trials, i.e. any button press triggered the effect with a probability of .75 after the relevant delay. Responses made during a delay period were also recorded and subjected to the reinforcement schedule. Each condition lasted 2 minutes. The two control conditions involved 24 background outcomes, which were not influenced by the participant pressing or not pressing the key. These background outcomes were randomly scheduled within each 2-minute condition with the restrictions that (a) one background event occurred in each of 24 5-second intervals and (b) background events were separated by at least 500ms. One control level employed P(e|c)=0 (so that participants' actions were ineffective), the other P(e|c)=.75 with an action-outcome delay of 0 seconds. I included the control conditions to check whether participants can distinguish between causal and non-causal conditions, and to provide them with experience of conditions where the effect indeed sometimes occurred on its own, as stated in the instructions (see below). In each scenario (Light bulb and Grenade) there were thus three experimental conditions with P(e|c)=.75 and 0, 2, or 5s delays, and two control conditions, one with P(e|c)=.75 and 0s delay, but an additional 24 background events, and one with P(e|c)=0 and 24 background events.

For the four levels with a programmed value of P(e|c)=.75, the mean value actually obtained was .76.

5.2.1.3. Materials and Procedure

I used a Macintosh 8500 computer, running Macromedia Director 7.0, to administer the experiment. Participants read instructions on the screen, telling them that in this experiment they had to learn to what extent their

actions caused something to happen on the computer screen. They then proceeded to a specific description of the first scenario. The Light bulb scenario asked participants to imagine that a switch and light bulb displayed on the screen were connected. It further instructed them about another switch in a different room invisible to them. Other persons might flick that switch sometimes, so that when the bulb lights up, it could be either due to the participant's or other people's action. The instructions for the Grenade scenario were structurally identical, but asked participants to imagine that a FIRE! Button operated a grenade launcher to fire shells into a training range, and that other people might also fire shells into the range (see Appendices C through E).

In the Light bulb conditions, the computer displayed a drawing of a light bulb centrally against a gray background. 2cm below the light bulb was a white rectangular push-button labeled "Lightswitch" which inverted its colors when participants clicked on it with the mouse. An effect was represented as the bulb illuminating in yellow for 500ms accompanied by a whistling sound. In the Grenade scenario, the screen centrally displayed a rectangular viewing window (5 x 12 cm). The window displayed a view of the horizon with lines of schematic trees on the left and right side of the landscape. Approximately 2cm below the window was a blue rectangular push-button labeled "FIRE!". An effect was implemented as a 500ms display of a red and orange mushroom cloud in the center of the landscape, accompanied by an explosion sound. To make the representations more realistic, a continually repeating 20-second sound loop of gunfire and battlefield noises provided background sound in the Grenade scenario. The Light bulb scenarios were accompanied by a continually looping 20-second piece of funky music. Each condition lasted 2 minutes. At the end of each condition, participants had to indicate whether or not there was a causal relation between clicking the button and the outcome by selecting one of two radio buttons. If a participant indicated and confirmed

that there was no causal relation, the answer was scored as 0. If participants indicated that there was a causal relation, they read the following prompt:

> Nobody else is (turning on the light / causing explosions).
>
> If you clicked the switch 100 times,
>
> (how often would the bulb light up? / how many explosions would occur?)

Participants then had to enter a number between 1 and 100, confirm their entry and then proceeded to the next condition. After having worked on all five conditions in one scenario (Lightbulb or Grenade), participants received instructions and worked on the five conditions for the other scenario. The order of conditions within a scenario was random, and the order of scenarios was counterbalanced between subjects. Participants worked individually and the experiment lasted about 30 minutes.

5.2.2. Results

Table 5-3 displays participants' mean ratings for all five levels of causality, broken down by scenario. Mean causal ratings in the experimental condition with P(e|c)=.75 and a 0s delay were close to the normative level for both the Lightbulb (80.00) and the Grenade (81.56) conditions. As the temporal delay increased, ratings decreased in both scenarios. Results from the control conditions show that when P(e|c)=.00, causal ratings were lower than in any of the three experimental conditions. However, introducing 24 "uncaused" events but maintaining P(e|c)=.75 with no delay did not produce substantially lower ratings than the corresponding experimental conditions.

Due to heterogeneity of variance between the conditions, parametric statistics are not warranted. For the Light bulb scenario the number of participants who rated the P(e|c)=.75 , 0s delay experimental condition *higher* than the corresponding control condition (P(e|c)=.75 , 0s delay, 24 background

events) was 9, and the number of participants who rated the experimental condition *lower* than the control condition was 9, with no ties. There thus seems to be no difference between this pair of control and experimental conditions. For the Grenade scenario, the number of participants who rated the experimental condition *higher* than the control condition was 9, the number of participants who rated the experimental condition *lower* than the control condition was 6, with 3 ties. A paired sign test gives the probability of this result as p=.61. The addition of 24 randomly scheduled outcomes thus did not significantly decrease participants' evaluations of causal strength. This finding is clearly at variance with associative learning and computational causal power approaches. Even though the continuous, trial-free, paradigm I used does not allow a formal assessment of $P(e|\neg c)$, it is evident that adding "uncaused" background events increases the base-rate, the probability of the effect occurring in the absence of the candidate cause. In contrast to the three experimental conditions, that all employed $P(e|c) = .75$ and $P(e|\neg c)=0$, this control condition employed a value of $P(e|\neg c)$ greater than 0. Numerous studies in the past (see Shanks, Holyoak et al., 1996 for an overview) have shown that increasing $P(e|\neg c)$ while keeping $P(e|c)$ constant (i.e. lowering the objective contingency) results in degradations of causal judgments. Section 5.3.3.1 will come back to this abnormal finding and discuss why it might have occurred.

Table 5-3. Mean causal ratings from Experiment IV. Control conditions included 24 background events. Standard deviations in brackets. N=18

			Scenario			
Condition	P(e|c)	Delay	Lightbulb		Grenade	
Control	0.00	n/a	3.44	(11.78)	11.39	(20.13)
	0.75	0s	71.67	(25.38)	69.17	(28.56)
Experimental	0.75	0s	80.00	(9.08)	81.56	(11.83)
		2s	59.17	(29.57)	56.11	(35.67)
		5s	50.17	(36.61)	52.39	(43.61)

The main point of interest in this study was how the causal assessment of identical contingencies is affected by cause-effect-delays, and how knowledge mediates the influence of delay, so further analyses exclude the control conditions. Inspection of Figure 5-1 suggests a main effect of delay, unmediated by knowledge about the timeframe of the causal relation. The variance between the experimental conditions varied considerably, however, so an ANOVA is not warranted. It may be worthwhile to take a look at the distribution patterns of causal ratings in the experimental conditions, to understand how these differences in variance came about. Figure 5-2 displays histograms of the causal ratings in the six experimental conditions. In both the Light bulb and Grenade scenarios ratings for the 0s delay conditions scattered closely around 80 and varied widely for the 2s delay conditions. For the 5s delay, ratings in the Light bulb scenario were also extensively distributed, but low ratings (0 to 10) are the largest category. However, in the grenade condition the 5s delay produced a bi-modal distribution of judgements: participants largely either rated the causal relation to be non-existent (0 -10) or to be perfectly deterministic (90-10), with five participants providing ratings in between those extremes.

Figure 5-1. Experiment IV: Mean ratings of causal strength for experimental conditions. Error bars indicate standard errors.

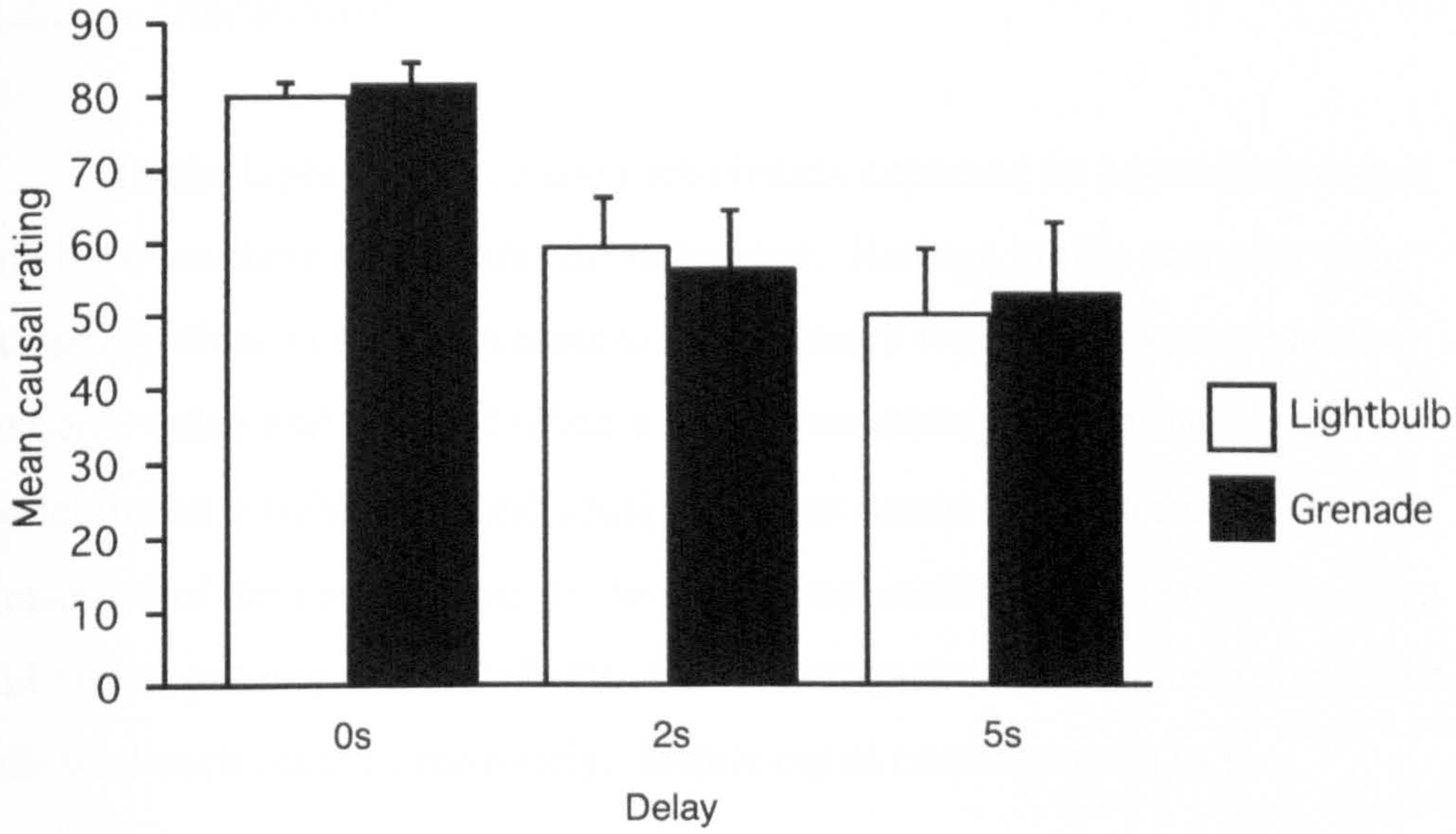
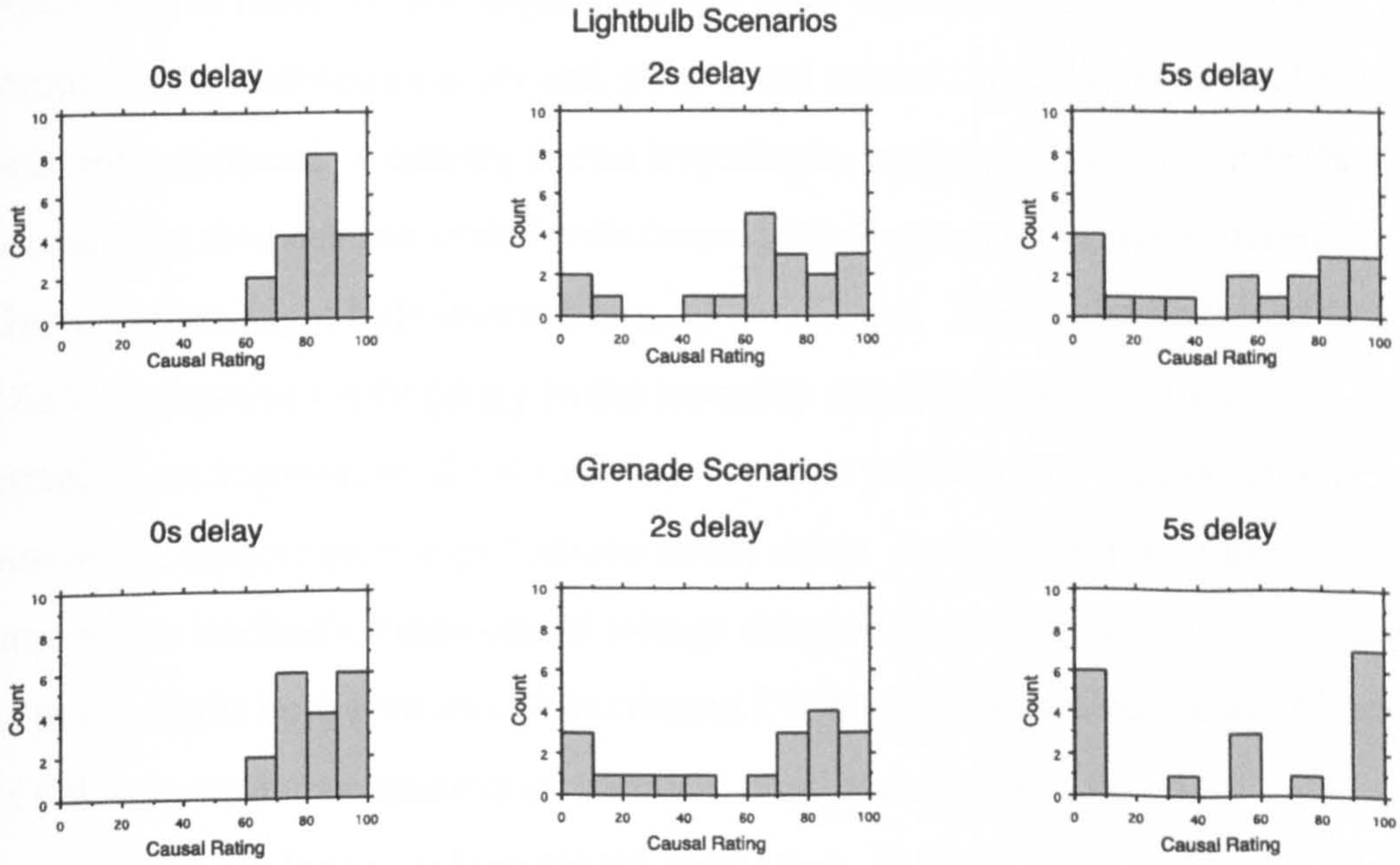


Figure 5-2. Experiment IV: Distribution patterns of causal ratings.



To investigate whether the impact of delay is mediated by knowledge, I computed difference scores for each participant's ratings in the 0s and 5s conditions of each scenario and subjected them to an paired t-test. The

influence of delay was the same in the Light bulb (*M*=29.83, *STD*= 34.14) and Grenade (*M*=29.18, *STD*=41.27) scenarios, *t*(17)=.08, *n.s.*

5.2.3. Discussion

In the Light bulb scenario participants expected an immediate causal link between their actions and the outcomes. Ratings in this scenario were extremely close to the implemented contingency when there was no action-outcome delay and dropped when a delay was introduced. In the context of switching on a light, temporal contiguity thus seems to be an important constraint of the causal relation: when an implausible delay separated cause and effect, participants gave lower causal ratings than when cause and effect followed each other immediately, despite equal contingencies in the conditions.

In the Grenade scenario participants expected a delayed causal relation. According to Einhorn and Hogarth (1986) such expectations should bridge temporal gaps between causes and effects and render immediate cause-effect pairings noncausal. Contrary to this hypothesis, participants did not rate the causal effectiveness associated with immediate contingencies lower under Grenade than Light bulb instructions, *t*(17)=.52, *n.s.* The immediate feedback of a high positive contingency in the no-delay conditions was sufficient to establish an impression of a causal link between participants' actions and the outcomes, despite prior expectations about delay. In the absence of this immediate feedback, mean causal ratings dropped just as much in the Grenade as in the Light bulb scenario. Experiment IV thus demonstrated a main effect of delay in causal assessment of identical contingencies, but failed to demonstrate evidence for knowledge mediation. It is not clear, however, why the Grenade cover story produced a bi-modal distribution of causal ratings in the 5s delay condition, but this point will be addressed again in section 5.3.3.1.

Overall, the detrimental effect of delay observed in Experiment IV was not as drastic as in Shanks et al. (1989), who reported that introducing a delay of 2 seconds into a .75 contingency already produced a drop from 73 to 40 on a rating scale from 0 to 100. Section 5.3.3.1 will discuss why delay might have had a lesser impact on causal judgments in Experiment IV than in comparable earlier studies.

5.2.3.1. Prior Experience of Contiguity and Within-Subjects Design

In retrospect the design employed in Experiment IV may have been sub-optimal to study knowledge-mediation in assessment of delayed causal relations for a number of reasons. Because the experiment was conducted on a computer, the knowledge manipulation induced by the cover stories may not have been convincing enough. Participants were very aware that the computer controlled the presentation of outcomes. Every participant rated each combination of delay and contingency twice, once with Light bulb, once with Grenade instructions. It seems therefore especially plausible that participants noticed the structural identity between the problems. Another reason for the absence of knowledge mediation in Experiment IV could be that (previously) experienced contiguity is a very powerful cue to causality. Participants generally underestimate delayed contingencies more severely if they previously encountered contiguous as compared to delayed contingencies (Buehner & Hagmayer, 2001). In other words, once participants have noticed that the critical covariation sometimes is contiguous, subsequent exposure to non-contiguous covariations fails to elicit appropriate estimates of causal strength (this rationale also explained the order effect I found in Experiment II). The logical way to circumvent both of these problems is to forfeit the economy of a within-subject design and collect data between-subjects instead.

5.3. Experiment V

Experiment V is a fully between-subject replication of Experiment IV, but included only the three experimental conditions. Every participant thus worked on only one of the six possible combinations of Scenario and Delay.

5.3.1. Method

5.3.1.1. Participants

124 volunteers (103 female, 21 male) participated. 73 of them were undergraduate students from the University of Sheffield and participated as part of a lab class, the remaining 51 participants were visitors to the Department of Psychology, Sheffield, and participated as part of an Open Day Demonstration.

5.3.1.2. Design, Materials, and Procedure

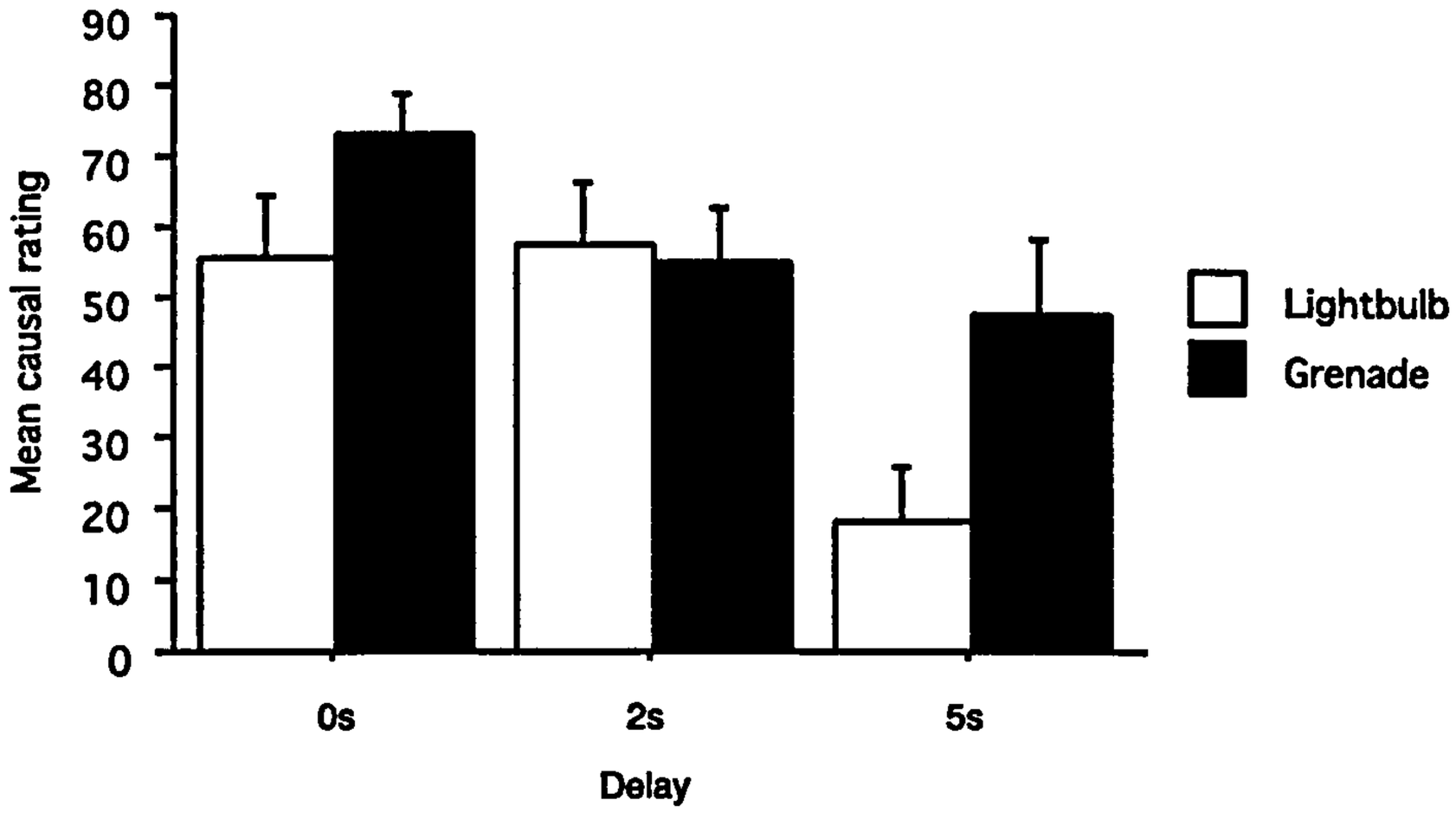
The design and materials were identical to Experiment V. Because the experiment was run in large groups, all the sound effects (background music/gunfire and whistle/explosion) were removed in order to keep mutual disturbance at a minimum. The initial instructions now informed participants that they would solve one problem, lasting about two minutes. Subsequently, participants read the specific instructions relevant for their assigned scenario and then proceeded to the experiment. Students enrolled in the lab class were split into three large groups of about 25 students each to be run on consecutive days. After students were seated at individual computers in the classroom, the experimenter divided the classroom into two groups (window side vs. side facing the wall) and gave each group instructions on where to locate the relevant experimental program on their computer. Participants then started the program themselves and followed the instructions on the screen. This

procedure assigned each participant to one of the six experimental conditions. Participants who took part during the Open Day demonstration were also led into the classroom and were seated in front of a computer each. Group size during the Open Day demonstration varied from about 12 to 20 persons. Each group was randomly assigned one of the six experimental conditions, but an effort was made to achieve approximately equal sample sizes overall in each group. The Lightbulb/0s cell comprised 19, the Lightbulb/2s 19, the Lightbulb/5s 18, the Grenade/0s 22, the Grenade/2s 27, and the Grenade/5s 19 participants. The experiment lasted about five minutes altogether. Once all participants were finished, they were dismissed from the computer lab and proceded to another classroom. Undergraduate students were debriefed by email a few days after data collection was completed, Open Day visitors were debriefed after completion of the experiment.

5.3.2. Results

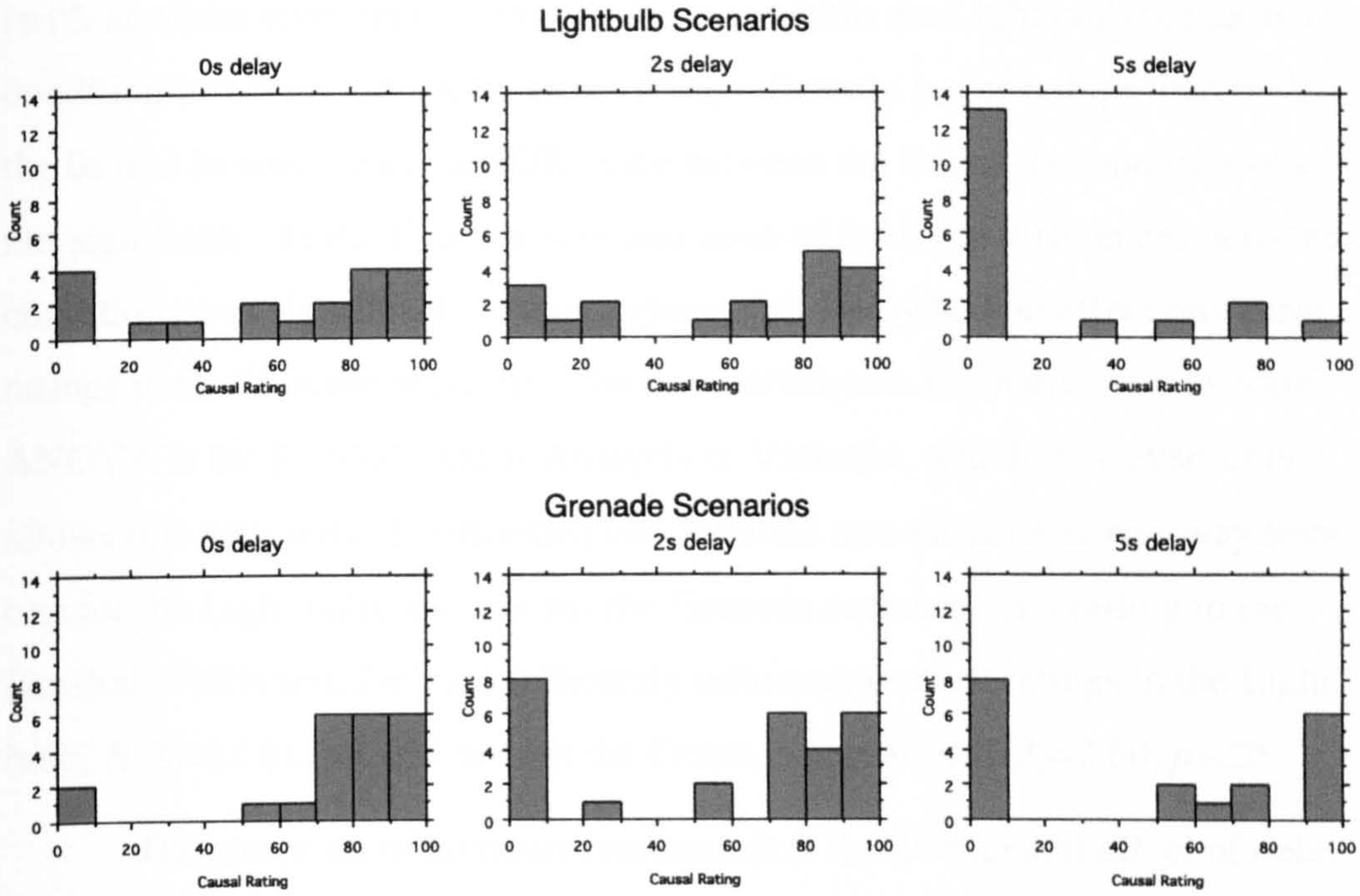
Figure 5-3 displays participants' mean causal ratings in each of the six conditions. As in Experiment IV, causal ratings generally decreased as the cause-effect delay increased. In contrast to Experiment IV, however, *Cover Story* also influenced participants' rating behaviour, particularly in the two conditions involving a 5s delay. Causal ratings in the 5s Grenade condition (M=47.63, STD=44.73) were considerably higher than in the 5s Light bulb condition (M=17.83, STD=32.06).

Figure 5-3. Experiment V: Mean ratings of causal strength. Error bars indicate standard errors.



A somewhat irregular finding is the pattern of ratings in the 0s delay conditions; the Light bulb scenario (*M*=55.79, *STD*=36.35) elicited lower causal ratings than the Grenade scenario (*M*=73.09, *STD*=26.93). This result not only contradicts the predictions of Einhorn & Hogarth's (1986) knowledge-mediation hypothesis (the difference is in the opposite direction than what would be expected, cf. introduction to 5.2), but also is at variance with the results from Experiment IV, where cover story had no influence whatsoever in the 0s delay condition. An inspection of the distribution patterns of causal ratings (Figure 5-4) reveals that this irregular result is due to four participants who provided a non-causal (zero) rating in the 0s Light bulb condition. Maybe some participants did not understand the rating instructions. It may well have been the case that the large group setting in which Experiment V was run made it harder for some participants to concentrate on the task, which in turn would result in more noise in the data. Figure 5-4 also shows that the distribution patterns in the 5s conditions of Experiment V are more extreme than they were in Experiment IV. The 5s Light bulb condition was rated as non-causal by an overwhelming majority of participants in Experiment V, compared to a more flattened distribution in Experiment IV.

Figure 5-4. Experiment V: Distribution patterns of causal ratings.



The 5s Grenade condition again gave rise to a bi-modal distribution, similar to Experiment IV.

Although the data from Experiment V do not show the problematic heterogeneity of variance as the data from Experiment IV, one still has to be cautious about applying and interpreting parametric statistics, because the data are clearly not normally distributed. I will present the outcome of an ANOVA and relevant post-hoc tests anyway, and then show that the key findings are similar when using non-parametric tests.

An ANOVA with alpha level of .05 revealed significant main effects of *Delay*, $F(2, 118)=7.804$, and *Cover Story*, $F(1,118)=4.991$. If knowledge mediates the timeframe of covariation assessment in causal reasoning, one would expect an interaction between *Delay* and *Cover Story*. The interaction would indicate that whether or how strongly delay influences causal judgment depends on people's assumptions regarding the timeframe of the causal

relation in question. The *Delay* x *Cover Story* interaction fell short of significance, however, $F(2,118)=2.063$, $p=.13$. Tukey HSD post hoc tests (with an alpha-level set to .05) revealed that within the Light bulb scenario, the condition involving a 5s delay received significantly lower ratings than both the 0s and 2s conditions; the difference between the 0s and 2s condition was not significant. In the Grenade scenario none of the three differences between conditions was significant. It thus appears that *Delay* had no effect on causal ratings in the Grenade scenario. The non-parametric alternative to a factorial ANOVA is the Kruskal-Wallis Analysis of Variance, which, however, only allows one-way tests. I performed two separate non-parametric one-way tests, one for the Light bulb, and one for the Grenade scenario. According to the Kruskal-Wallis test, *Delay* significantly influenced causal ratings in the Light bulb, $H(2)=11.81$, $p<.003$ but not the Grenade scenario, $H(2)=2.60$, $p>.25$.

The above set of comparisons considers the detrimental effect of delay on causal judgments derived from constant contingencies, and allows one to decide whether the effect disappears or is weakened in scenarios where people expect a delay. Another way to analyse whether and how knowledge and delay interact in influencing causal judgment is to compare causal ratings between conditions involving identical delays and check for an influence of *Cover Story*. Such comparisons are not focussed on detrimental effects of delay. Instead, they inform us whether the same objective evidence (identical covariations manifested with equal delays) is evaluated differently, depending on participants' assumptions about the timeframe of the causal relation. Tukey's HSD (with alpha level .05) revealed that *Cover story* produced significantly different causal ratings in the 5s conditions only: participants were more willing to judge a delayed contingency as causal if they expected a delay (in the Grenade scenario) than if they expected immediacy (in the Light bulb scenario). A non-parametric alternative to this set of comparisons would be to compute separate Mann-Whitney U statistics for each of the three *Delay*

conditions and check for an effect of *Cover Story*. The Mann-Whitney-U statistics were not significant in any of the three comparisons, however.

5.3.3. Discussion

Experiment V replicated the main finding of Experiment IV (and Experiments II and III), in that cause-effect delays generally impaired causal ratings in conditions with constant contingencies. In Experiment IV this detrimental effect of *Delay* was universal, and unmediated by knowledge as manipulated implicitly via *Cover Story*. I have argued that the failure to obtain any effect of *Cover Story* could have been the result of the within-subjects nature of Experiment IV. Participants were exposed to each reinforcement schedule (e.g. 75 contingency manifested with 5s delay) twice, once with Light bulb and once with Grenade instructions. It is thus very likely that they noticed the fundamental structural identity between the pairs and consequently learned that *Cover Story* concerned only a superficial alteration. This kind of transfer or learning effect was not possible in the between-subjects design of Experiment V, where each participant only worked on one problem, and indeed *Cover Story* did prove to be effective to influence how participants evaluated covariational evidence.

This methodological improvement of abolishing learning and transfer effects came with a price, however: the between-subject design (or the large group setting in which it was implemented) of Experiment V resulted in more extensively distributed data. This greater overall noise made it difficult to draw any firm conclusions from the data. Inspection of the distribution patterns of causal ratings (Figure 5-4) does reveal a clear influence of both *Delay* and *Cover Story*, however, and suggests that this influence was particularly strong in the conditions involving a 5s delay. Inspection of the means (Figure 5-3) further suggests that participants in the Grenade scenario were more willing to judge a delayed contingency as causal than participants

in the Light bulb scenario. One cannot interpret too much into this finding, however, particularly since it rests on a bi-modal distribution of ratings in the 5s Grenade condition. This bi-modal distribution pattern was also found in Experiment IV. The repeated occurrence of such an abnormality admonishes explaining it away by mere statistical noise. Rather, it implies that there must have been a fundamental ambiguity in the experimental materials that led to this bi-modal distribution. There are in principle two ways how a causal judgment experiment can be afflicted by ambiguity: the evidence presented from which subsequent causal judgments are meant to be derived can be problematic, or the dependent variable used to probe causal judgments can elicit more than one interpretation. Experiment II identified a serious problem inherent in the Free-Operant procedure, which resulted in unstable evidence for the causal relation. Given that Experiment III showed that a simple improvement to the procedure considerably alleviated the problem (and subsequent experiments employed this improved procedure), the first alternative seems not to be a likely candidate. This leaves the dependent variable used to probe causal ratings, and the next subsection will examine how the question I asked participants in Experiments IV and V might have confused some participants under certain conditions.

5.3.3.1. Perceptual or Causal Judgments?

The dependent variable employed in Experiments IV and V was based on a frequency estimate couched in a counterfactual question:

> Nobody else is (turning on the light / causing explosions).
> If you clicked the switch 100 times, (how often would the bulb light up? / how many explosions would occur?)

I chose this question format over the more traditional rating scale "How strongly do you think clicking the switch causes the bulb to light up / explosions?" because my earlier work (Buehner et al., 2001) showed that standard rating scales are often problematic in causal reasoning research. Buehner et al. found that when people base causal estimates on a traditional rating scale, judgments can actually reflect a conflation of causal strength and the reliability of the information provided.

Consider a participant who provides a low rating in answer to the above question. Her low rating is equally consistent with a strong belief that clicking the switch does not cause explosions at all as it is with a weak belief that clicking the switch does cause explosions, possibly even strongly, but she just does not know for certain. Employing a standard rating scale can then be especially problematic when both reliability and strength vary within the same experiment, as was the case in Buehner et al.'s study, where participants worked on multiple conditions with varying contingencies and a constant number of trials per condition. Consider, for example, causal ratings derived from two non-contingent conditions, one in which the effect never happened at all ($P(e|c)=P(e|\neg c)=0$), and one in which it happened equally often in the experimental and control groups ($P(e|c)=P(e|\neg c)=.75$). If the number of learning trials in each condition is constant, say eight, it follows that one would be more confident of the non-causal status in the former than the latter condition. Assuming that alternative causes are constant between the experimental and control groups, the cause would have had eight trials to "prove" its power in the first condition, and it failed on all eight of them. By that same rationale the cause would have had only two trials left to show its power in the second condition (alternative causes already produce the effect 75% of the time, i.e. 6 out of 8 times), and it failed on both. Providing a constant number of trials across conditions therefore leads to varying reliability between these conditions: because there were more trials on which the cause could have but in fact failed to prove its power in the former

condition than in the latter, participants might be more confident of a noncausal rating in the former condition, leading to a causal rating closer to 0.

The frequency estimate couched in a counterfactual question constituted an enormous improvement in Buehner et al.'s paradigm, as it was no longer susceptible to the conflation of reliability and strength. Using frequency as the dependent measure increased the concreteness of the reasoning task and improved overall accuracy. Gigerenzer and Hoffrage (1995), for instance, showed that performance on a fairly complex reasoning task requiring Bayesian inference improved dramatically, if the information (and question) was phrased in frequency format, rather than in probabilities. In the current experiment, reliability was not an issue, however, as there was only one contingency. Although it may be true that frequency estimates in general allow more accurate representations of people's (causal) beliefs than ratings on a scale, the frequency estimate procedure may not be the best way to assess causal judgments in the context of delays.

Imagine a participant who experiences a high contingency implemented with an implausible long action-outcome delay. Although the participant may think that the causal relation is weak, she still observes that effects are frequently preceded by candidate causes. Consequently, a high frequency estimate would be consistent with the perceptual quality of the learning experience. In other words, the number provided could be based on a projection from causal beliefs about the candidate cause (the measure I intended it to be), or on a projection from the belief about how strongly alternative causes produce the effect, or a combination of the two. Because the question format used in Experiments IV and V was ambiguous, some participants may have provided perceptual judgments while others gave actual causal ratings. This would also explain the heterogeneity of variance in the experimental conditions of Experiment IV, and the abnormal finding in the control conditions of Experiment IV. In the control conditions of Experiment IV the abnormal finding was that adding a substantial number of "uncaused"

effects while maintaining P(e|c) at .75 did not lower causal ratings. This result is at variance with all existing accounts of causal learning, both associative and causal power based accounts. If the probability of the effect given the cause P(e|c) stays the same, but the probability of the effect given the absence of the cause P(e|¬c) increases – as was the case when inserting "uncaused" background events – all theories predict that causal judgments should decrease. However, if the judgments given by participants are not pure estimates of causal strength, but are confounded with estimates of the base-rate, then there is no reason to expect that judgments should decrease. On the contrary, they might even *increase*, which in fact they did in some participants.

Another problem with the rating procedure in Experiments IV and V has to do with the wording of the counterfactual question: "Nobody else is (turning on the light / causing explosions)". While I intended it to be interpreted as a counterfactual, it may well be that some participants did not process the statement as a counterfactual, but as a true statement instead. It would have been better to phrase the statement as "Suppose nobody is...." Without this addition, some participants may have taken the statement at face value, which of course would have violated what they learnt in the instructions, namely that another person is also trying to cause the effect. Because the statement may thus have overridden the initial instructions, some participants could have consequently assumed that P(e|¬c) was 0 all along. This in turn would have led them to attribute all occurrences of the effect to their button presses, as – according to their belief – no other plausible cause existed. This interpretation can explain both why increases in P(e|¬c) while keeping P(e|c) constant did not lower causal ratings, and why the overall effect of delay was not as substantial as in comparable earlier studies (e.g. Shanks et al., 1989).

The frequency estimate procedure is also potentially confusing with respect to the timeframe, both over which one has to imagine pressing the button 100 times and over which one has to imagine the occurrence of

potential effects. Given the reinforcement procedure employed in these experiments (a .75 contingency implemented with 0, 2, or 5 seconds delay), it is not straightforward to envision what would happen if one pressed the button 100 times. Because these experiments dealt with causal relations elapsing in continuous time, frequency estimates are probably conceptually harder to represent than ratings on a scale. Frequency estimates work fine in experiments that do not deal with evidence sampled over continuous time. Buehner et al.'s (2001) participants, for instance, had to learn whether particular medications produced headaches as side-effects in populations of allergy patients. After studying a datasheet providing covariational data in visual format, participants had to imagine 100 patients all of whom did not have a headache, and were then asked how many of them would have a headache if they had taken a particular medication. It is evident that frequency estimates in this context are much easier to grasp than in Experiments IV and V. For all these reasons I decided to change the dependent variable used to probe causal estimates in the last experiment to a standard rating scale.

5.4. Experiment VI

Experiment VI was a replication of Experiment V, but used a rating scale to probe for causal ratings, and employed the 0s and 5s conditions only in order to obtain a larger sample in these two maximally informative conditions.

5.4.1. Method

5.4.1.1. Participants

160 (116 female, 44 male, median age 18) visitors to the Department of Psychology, Sheffield participated as part of an Open Day Demonstration, run on 4 separate days. Participants were randomly assigned to the four

conditions, but an effort was made to achieve approximately equal numbers in each cell. The Lightbulb/0s cell comprised 57, the Lightbulb/5s 36, the Grenade/0s 36, and the Grenade/5s 31 participants.

5.4.1.2. Design, Materials, and Procedure

The materials were identical to Experiment V, with two exceptions: the number of conditions included in the study (only the 0s and 5s conditions were included), and the dependent variable used to probe causal ratings. After having sampled evidence from the continuous paradigm for two minutes, participants had to rate causal strength on a scale from 0 to 100. I wanted to keep the possibility that participants base their ratings on the perceptual quality of the feedback at a minimum. To this end, I labelled the extreme ends and midpoints of the scale in such a way that would encourage participants to provide ratings based on their causal beliefs, i.e. 0 means that clicking the switch has no influence on whether or not the effect occurs, 50 means that clicking the switch moderately causes the effect, and 100 means that clicking the switch strongly causes the effect (see also Appendix F).

A comparison of the rating scales employed in Experiment VI with the materials used in Experiments II and III reveals that they are nearly identical. The reason for this is that although conceptually Experiments II and III preceded Experiments IV through VI, data collection for Experiments IV through VI took place before Experiments II and III. It was thus possible to learn from the flaws of Experiments IV and V and employ the same rating procedure in Experiments II and III as in Experiment VI.

5.4.2. Results

Figure 5-5. Experiment VI: Mean ratings of causal strength. Error bars indicate standard errors.

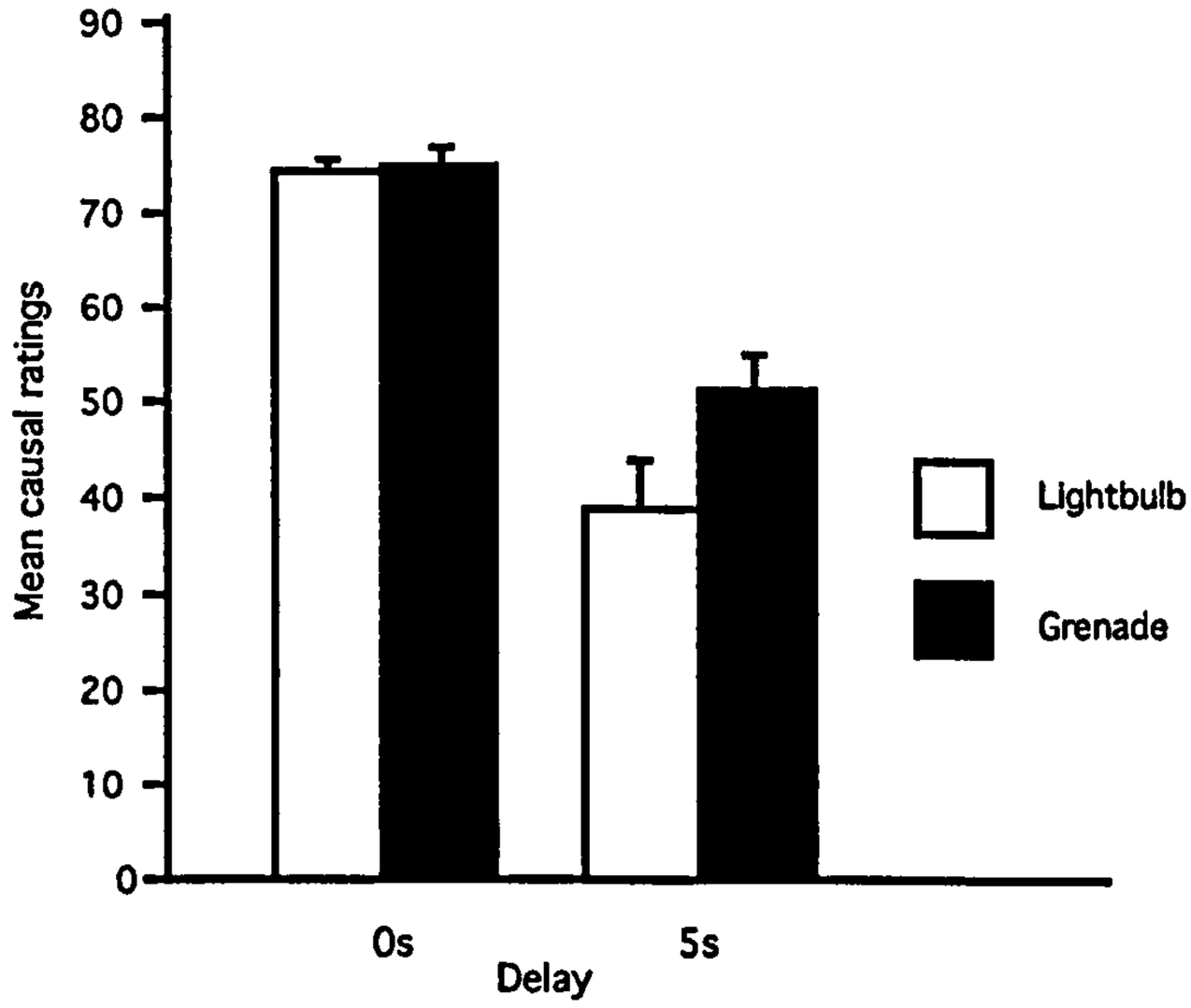
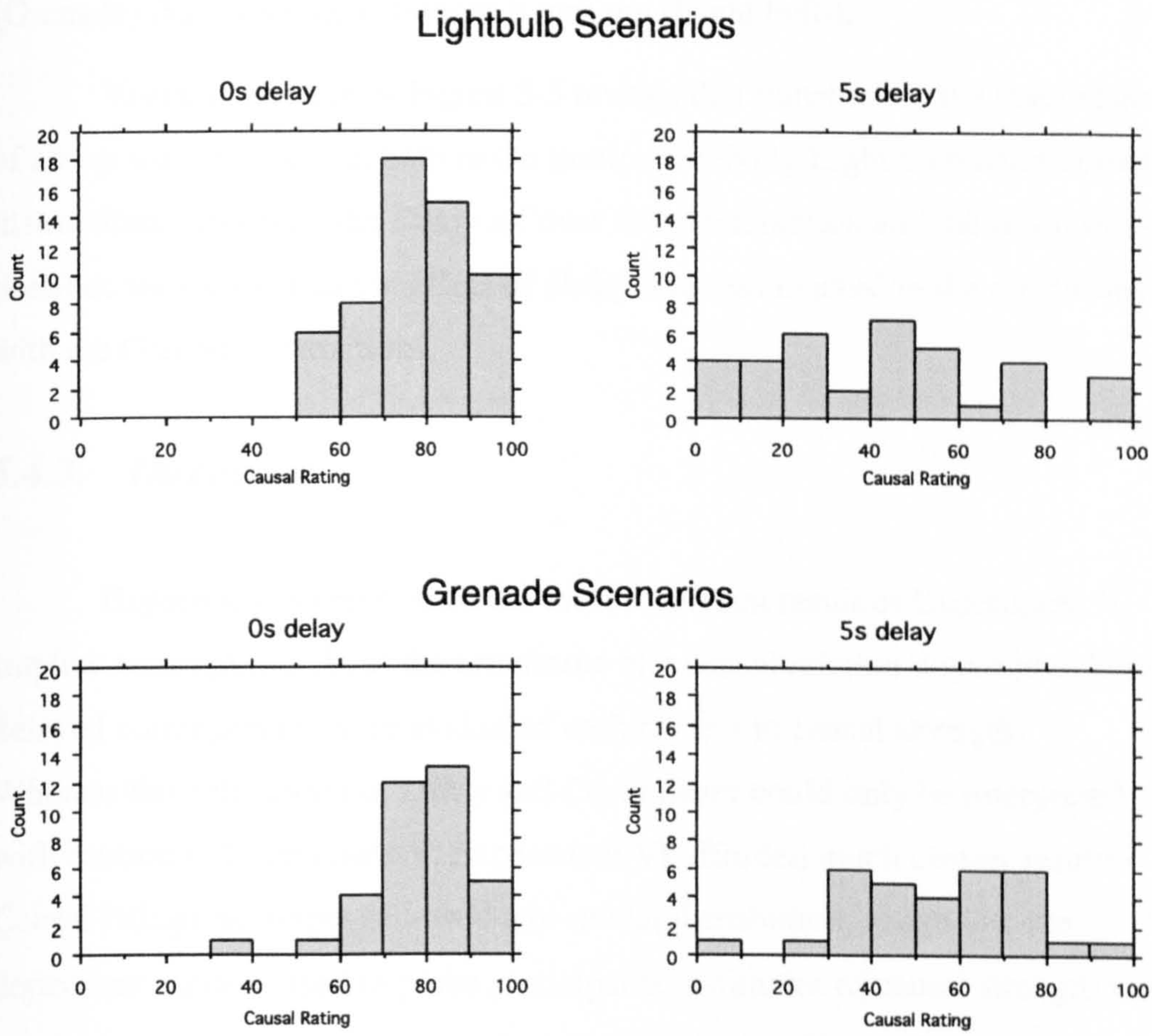


Figure 5-5 displays mean ratings of causal strength in the four groups. As in Experiments IV and V, participants interpreted contiguous contingencies as highly causal in both the Light bulb (*M*=74.1, *STD*=12.51) and Grenade (*M*=74.7, *STD*=12.56) scenarios. The 5s delay generally resulted in lower ratings of causality; this finding was more pronounced in the Light bulb (*M*=38.9, *STD*=27.44) than the Grenade (*M*=51.2, *STD*=20.82) group.

Visual inspection of Figure 5-6 reveals that the distributions of causal ratings were considerably more normal than in Experiments IV and V. Causal ratings in the 5s Grenade condition in particular did not follow a bi-modal distribution anymore, but instead scattered mostly between 30 and 80, producing a flattened normal distribution.

Figure 5-6. Experiment VI: Distribution patterns of causal ratings.



The data in Experiment VI satisfy all key assumptions for parametric statistics, so I will report an ANOVA on the causal ratings with *Delay* and *Cover Story* as independent variables. The significance level was again set to .05. The ANOVA revealed main effects of *Delay*, $F(1,156)=95.272$, and *Cover Story*, $F(1,156)=4.556$, and a marginal *Delay x Cover Story* interaction, $F(1,156)=3.811$. The *Delay* x *Cover Story* interaction indicates that the impact of delay on causal ratings was mediated by assumptions about the timeframe of the causal mechanism. Tukey HSD post-hoc tests (with an alpha level of .05) revealed that estimates of causal strength in the conditions involving a 5s delay were significantly higher in the Grenade scenario than in the Light bulb scenario. When implemented with 0s delay, estimates were identical in both

scenarios. This result shows that participants were more willing to judge a delayed contingency as causal when they thought the delay was plausible (Grenade) than when they thought it was not (Light bulb).

Visual inspection of Figure 5-5 reveals that nonetheless the main effect of *Delay* was very strong, both in the groups receiving Light bulb and Grenade instructions. However, the *Delay x Cover Story* interaction and the result of the post-hoc tests show that the effect of *Delay* was less marked in the conditions with the Grenade instructions.

5.4.3. Discussion

Experiment VI replicated the most important result of Experiment V. Implicit assumptions about the timeframe of a causal relation determined how delayed contingencies were evaluated with respect to causal strength. Whereas this interaction of *Delay* and *Cover Story* could only be interpreted with caution in Experiment V, Experiment VI afforded much clearer results. Causal ratings no longer followed a bi-modal distribution, a sign that the dependent variable used to probe participants' estimates of causal strength was much clearer and less ambiguous than in Experiments IV and V.

The *Delay x Cover Story* interaction was straightforward to interpret: given a constant cause-effect contingency of .75, a 5s delay always impaired causal judgments relative to no delay; this detrimental effect of delay was less pronounced when participants thought a delay was plausible (Grenade instructions) than when they expected immediate cause-effect pairings (Light bulb instructions). This finding partially supports Einhorn & Hogarth's (1986) knowledge-mediation hypothesis: participants bridged temporal gaps between causes and effects better when they assumed a causal mechanism that took time to unfold, than when they contemplated an immediate mechanism. A stronger and complete support of Einhorn & Hogarth's hypothesis would additionally require that contiguous contingencies be *not* interpreted as

evidence for a causal relationship if reasoners assume a causal mechanism that requires a delay between cause and effect. The questionnaire data I reported in section 5.1 showed that people have fairly strong expectations that about five to ten seconds should pass between firing off a grenade and observing an explosion several miles away. Nonetheless, participants disregarded these temporal assumptions when they evaluated contiguous contingencies in the Grenade scenario; they provided equally high estimates of causal power in the Grenade and Light bulb scenarios, even though they assumed a delayed mechanism in the former and an immediate mechanism in the latter scenario.

5.5. Discussion and Summary of Experiments IV through VI

There were two motivations behind Experiments IV through VI. First, to test whether the results from Experiments II and III replicate in a more ecologically valid paradigm, and second, to allow a strong test of Einhorn & Hogarth's (1986) Knowledge Mediation hypothesis. The experimental procedure employed in Experiments IV through VI was similar to Experiments II and III in that both paradigms were based on a Free-Operant procedure implemented with probabilistic feedback. Participants were exposed to a reinforcement schedule for a fixed time, during which they could make responses (mouse clicks) whenever they wanted. They were then asked to indicate how strongly they thought their actions produced an outcome on the computer screen. The crucial difference between the two sets of studies was that Experiments IV through VI employed concrete stimulus materials and asked participants to imagine real world causal mechanisms, whereas Experiments II and III employed very impoverished stimuli, that bore no relevance to real world causal mechanisms. This distinction meant that in Experiments IV through VI it was no longer necessary to explicitly inform participants about potential delays; instead, these experiments employed two

different cover stories and scenarios, and relied on participants' implicit assumptions about the timeframe of the causal relation in question.

Experiment IV failed to replicate the principal finding from Experiments II and III – an interaction between *Instructions* and *Time.* Instead, causal judgments decreased as the delay increased, regardless of participants' assumptions about the timeframe of the relation in question. Experiments V and VI were between-subject replications of Experiment IV; Experiment VI also employed a less ambiguous rating procedure as Experiments IV and V. With these methodological improvements, it was possible to replicate the principal finding from Experiments II and III: knowledge (as manipulated implicitly via instructions) and degree of contiguity interacted to influence participants' estimates of causal strength. Experiments V and VI thus could show that knowledge mediates the timeframe of covariation assessment in causal reasoning, and extended the findings from Experiments II and III to a more ecologically valid paradigm. Participants in Experiments V and VI bridged temporal gaps between candidate causes and effects without being explicitly told to do so (as in Experiments II and III); instead, they recruited their world knowledge about the causal mechanisms in question to help them bridge the gaps.

The second motivation behind Experiments IV through VI was to allow a strong test of Einhorn and Hogarth's (1986) Knowledge Mediation hypothesis. I intentionally picked the Grenade scenario, because it implies that a delay between cause and effect is not only plausible, but even necessary. A grenade cannot cause an explosion several miles away the instant it is fired off; it has to fly through the air, and this takes several seconds (participants' estimates of the cause-effect delay were around 8s, see section 5.1). Knowledge Mediation predicts that immediate cause-effect pairings, which contradict expectations of a delayed causal mechanism, should not be interpreted as evidence for a causal relation. Experiments IV through VI failed, however, to provide any support for this strong prediction of the

Knowledge Mediation hypothesis. Contingencies implemented with 0s delay always gave rise to high causal ratings, regardless of the assumptions about the timeframe as manipulated through the cover story. Note that even though instructions likewise had no influence on causal ratings derived from contiguous conditions of Experiments II and III, this finding bore no relevance to the strong interpretation of the Knowledge Mediation hypothesis, as the instructions in Experiments II and III merely stated that delays were *possible/plausible* – the cover stories employed in Experiments IV through VI implied that a delay was *necessary* in the Grenade scenario

Does this imply that contiguity is more important than expectations about the timeframe of the causal relation? That contiguity overrides temporal assumptions? Not necessarily. It is true that all previous research on this topic (including Shanks et al., 1989; and Schlottmann, 1999) showed that contiguity is a very powerful cue to causality. It facilitates the discovery and appropriate assessment of contingencies. Previous studies, however, uniformly showed that the assessment of causal relations in the absence of contiguity is very hard, so much that delays of more than 2 seconds cannot be tolerated (Shanks et al., 1989). My results showed that this hardship can be alleviated by explicit or implicit knowledge about the timeframe of the candidate relation. As Schlottmann's results from the 5 and 7 year olds and my results from the contiguous conditions showed, experienced contiguity may even appear to override knowledge-based expectations about the timeframe of the causal relation in question, suggesting a bottom-up contiguity bias. What is unclear as yet is whether adults, when reasoning about complex probabilistic causal relations will always behave like 5-7 year olds, or whether they could in principle appreciate the *necessity* of delays.

One explanation for the strong effect of temporal contiguity in Experiments IV through VI could be that the task was administered on a computer. As discussed earlier, people have very rich conceptions about (and, in most cases, a great deal of experience with) interactions with computers.

Consequently, participants in my experiments must have been aware that the computer controlled all aspects of the stimulus display. I tried to minimize the impact of this problem by employing a between-subjects design in Experiments V and VI, but apparently did not succeed completely. This change ensured that participants could not notice the structural identity between Light bulb and Grenade problems. However, even when solving only one short problem of 2-minute duration, participants in the Grenade scenario had no reason to believe that a delay between cause and effect was *necessary*. I managed to induce participants in the Grenade scenario of Experiments V and VI to think that a delay between their actions and the display of outcomes was *plausible*, but could not get them to assume it was *essential*.

Another possible reason for why Experiments IV through VI failed to provide support for the strong claim of the Knowledge Mediation hypothesis could be a discrepancy in the amount of experience participants had with the (imagined) causal mechanisms. A typical undergraduate student probably is exposed to the Light bulb problem in the real world several times throughout the day. We are surrounded by electric light nearly everywhere we go, and we use it regularly. Consequently, we have a vast amount of experience with the simple causal connection between a light switch and the bulb lighting up. The Grenade scenario, in contrast, is probably unfamiliar to most students. In fact, probably none of my participants ever had any direct experience with this causal mechanism in real life. As a consequence, participants' beliefs about the timeframe of the causal mechanism in question could have been more stable in the Light bulb than in the Grenade scenario. In other words, when participants were subjected to a mismatch between their assumptions about the timeframe of the relation in question and the experienced temporal feedback, they might have been more willing to update their beliefs about the Grenade than about the Light bulb scenario.

Despite of these explanations, it is important to stress here that participants in Experiments IV through VI were "correct" (in a normative

sense) in assigning high causal ratings to contiguous conditions in both the Light bulb and Grenade scenarios. After all, their actions really did cause the outcomes with a probability of .75, and the computer never produced the outcome *unless* they pressed the button. Because participants knew that the computer controlled stimulus display and feedback, it would have even been irrational for them to deny the existence of a causal link between their actions and the outcomes in the 0s Grenade scenario. Participants would have had to explicitly disregard objective evidence that suggested a strong causal link in order to conclude that the 0s Grenade condition was non-causal. If one wanted to demonstrate knowledge-mediation on contiguous contingencies it seems imperative to create a scenario where delays indeed are judged to be necessary, and it seems likely that a Free-Operant procedure on a computer will never be able to meet this requirement. Real physical causal mechanisms, like the ones in Schlottmann's (1999) study are probably better suited for strong tests of Einhorn & Hogarths (1986) hypothesis. Adult participants in her study easily learnt that a five second delay was necessary when the slow toy was inside the box.

6. General Discussion and Outlook

6.1. Summary

I began this thesis with a review of the most influential theories of human causal induction. With the exception of the perceptual causality approach (1946/1963) most theories have taken up Hume's (1739/1888) analysis of causal induction: causal relations are unobservable; they are the result of a mental process which operates on observable evidence in the form of co-occurrences between candidate causes and effects. The main focus of recent research in this area has been on illuminating the exact nature of the mental leap from covariation to causation. One ongoing debate is, for example, whether the nature of the process is rule-based or associative. The rival accounts can be distinguished by their predictions on how manipulating certain aspects of the covariational evidence (e.g. base-rate, or direction of learning) affect causal judgments. As a consequence, most of the recent empirical work has been aimed at supporting one, and refuting the other account (e.g. Baker et al., 1996; Baker, Vallee Tourangeau, & Murphy, 2000; Buehner & Cheng, 1997; Buehner et al., 2001; Lober & Shanks, 2000; Perales & Shanks, 2000; Shanks & Lopez, 1996; Waldmann & Holyoak, 1992, 1997; Waldmann, 2000; for an overview see Shanks, Holyoak et al., 1996). Although this debate was, and still is, highly fruitful and productive, it also meant that one important precursor for causal induction from covariation has largely been forgotten or ignored. In order to derive causal knowledge from covariation, a reasoner first has to notice that cause and effect have co-occurred. Most empirical studies have circumvented this problem by providing participants with covariational evidence (condensed in contingency tables, or presented in discrete trial structures). What little evidence relevant to this question existed (Michotte, 1946/1963; Shanks et al., 1989; Reed, 1992, 1999), however, painted a rather unflattering picture of human causal

induction: people could not correctly identify causal relations in laboratory tasks, if cause and effect were separated by more than two seconds.

These findings are at variance with our intuitions about every day causal inference where people apparently can reason about causal relationships that involve considerable delays. I have identified how two major theoretical frameworks account for the paradoxical laboratory findings. Associationism stresses the importance of temporal contiguity for causal induction and states that non-contiguous action sequences should always give rise to lower impressions of causal strength than comparable contiguous pairings. Causal power, on the other hand, does not bestow a privileged role to contiguity; whether or not a particular covariation gives rise to a causal impression crucially hinges on the reasoner's assumptions about the mechanism linking cause and effect. Knowledge about the causal mechanism also entails assumptions about its timeframe; consequently, whether or not delayed causal relations are identified correctly depends on the reasoner's beliefs about the causal mechanism in question.

I have next pointed out that existing empirical data cannot distinguish between these two rivalling accounts. Relevant previous experiments had always employed causal mechanisms that created assumptions of immediacy in participants. When these expectations were paired up with experienced delays, participants could not identify causal relations correctly. Delayed cause effect pairings may have failed to elicit causal impressions either because contiguity is essential to causal induction (as associationism argues), or because of a mismatch between belief and experience (the Knowledge Mediation account).

In its empirical part, this thesis contained six experiments that investigated the role of temporal contiguity in human causal induction. Experiment I sought to clarify whether people (erroneously) attach more importance to temporal contiguity than to covariation. This was not the case; participants' causal judgments were solely determined by the true underlying

causal structure, as manifested in two different contingencies. Identical contingencies were interpreted equivalently, regardless of whether they were implemented with contiguous or delayed cause effect pairings. Temporal contiguity thus plays no privileged role in human causal induction. Experiment I also revealed that humans understand the importance of time in causal induction. When the same physical event had different causal powers with regard to the effect, depending on its temporal position relative to the effect, participants used this temporal position to correctly indicate how likely they thought the effect would occur.

Experiments II through VI were dedicated to direct tests of Einhorn and Hogarth's (1986) Knowledge Mediation hypothesis. Experiments II and III used the same paradigm as Shanks et al.'s (1989) original study, where participants evaluated the causal effectiveness of various instrumental contingencies. The cause-effect contingency always remained at a constant high value, but the cause-effect contiguity was either immediate or delayed. The important modification from Shanks et al.'s procedure was that my experiments manipulated whether or not participants were explicitly instructed that the causal relation sometimes might involve a delay. When participants were alerted to delays, their causal ratings derived from delayed contingencies were significantly higher compared to ratings from a group of participants who were ignorant of the possibility of delays. Furthermore, participants no longer distinguished between immediate and delayed contingencies if the conditions were optimal (no order effects, no confounding of contiguity and contingency). Experiment II also revealed a serious methodological problem associated with the use of Free-Operant instrumental paradigms in causal reasoning research: Decreasing the cause-effect contiguity automatically results in lower cause-effect contingencies, and thus weaker objective evidence for the relation in question. Experiment III (and all subsequent experiments) employed a modified Free-Operant procedure that considerably alleviated this problem.

Experiments IV through VI sought to extend the findings from Experiments II and III to a more ecologically valid paradigm, and were designed to allow a stronger test of Einhorn & Hogarth's (1986) hypothesis. Rather than instructing participants explicitly about the timeframes of causal relations, these experiments used two different cover stories. The Light bulb scenario was aimed at eliciting expectations of immediate, the Grenade scenario expectations of delayed cause-effect pairings. Unlike the instructions in Experiments II and III, which stated that delays are merely plausible, the Grenade cover story was intended to create the assumption that a delay was necessary. Various methodological problems in the procedures of Experiment IV and V prevented a successful replication of the principal findings of Experiments II and III. I curtailed these problems in Experiment VI and could replicate the Knowledge Mediation effect. The results did not pass the strong test, however, as contiguous contingencies were always rated as highly causal, even when the cover story implied a delayed causal mechanism. This failure is probably not indicative of the superordinate nature of contiguity, however, but rather reflects participants' rational inference strategies (discounting the assumptions elicited by the cover story and attending to the objective evidence of a strong instrumental contingency).

6.2. Re-considering the Paradox between Real World and Laboratory Causal Cognition

My motivation in this thesis was to resolve the paradox between previous experimental results – humans fail to identify causal relations involving more than a few seconds of delay – and everyday causal cognition, where people apparently can reason about delayed causal relations with relative ease. The analysis of the literature identified two theoretical explanations for participants' poor and apparently irrational performance in previous tasks.

According to the associationist perspective, cause-effect delays always result in weaker increments of associative strength. Everything else being equal, delayed event sequences thus give rise to weaker impressions of causality than do immediate cause-effect pairings. This framework of course has been inspired by theories of animal learning, with the openly admitted agenda of scrutinizing similarities between animal conditioning and human learning (see e.g. Shanks & Dickinson, 1987). In the very specific environment an animal faces in a typical Skinner box, responding less on delayed than on immediate reinforcement schedules may be a perfectly rational behaviour. After all, the utility or expected gain from responding typically decreases as the interval between response and outcome increases. Utility and causality are not equivalent concepts, however. Consider the following two reinforcement schedules: schedule A delivers one unit of reward (say, a food pellet) with a probability of 75% given a response; schedule B delivers three units of reward with the same probability. Although the two schedules have identical underlying causal structures, the expected gain from responding is three times higher on schedule B as compared to A. Consequently, schedule B should elicit higher levels of responding (cf. Shanks, 1993b). A similar argument can be made about delay of reinforcement. Although two schedules may have identical underlying probabilistic structures, one employing a reinforcement delay will produce a lower rate of reinforcement (conditional on responding) relative to a schedule that delivers immediate reinforcement. Anderson and Sheu (1995) have pointed out this important distinction between probabilities and rates, and showed that human causal judgment in their continuous paradigm was sensitive to rates rather than probabilities (for a rate based account of conditioning, see Gallistel & Gibbon, 2000). This sensitivity to rates presents a key to unravelling the paradox between experimental results and everyday causal cognition.

Anderson and Sheu (1995) suggested that judgments of causal strength in a continuous paradigm are a function of the contrast between the rates of the effect occurring given the presence versus the absence of the cause. Such conditional rates R(e|c) and R(e|¬c) of course can only be defined relative to a specific timeframe. The observer needs to have some basis for attributing effects either to the candidate cause (thus increasing estimates of R(e|c)), or to the background of alternative causes (increasing estimates of R(e|¬c)).

This nicely resonates with the second explanation for participants' poor performance I found in the literature: Einhorn and Hogarth's (1986) knowledge mediation account. According to this theory, the impact of delay on causal learning depends on the expectations about the timeframe of the relation in question. If reasoners assume that a candidate cause should exert its influence immediately, they will attribute delayed occurrences of the effect to alternative causes, and not to the candidate in question. Such a mismatch between knowledge and experience often results in a perfectly rational rejection of the causal relationship in question, for instance when people do not "perceive" a causal relationship in delayed manifestations of Michotte's (1946/1963) launching paradigm. However, if experimenters do not take sufficient care to check whether the properties of the reasoning task they employ indeed match up with participants' expectations, seemingly irrational behaviour and judgment may arise.

Whether or not it was irrational for Shanks et al.'s (1989) participants, for instance, to deny the existence of a causal relationship even when in fact they were in control of the stimulus presentation lies of course in the eye of the beholder. The experimenter is focused on the fact that the contingency stays constant across problems despite increases in the delay, and hence thinks the true causal effectiveness of key presses is likewise constant (this need not even be the case, as the behavioural data from Experiment II showed, section 4.1.2.1). Participants, on the other hand, come to the experiment with rich conceptions about computers; they expect that their actions have immediate

consequences (cf. Shanks et al., 1989 p.155). Furthermore the instructions in Shanks et al.'s paradigm (and all the relevant replications, including this one) explicitly mentioned that the apparatus sometimes might produce the effect on its own, independent of participants' behaviour. This statement was only fulfilled in the yoked control conditions (and that seems to be the only purpose of the yoked conditions), but participants do not know this. In fact, one can even point out the similarity to the blocking paradigm (see section 1.2) that already crippled Siegler & Liebert's (1974) study (cf. section 2.1.3.2). The instruction phase explicitly introduced the computer as a cause that sometimes brings about the effect, and one can never know whether a given flash of the triangle in the Free-Operant phase was brought about by the computer, or by one's key press. If the expectations about immediate feedback are fulfilled, it is easy to recognize the causal relation between key presses and the outcome; if, however, the expectations are violated, it is only rational to attribute causality to the background, which was already been established as a predictor

My experiments have shown that the seemingly irrational detrimental effect of delay disappears if the quality of the feedback created through the experimental paradigm does not diverge from participants' expectations. The paradox thus is not one between participants' performance in laboratory tasks and their performance in the real world, but rather one between experimenters' conceptions of the reasoning processes involved when participants solve causal induction tasks in laboratories and the apparent complexity of the computations required to solve such tasks, both in the real world and in laboratories. Human causal induction was probably perfectly rational all along, but psychological theories of causal induction might sometimes have been irrational.

6.3. Can Associationism Account for Effects of Knowledge Mediation?

I set up this thesis around two competing explanations for the detrimental effect of delay in causal learning found in laboratory tasks: associative learning theory and the knowledge mediation account. While the former postulates that experienced delays *always* result in weaker associations relative to immediate pairings (see e.g. Dickinson, 2001 for further implications of such an account), the latter proposes that the influence of delay depends on prior knowledge or experience. Such complex knowledge falls outside the scope of associative learning theory. Consequently, it predicted no influence whatsoever of experimentally induced delay assumptions. Instead, the experience of cause-effect delays should have uniformly weakened causal ratings, irrespective of knowledge. Experiments II, III, and VI clearly contradicted this prediction. The influence of delay was significantly less prominent when participants were aware of potential delays, just as predicted by Einhorn & Hogarth (1986). I could thus, for the first time, demonstrate that participants evaluate identical covariations experienced in a continuous free-operant paradigm differently, depending on the temporal assumptions participants bring to bear. Knowledge thus mediates covariation assessment in human causal induction. Proponents of the associative learning account of course would not deny the existence of knowledge in intelligent organisms, but their approach limits them considerably as to how knowledge could be acquired and represented: "...experience is stored as a small number of associative strengths. ... information about past events is lost in the computation. In other words, these models do not have episodic memory."(Baker et al., 1996 p.1).

The prospects for associative learning need not be so grim, however. A meta-analysis of a range of data in the animal learning literature (Gallistel & Gibbon, 2000), for instance, has shown that animals react to changes in rates,

acquire information about the timeframe of causal relations, and use this information as a basis for their behaviour. Gallistel and Gibbon's rate-based account thus could explain how prior experience of delayed relations improves subsequent performance on similarly delayed relations, without the need to draw on complex conceptions of causal mechanism as envisioned by Einhorn and Hogarth (1986). Could such a low-level psychophysical account explain the results in our experiments? One can certainly raise the point that participants in Experiments II and III who received instructions about the delay actually *experienced* the delayed relation repeatedly in the instruction phase. This could have sparked some low-level learning about the timeframe of the relation, which in turn may have helped subsequent assessment of delayed relations. The order effect I found in Experiment II would fit in nicely with such an account: the beneficial effect of delay instructions was drastically reduced if participants experienced a contiguous relation between the delay instructions and working on the delayed scenario. Other related studies done by me (Buehner & Hagmayer, 2001) have shown that evaluations of delayed target contingencies change dramatically depending on whether participants experienced contiguous or delayed contingencies in a prior priming phase. Participants evidently acquired specific notions about the timeframe of candidate relations and applied those notions to the evaluation of the target contingencies.

However, Experiments V and VI also demonstrated the Knowledge Mediation effect in the absence of specific prior experience; in these experiments, the instructions simply *described* an immediate (light switch -> light bulb) or delayed (grenade launcher -> explosion) causal mechanism. This difference between cover stories alone lead to different interpretations of exactly the same evidence sampled from Free-Operant procedures. The most parsimonious explanation for the whole range of results then seems to be that knowledge about the timeframe of the relation in question (be it acquired via experience or description) influences the parsing of causal episodes.

Knowledge Mediation and associative learning theory need not necessarily be mutually exclusive accounts, however. It may well be that the former operates on top of the latter. In such a unified framework, knowledge would define the timeframe over which events were to be classified as co-occurrences, while the actual event parsing algorithm could be associative in nature (although of a radically different nature than most conditioning accounts; causal learning algorithms need to be symbol manipulating, see e.g. Holyoak & Hummel, 2000; Cheng & Buehner, 2000).

A similar argument holds for the power PC theory (Cheng, 1997), which is a computational level description (Marr, 1982) of rational causal inference(cf. section 1.4). This theory takes covariational data as its input, but makes no assumptions or restrictions about where the data is sampled from (trial-by-trial learning, summary data, described situations, etc.). It is thus evident that the power PC theory on its own makes no predictions about the influence of delay or knowledge-mediation. Rather, assumptions about the timeframe of the causal relation in question would determine whether a specific occurrence of an effect would be attributed to the candidate cause in question, or to alternative causes. Once this allocation has taken place, the power PC theory specifies how the evidence is interpreted.

6.4. Direct Detrimental Effects of Delay on Causal Judgment? From Free-Operant to Classical Conditioning Procedures

Towards the beginning of this thesis I presented an apparent paradox between experimental results and untutored everyday causal inference. I have tried to resolve the paradox and have shown that what is paradoxical is that experimenters have tried to apply what appear to be oversimplified conceptions about causal induction to describe an evidently complex computational problem. My experiments have demonstrated that there could have been several reasons why participants in previous causal induction

experiments reported degraded judgments of causal strength if cause and effect were separated by a delay. Experiments II through VI gave clear evidence that the impact of delay depends on people's assumptions about the timeframe of the causal relation in question. Experiment II has additionally shown that an action-outcome delay in a free operant procedure may result in weaker objective evidence for a causal relation by lowering the actual values of P(e|c). When I took efforts to guarantee stable values of P(e|c) across delays in Experiment III, the impact of delay in the instruction group was no longer significant.

The logical question then is, whether cause-effect delays actually do exert a direct detrimental influence on causal reasoning at all. Previous reports claiming the existence of such direct detrimental effects of delay may have employed procedures that confounded delay with weaker evidence, or presented participants with a mismatch between expectations and experience (or a combination of the two). Shanks and Dickinson (1987) have already identified that one result of delay might be to weaken the subjective evidence. They reasoned as follows:

> Presumably delaying an outcome makes it possible that the subjects will classify this outcome as being one that occurred in the absence of the action rather than in conjunction with it. This classification serves to decrease the subjects' estimates of ΔP both by reducing their perceived value of P(O/A) and by enhancing that of P(O/-A). [P(e|c) and P(e|¬c), respectively] (Shanks & Dickinson, 1987, p.234).

Shanks and Dickinson developed a procedure aimed at testing their argument. Participants had to judge how effective the pressing of certain keys was in producing an outcome on the screen. In contrast to the paradigm employed in their other paper (Shanks et al., 1989), participants now could alternate between pressing one of two keys (A3 and A4). Both keys produced the outcome with P(e|c)=.75, but A4 did so only after a 4s delay. P(e|¬c) was set

to .25 (presumably defined relative to a 1s time-bin, although this is not explicitly mentioned). In line with their other findings, participants evaluated pressing the A4 key as significantly less causal than pressing A3. Shanks and Dickinson interpreted this finding as evidence *against* the above argument:

> If the effect of contiguity is mediated by a change in the perceived contingency, we should expect to have observed a decrement in the judgments not only for A4 but also for A3. ...delaying the outcome for A4 does not just decrease the perceived P(O/A4) but also correspondingly increases P(O/-A4). As these delayed outcomes were unlikely to have occurred in close association with A3, they should also have served to increment P(O/-A3), thus reducing the perceived contingency for A3 as well as A4. (Shanks & Dickinson, 1987 p. 235)

Unfortunately, Shanks and Dickinson did not provide an analysis of the behavioural data that would corroborate their claim. There is no reason to believe that participants did not attribute the (delayed) outcomes they produced by pressing A4 to their subsequent presses of A3. Under such an interpretation, delaying the outcome for A4 would still have decreased the perceived P(O/A4), but crucially, it would also have *increased* the perceived P(O/A3). Without an analysis of the response-outcome structure, Shanks & Dickinson's contention remains an untested speculation. It is, however, weakened by the fact that participants overestimated the causal effectiveness of A3 (see Figure 3 in Shanks & Dickinson, 1987 p.236), both in comparison to two identical control conditions (A1, A2) which were not paired with a delayed condition, and relative to the objective contingency. The overestimation of A3 suggests that participants did in fact over-estimate P(O/A3), contrary to Shanks and Dickinson's claims.

Be that as it may, Shanks and Dickinson's (1987) argument was one about the *subjective* experience derived from the evidence at hand. While this is an important problem, the ambiguity in both Shanks and Dickinson's and my data from Experiment III shows that controlling the subjective experience

participants derive from Free-Operant procedures may not be as simple as one would like. Experiment II showed, however, that delays in a standard free-operant procedure (as employed in Shanks et al., 1989, Experiment 1; and Reed, 1992, Experiments 1 and 2) result in weaker *objective* evidence for the causal relation in question, which is of course a far more serious problem. Analysis of the behavioural data from experiments II and III have shown that a Free Operant procedure may not be the best paradigm to systematically study the influence of delay on human causal induction; one easily falls into the trap of providing weaker objective or subjective evidence for the causal relation when introducing a delay.

A more promising approach could be to apply methods borrowed from classical conditioning. Rather than asking participants to interact with the apparatus (instrumental paradigm), one would program the apparatus to display candidate causes and effects according to some programmed schedule; participants would be first asked to observe the evidence, and then to derive causal estimates from it. The evidence could take the form of a movie where some events regularly follow one another. The experimenter could still vary the essential parameters contingency and contiguity. Such a strategy would preserve ecological validity by allowing the repeated presentation of information (and thus still falls within the scope of associative learning theory), but would also permit the experimenter full control over the stimulus presentation; the apparatus could readily be programmed to provide stable evidence for the causal relation across delays. Future research, based on such new paradigms that fully eliminate any confounds of delay with weaker evidence for the relation in question, may allow even stronger tests of the Knowledge Mediation hypothesis.

6.5. From Probabilities to Rates

The experiments reported in this thesis combined and extended earlier studies investigating the influence of delay in human causal induction. As in Schlottmann's (1999) developmental study, I manipulated participants' expectations about the timeframe of the causal relation in question. To allow a full comparison with the predictions of the associative learning account, I employed a paradigm similar to the one used in Shanks et al.'s (1989) seminal study, amended with a few improvements. For reasons of ecological validity I abolished the notion of pre-defined learning trials. This of course made any implementation of $P(e|\neg c)$ other than 0 impossible, because conditional probabilities are always defined relative to some event. While it was still possible to implement a probability that a given response triggered an outcome, it was no longer possible to implement a probability that "no response" triggered an outcome, because there was no pre-defined length of time that could be classified as a period of no response. Consequently, it is no longer possible to refer to action-outcome pairings in general as manifestations of a contingency schedule in the traditional sense.

This was not a problem for the experimental conditions in Experiments II through VI because the outcome never occurred unless it was triggered by a response, so the contingencies were in fact identical across all experimental conditions. The experimental conditions deliberately employed $P(e|\neg c)=0$, because my focus was on observing the influence of delay in its purest form, in conditions where causal power would be easiest to infer. I therefore thought it best not to place extra computational workload (dedicated to the discounting of effects caused by the background) on the participants.

However, future research in a continuous paradigm such as the one I used may wish to address more complex situations, ones that also allow a systematic examination of the impact of effect base-rates greater than zero. A probabilistic notion of causality does not really apply to such scenarios, so

current theories based on such notions (including associative learning) will need to be superseded by a framework that embraces the continuous paradigm more adequately. Such a framework will probably be based on *rates* rather than *probabilities* (for a rate-based account of conditioning see Gallistel & Gibbon, 2000). Another important limitation of probabilities is that they are restricted by an upper bound of 1, whereas rates are unconstrained (except by perceptual limitations of the reasoner, see Gallistel, 1990; and Cheng, 1997). This problem particularly applies to evaluations of generative causes. Consider a task commonly employed in causal reasoning experiments (Wasserman et al., 1993, p.176):

> To implement the ... contingencies, I defined 1-s sampling intervals. If the subject tapped the telegraph key at least once at any time during the sampling interval, the light [effect] occurred with the conditional probability of P(O|R) at the end of the interval; otherwise, the conditional probability of the light was P(O|NoR).
>
> Both conditional probabilities in Wasserman et al.'s experiment ranged from .00 to 1.00. There is, however, no reason why participants should know that the rate was artificially limited to one effect per second. The light could conceivably have flashed more often than that. Applying a probabilistic framework to such continuous paradigms thus may produce drastically different conceptions of task characteristics between experimenters and participants. Anderson and Sheu (1995) have shown that humans indeed base their causal judgments on rates rather than probabilities, when sampling evidence from a continuous paradigm.

Analogously to the probabilistic contrast $P(e|c) - P(e|\neg c)$ used in contemporary causal reasoning theories (see chapter 1), a rate-based framework could involve a contrast of the effect-rate conditional on the presence or absence of the candidate cause, $R(e|c) - R(e|\neg c)$ (c.f. Anderson & Sheu, 1995). Regardless of whether one adopts a probabilistic or a rate-based

notion of causality, however, knowledge about the timeframe of the causal relation in question will always be critical for the computation of an adequate contrast, as the reasoner needs to have some basis for deciding whether a particular effect should be attributed to the candidate or to alternative causes.

6.6. Knowledge-based Causal Induction

I have demonstrated that different beliefs about the timeframes of causal relations result in different interpretations of identical covariations. Prior knowledge mediates causal inference in many other interesting ways. Michael Waldmann, for example, showed that different assumptions about causal models (predictive vs. diagnostic learning) likewise result in different interpretations of identical covariations (see Waldmann, 1996 for an overview). Both his work and the experiments presented here demonstrated an interaction between bottom-up (covariation assessment) and top-down (knowledge mediation) components in causal induction. Associationism is at a loss explaining such interactions, as it disallows any influence of knowledge (be it assumptions about structure or delay) beyond pre-existing associations. The causal power approach, although it accounts for knowledge mediation, suffers circularity: it cannot explain how knowledge is acquired in the first place. Cheng's (1997) power PC theory combines bottom-up and top-down components, and suggests that all aspects of causal knowledge can ultimately be derived from observation. Lien & Cheng (2000), for instance, demonstrated that humans are able to derive abstract categories of causal and non-causal entities from experienced covariations, and use this category knowledge when classifying novel objects as genuine or spurious causes. Power assumptions that distinguish causal from non-causal covariations thus are not shrouded in mystery or innateness but can themselves be inferred from covariation.

What lies ahead is to determine how people acquire assumptions about the timeframe of causal relations (other than extracting them from experimental instructions). Analogously to Lien & Cheng's findings, future research may show humans to be capable of deriving surprisingly well-formulated temporal assumptions from statistical information. Once again David Hume can inspire our search: "In vain, therefore, should we pretend to determine any single event, or infer any cause and effect, without the assistance of observation and experience" (1777/1902 p.30).

References

Ahn, W.-K., Kalish, C. W., Medin, D. L., & Gelman, S. A. (1995). The role of covariation vs. mechanism information in causal attribution. *Cognition, 54*, 299-352.

Allan, L. G., & Jenkins, H. M. (1980). The judgment of contingency and the nature of response alternatives. *Canadian Journal of Psychology, 34*(1), 1-11.

Anderson, J. R., & Sheu, C. F. (1995). Causal inferences as perceptual judgments. *Memory and Cognition, 23*(4), 510-524.

Baker, A. G., Murphy, R. A., & Vallée-Tourangeau, F. (1996). Associative and normative models of causal induction: Reacting to versus understanding cause. In D. R. Shanks & K. J. Holyoak & D. L. Medin (Eds.), *Causal Learning* (Vol. 34, pp. 1- 45). San Diego, CA: Academic Press.

Baker, A. G., Vallee Tourangeau, F., & Murphy, R. A. (2000). Asymptotic judgment of cause in a relative validity paradigm. *Memory and Cognition, 28*(3), 466-479.

Buehner, M. J., & Cheng, P. W. (1997). Causal induction: The power PC theory versus the Rescorla-Wagner model. In M. G. Shafto & P. Langley (Eds.), *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society* (pp. 55-60). Hillsdale, NJ: Erlbaum.

Buehner, M. J., Cheng, P. W., & Clifford, D. (2001). From covariation to causation: A test of the assumption of causal power. *Manuscript submitted for Publication.*

Buehner, M. J., & Hagmayer, Y. (2001). Temporal distance between cause and effect: The role of prior experience in non-contiguous causal induction. *Manuscript in Preparation.*

Bullock, M., Gelman, R., & Baillargeon, R. (1982). The development of causal reasoning. In W. J. Friedman (Ed.), *The developmental psychology of time* (pp. 209-254). New York: Academic Press.

Chapman, G. B., & Robbins, S. J. (1990). Cue interaction in human contingency judgment. *Memory and Cognition, 18*(5), 537-545.

Cheng, P. W. (1993). Separating causal laws from casual facts: Pressing the limits of statistical relevance. In D. L. Medin (Ed.), *The psychology of learning and motivation. Advances in research and theory* (Vol. 30, pp. 215-264). San Diego, CA, USA: Academic Press, Inc.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review, 104*(2), 367-405.

Cheng, P. W., & Buehner, M. J. (2000). Covariation need not imply causation: The neccessity of causal power in accounts of causal induction. *Manuscript submitted for Publication.*

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments and Computers, 25*(2), 257-271.

Dickinson, A. (2001). Causal Learning: An associative analysis. *Quarterly Journal of Experimental Psychology Section B- Comparative and Physiological Psychology, 54*(1), 3-25.

Einhorn, H. J., & Hogarth, R. M. (1986). Judging probable cause. *Psychological Bulletin, 99*(1), 3-19.

Gallistel, C. R. (1990). *The organization of learning*. Cambridge, MA: MIT Press.

Gallistel, C. R., & Gibbon, J. (2000). Time, rate, and conditioning. *Psychological Review, 107*(2), 289-344.

Garcia, J., Ervin, F. R., & Koelling, R. A. (1966). Learning with prolonged delay of reinforcement. *Psychonomic Science, 5*(3), 121-122.

Garcia, J., & Koelling, R. A. (1966). Relation of cue to consequence in avoidance learning. *Psychonomic Science, 4*(3), 123-124.

Garner, W. R. (1974). *The processing of information and structure*. Potomac, MD: Lawrence Erlbaum Associates.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102*(4), 684-704.

Hagmayer, Y., & Waldmann, M. R. (2001). Testing complex causal hypotheses. In M. May & U. Oestermeier (Eds.), *Interdisciplinary Perspectives on Causation*. Norderstedt, Germany: Libri.

Holyoak, K. J., & Hummel, J. E. (2000). The proper treatment of symbols in a connectionist architecture. In E. Dietrich & A. Markman (Eds.), *Cognitive Dynamics: Conceptual change in humans and machines.* (pp. 229-263). Mahwah, NJ: Erlbaum.

Howell, D. C. (1997). *Statistical methods for psychology* (4th ed.). Belmont, CA: Wadsworth.

Hume, D. (1739/1888). A treatise of human nature. In L. A. Selby-Bigge (Ed.), *Hume's treatise of human nature*. Oxford, UK: Clarendon Press.

Hume, D. (1777/1902). An enquiry concerning human understanding. In L. A. Selby-Bigge (Ed.), *Hume's Enquiries*. Oxford, England: Clarendon Press.

Jenkins, H., & Ward, W. (1965). Judgment of contingencies between responses and outcomes. *Psychological Monographs, 7*, 1-17.

Kamin, L. J. (1969). Predictability, surprise, attention and conditioning. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behavior*. New York: Appleton Century Crofts.

Kant, I. (1781/1965). *Critique of pure reason*. London: Macmillan.

Lien, Y. W., & Cheng, P. W. (2000). Distinguishing genuine from spurious causes: A coherence hypothesis. *Cognitive Psychology, 40*(2), 87-137.

Lober, K., & Shanks, D. R. (2000). Is causal induction based on causal power? Critique of Cheng (1997). *Psychological Review, 107*(1), 195-212.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information.* San Francisco: Freeman.

Melz, E. R., Cheng, P. W., Holyoak, K. J., & Waldmann, M. R. (1993). Cue competition in human categorization: Contingency or the Rescorla-Wagner Learning Rule? Comment on Shanks (1991). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(6), 1398-1410.

Mendelson, R., & Shultz, T. R. (1976). Covariation and temporal contiguity as principles of causal inference in young children. *Journal of Experimental Child Psychology, 22*(3), 408-412.

Michotte, A. E. (1946/1963). *The perception of causality* (T. R. Miles, Trans.). London, England: Methuen & Co.

Miller, R. R., Barnet, R. C., & Grahame, N. J. (1995). Assessment of the Rescorla-Wagner model. *Psychological Bulletin, 117*(3), 363-386.

Over, D. E., & Green, D. W. (2001). Contingency, causation, and adaptive inference. *Psychological Review, 108*(3), 682-684.

Pearce, J. M. (1987). A model for stimulus generalization in Pavlovian conditioning. *Psychological Review, 94*(1), 61-73.

Perales, J. C., & Shanks, D. R. (2000). Normative and descriptive accounts of the influence of power and contingency on causal judgments. *submitted for publication.*

Piaget, J. (1969). *The child's conception of physcial causality.* Totowa, N.J.: Littlefield, Adams.

Reed, P. (1992). Effect of a Signaled Delay Between an Action and Outcome On Human Judgment of Causality. *Quarterly Journal of Experimental Psychology Section B- Comparative and Physiological Psychology, 44B*(2), 81-100.

Reed, P. (1999). Role of a stimulus filling an action-outcome delay in human judgments of causal effectiveness. *Journal of Experimental Psychology-Animal Behavior Processes, 25*(1), 92-102.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovial conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical Conditioning II: Current theory and research* (pp. 64-99). New Yowk: Appleton-Century Crofts.

Schafe, G. E., Sollars, S. I., & Bernstein, I. L. (1995). The CS-US interval and taste aversion learning: A brief look. *Behavioral Neuroscience, 109*(4), 799-802.

Schlottmann, A. (1999). Seeing in happen and knowing how it works: How children understand the relation between perceptual causality and underlying mechanism. *Developmental Psychology, 35*(5), 303-317.

Schustack, M. W., & Sternberg, R. J. (1981). Evaluation of evidence in causal inference. *Journal of Experimental Psychology: General, 110*, 101-120.

Shanks, D. R. (1985). Continuous Monitoring of Human Contingency Judgment Across Trials. *Memory & Cognition, 13*(2), 158-167.

Shanks, D. R. (1987). Acquisition Functions in Contingency Judgment. *Learning and Motivation, 18*(2), 147-166.

Shanks, D. R. (1991). Categorization By a Connectionist Network. *Journal of Experimental Psychology-Learning Memory and Cognition, 17*(3), 433-443.

Shanks, D. R. (1993a). Associative Versus Contingency Accounts of Category Learning - Reply. *Journal of Experimental Psychology-Learning Memory and Cognition, 19*(6), 1411-1423.

Shanks, D. R. (1993b). Human Instrumental Learning - a Critical-Review of Data and Theory. *British Journal of Psychology, 84*, 319-354.

Shanks, D. R., & Dickinson, A. (1987). Associative Accounts of Causality Judgment. In G. H. Bower (Ed.), *Psychology of Learning and Motivation-Advances in Research and Theory* (Vol. 21, pp. 229-261). San Diego, CA: Academic Press.

Shanks, D. R., Holyoak, K. J., & Medin, D. L. (Eds.). (1996). *The psychology of learning and motivation (Vol.34):Causal Learning*. San Diego, CA: Academic Press.

Shanks, D. R., & Lopez, F. J. (1996). Causal order does not affect cue selection in human associative learning. *Memory & Cognition, 24*(4), 511-522.

Shanks, D. R., Lopez, F. J., Darby, R. J., & Dickinson, A. (1996). Distinguishing associative and probabilistic contrast theories of human contingency judgment. In D. R. Shanks & K. J. Holyoak & D. L. Medin (Eds.), *Causal Learning* (Vol. 34, pp. 265-311). San Diego, CA: Academic Press.

Shanks, D. R., Pearson, S. M., & Dickinson, A. (1989). Temporal Contiguity and the Judgment of Causality By Human Subjects. *Quarterly Journal of Experimental Psychology Section B- Comparative and Physiological Psychology, 41*(2), 139-159.

Shimazaki, T., Tsuda, Y., & Imada, H. (1991). Strategy changes in human contingency judgments as a function of contingency tables. *Journal of General Psychology, 118*(4), 349-360.

Siegler, R. S., & Liebert, R. M. (1974). Effects of contiguity, regularity, and age on children's causal inferences. *Developmental Psychology, 10*(4), 574 579.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124-1131.

Wagner, A. R., Logan, F. A., Haberlandt, K., & Price, T. (1968). Stimulus selection in animal discrimination learning. *Journal of Experimental Psychology, 76*, 171-180.

Waldmann, M. R. (1996). Knowledge-based causal induction. In D. R. Shanks & K. J. Holyoak & D. L. Medin (Eds.), *Causal Learning* (Vol. 34, pp. 47-88). San Diego, CA: Academic Press.

Waldmann, M. R. (2000). Competition among causes but not effects in predictive and diagnostic learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(1), 53-76.

Waldmann, M. R., & Hagmayer, Y. (1999). How categories shape causality. In M. Hahn & S. C. Stoness (Eds.), *Proceedings of the Twenty-first Annual Conference of the Cognitive Science Society* (pp. 761-766). Mahwah, NJ: Erlbaum.

Waldmann, M. R., & Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models: Asymmetries in cue competition. *Journal of Experimental Psychology: General, 121*(2), 222-236.

Waldmann, M. R., & Holyoak, K. J. (1997). Determining whether causal order affects cue selection in human contingency learning: Comments on Shanks and Lopez (1996). *Memory & Cognition, 25*(1), 125-134.

Wasserman, E. A., Elek, S. M., Chatlosh, D. L., & Baker, A. G. (1993). Rating causal relations: Role of probability in judgments of response-outcome contingecy. *Journal of Experimental Psychology:Learning, Memory, and Cognition, 19*, 174-188.

Wasserman, E. A., & Neunaber, D. J. (1986). College students' responding to and rating of contingency relations: The role of temporal contiguity. *Journal of the Experimental Analysis of Behavior, 46*(1), 15-35.

White, P. A. (1995). Use of prior beliefs in the assignment of causal roles: Causal powers versus regularity-based accounts. *Memory and Cognition, 23*(2), 243-254.

Wu, M., & Cheng, P. W. (1999). Why causation need not follow from statistical association: Boundary conditions for the evaluation of generative and preventive causal powers. *Psychological Science, 10*(2), 92-97.

Appendices

Appendix A Scenarios for Questionnaire preceding Experiments IV through VI

Q1. Imagine you are an officer in a military training range. Your job is to test ammunition. How much time would you expect to pass between firing off a shell and being able to see the explosion in the training range?

Q2. How much time do you expect to pass between flicking a light switch and the light going on?

Q3. Imagine you are at a pedestrian crossing. How much time do you expect to pass between you pressing the button and the light signal turning green?

Q4. Imagine you are an engineer. Your job is to test an infrared remote control for a simple device located in the same room as you. How much time do you expect to pass between pressing the button on the remote control and the device to operate?

Q5. Imagine you are a journalist reporting back to Europe from Australia. You have set up a satellite link both to send recordings to your studio in Europe and to see on TV in "real-time" what your colleagues in Europe receive from you. How much time do you expect to pass between you sending a signal and being able to see it on your TV?

Q6. Imagine you are standing in front of an elevator. How much time do you expect to pass between pressing the button and the elevator to arrive?

Q7. How much time do you expect to pass between pressing a key on the computer keyboard and the corresponding character to appear on the screen?

Q8. How much time do you expect to pass between pressing the button of a camera and being able to hear the shutter click?

Q9. How much time do you expect to pass between pushing the button of a doorbell and the bell ringing?

Q10. Imagine you are searching a database on the internet. How much time do you expect to pass between clicking on a button (e.g. "submit") and the corresponding action becoming visible on the screen?

Appendix B Revised Scenarios for Questionnaire

Q1. Imagine you are an officer in a 5 by 5 mile military training range. Your job is to test a grenade launcher. How much time would you expect to pass between launching a missile and being able to see the explosion in the training range?

Q4. Imagine you are an engineer. Your job is to test an infrared remote control for a TV located in the same room as you. How much time do you expect to pass between pressing the button on the remote control and TV to operate?

Q5. Imagine you are a journalist reporting back to Europe from Australia. You have set up a satellite link to send recordings to your studio in Europe. On your TV you will see what your colleagues in Europe receive from you. How much time do you expect to pass between you sending a signal from Australia to Europa and being able to see it on your local TV back in Australia?

Q10. Imagine you are looking for information on the internet. How much time do you expect to pass between performing an action (e.g. clicking on a button) and the consequences (e.g. the button lighting up) becoming visible on the screen?

Appendix C General Instructions for Experiments IV through VI

In this experiment you have to evaluate the extent to which your actions can cause something to happen. There will be a button on the computer screen and your task is to observe whether clicking it causes something to happen on the screen.

You can choose at any time whether or not to click the button. You can click it as often or as little as you like. However, because of the nature of the task it is to your advantage to click it some of the time and not to click it some of the time.

The effectiveness of you clicking the button stays the same within a particular condition but may well vary between problems.

At the end of each problem you will be asked whether clicking the button causes the outcome and if so, how strongly it causes the outcome.

You will work on two different scenarios with five problems per scenario, each lasting about two minutes.

Appendix D Specific Instructions for Light bulb scenario

In the upcoming five problems there will be a lightbulb and a lightswitch on the screen. Your task is to judge the extent to which clicking the switch causes the bulb to light up.

Imagine that the lightbulb is connected to the switch you can click on and to a switch in another room that other persons can flick without you being aware of it. Thus, if the bulb lights up, it may be because you clicked the switch or because a person in the other room flicked the second switch.

You can choose at any time whether or not to click the switch. You can click it as often or as little as you like. However, because of the nature of the task it is to your advantage to click it some of the time and not to click it some of the time.

You will work on five different problems with the lightbulb, each lasting for 2 minutes. The relationship between your clicking the switch and the bulb lighting up will be constant within each problem but may well differ from one problem to the next.

At the end of each problem you will be asked whether and how strongly clicking the lightswitch makes the bulb light up.

Appendix E Specific Instructions for Grenade scenario

In the upcoming five problems you will view a military training range from a command post several miles away. Your task is to find out whether clicking on a "FIRE!" button produces explosions in the range. Imagine that the "FIRE!" button operates a grenade launcher which is situated in your post and fires grenades into the training range. When a grenade you've fired hits the training range, you will see an explosion. However, an officer in another post is also firing into the range. Thus if you see an explosion, it may be because you clicked the "Fire!" button or because the second officer in the other post launched a grenade.

You can choose at any time whether or not to click on "Fire!". You can click it as often or as little as you like. However, because of the nature of the task it is to your advantage to click it some of the time and not to click it some of the time.

You will work on five different problems with the grenade launcher, each lasting for 2 minutes. The relationship between your clicking "Fire!" and explosions in the range will be constant within each problem but may well differ from one problem to the next.

At the end of each problem you will be asked whether and how strongly clicking "FIRE!" causes explosions in the range.

Appendix F Rating Instructions in Experiment VI

1 Light bulb Scenario

We will now ask you how clicking the switch affects whether or not the bulb lights up.

Please use the rating scale on the bottom for your judgment.

0 means that clicking the switch has **no influence** on whether or not the bulb lights up,

i.e. the bulb lighting up is completely independent from your clicking

50 means that clicking the switch **moderately causes** the bulb to light up.

i.e. half of your clicks make the bulb light up

100 means that clicking the switch **strongly causes** the bulb to light up,

i.e. everytime you click the switch, the bulb lights up.

Please use values in between if your estimate is between two numbers.

What effect does your clicking the switch have on the bulb lighting up?

2 Grenade Scenario

We will now ask you how clicking the switch affects whether or not explosions happen.

Please use the rating scale on the bottom for your judgment.

0 means that clicking the switch has **no influence** on whether or not explosions happen,

i.e. the explosions are completely independent from your clicking

50 means that clicking the switch **moderately causes** explosions,

i.e. half of your clicks produce explosions

100 means that clicking the switch **strongly causes** explosions,

i.e. everytime you click the switch, an explosion happens.

Please use values in between if your estimate is between two numbers.

What effect does your clicking the switch have on the explosions?