Illusory Correlation and Valenced Outcomes

by

Cory Derringer

BA, University of Northern Iowa, 2012

MS, Missouri State University, 2014

Submitted to the Graduate Faculty of

The Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2019

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Cory Derringer

It was defended on

April 12, 2019

and approved by

Timothy Nokes-Malach, Associate Professor, Department of Psychology

Julie Fiez, Professor, Department of Psychology

David Danks, Professor, Department of Psychology, Carnegie Mellon University

Thesis Advisor/Dissertation Director: Benjamin Rottman, Associate Professor, Department of Psychology

Illusory Correlation and Valenced Outcomes

Cory Derringer, PhD

University of Pittsburgh, 2019

Accurately detecting relationships between variables in the environment is an integral part of our cognition. The tendency for people to infer these relationships where there are none has been documented in several different fields of research, including social psychology, fear learning, and placebo effects. A consistent finding in these areas is that people infer these illusory correlations more readily when they involve negative (aversive) outcomes; however, previous research has not tested this idea directly. Four experiments yielded several empirical findings: Valence effects were reliable and robust in a causal learning task with and without monetary outcomes, they were driven by relative rather than absolute gains and losses, and they were not moderated by the magnitude of monetary gains/losses. Several models of contingency learning are discussed and modified in an attempt to explain the findings, although none of the modifications could reasonably explain valence effects.

Table of Contents

List of Tables

List of Figures

List of Equations

Preface

Over the past five years I have come to realize that earning a PhD is very much a team sport. I would like to express my deepest gratitude to the people who made this project possible. This dissertation would never have been completed without the support and mentorship of my advisor, Dr. Ben Rottman. I would also like to thank my committee members—Dr. Tim Nokes-Malach, Dr. Julie Fiez, and Dr. David Danks—for their thoughtful advice, guidance, and feedback.

Additionally, I would like to extend my sincere thanks to my colleagues in the Causal Learning and Decision Making Lab—Dr. Kevin Soo, Ciara Willett, and Zac Caddick—for keeping me sane and (mostly) motivated throughout my time at Pitt.

Although this journey culminated at Pitt, it began a long time ago in a state far, far away. I cannot begin to express my thanks to my parents, Rodney and Lori Derringer, for their bottomless patience and support. Similarly, this dissertation would not exist without the endless encouragement (and occasional pep talk) from my wife, Kelsey Derringer.

Finally, I may not have pursued a doctorate in the first place if not for very sound advice from my grandfather, Charles Wooldridge. This work is dedicated to his memory; he always wanted me to be a professional student.

1.0 Introduction

Distinguishing genuine relationships between two events from statistical noise is something people do every day, and is an important part of how we make many of our decisions. At a small scale, one might believe that the line at a coffee shop is longer on rainy days. In this example, the stakes are relatively low. If the two are correlated, one could plan around this inconvenience. If they are not, no one is really harmed outside of the time lost to unnecessary planning. However, people do not only form contingency beliefs about benign relationships such as this one. Sometimes we fail to form contingency beliefs at our peril (e.g., if a patient stops taking their medication because they fail to notice the effect, or because it has a delayed effect).

The present work focuses on the opposite situation; the formation of a belief that two events or variables are related even when they are not related. These illusory correlations can also manifest on a small scale (e.g., if a sports fan believes in their lucky jersey) or on a large scale with more nefarious outcomes (e.g., the false belief that immigrants commit crimes at a higher rate than native born citizens).

The tendency to infer illusory correlations has been documented in various forms across a variety of fields such as causal learning (see Matute et al., 2015 for a review), placebo effects (e.g., Au Yeung, Colagiuiri, Lovibond, & Colloca, 2014; Colagiuiri, Quinn, & Colloca, 2015), fear learning (Pauli, Montoya, & Martz, 1996; 2001), and stereotype formation in social psychology (e.g., Acorn, Hamilton, & Sherman, 1988; Hamilton & Gifford, 1976; Mullen & Johnson, 1990).

Some of the earliest and most widely replicated effects in illusory correlations come from the domain of social stereotyping research (e.g., Acorn et al., 1988; Hamilton, Dugan, & Trolier, 1985; Hamilton & Gifford, 1976; Schaller & Maass, 1989). In a typical design, participants are

shown trial-by-trial data regarding the states of two variables that are in fact uncorrelated (e.g., Table 1). For example, Hamilton and Gifford showed participants a number of statements in which a fictional person was a member of a social group, and exhibited a prosocial or antisocial behavior (e.g., "John, a member of Group A, visited a sick friend in the hospital."). They then asked participants at the end of the dataset to judge the relationship between group membership and behavior.

Table 1. Frequencies of each combination of cause (Group A/B) and effect (Prosocial/Antisocial Behavior) variables in typical Negative-Distinctive and Positive-Distinctive illusory correlation datasets.

	Negative-Distinctive				*Positive-Distinctive*		
	Prosocial (common)	Antisocial (rare)	Total		Antisocial (common)	Prosocial (rare)	Total
Group A (common)	24	12	36	Group A (common)	24	12	36
Group B (rare)	8	4	12	Group B (rare)	8	4	12
Total	32	16	48	Total	32	16	

Using the data in Table 1 as an example, subjects typically infer an illusory correlation between group membership and attribute such that the common group is associated with the common attribute and the rare group with the rare attribute. Furthermore, and critical for the current work, these illusory correlations are usually stronger if the rare attribute is has negative valence (Negative-Distinctive; ND) than if it has positive valence (Positive-Distinctive; PD). Throughout the rest of the present work, this pattern of stronger illusory correlations for negative than positive outcomes will be called a valence effect. The overall goal of this work was to

systematically test valence effects in a more precise and detailed way than has been done previously.

The next sections contain a brief review of evidence from multiple research domains suggesting the existence of a valence effect in illusory correlations. This is followed by the goals of the current research, and the specific research questions it is intended to answer. Four experiments will be discussed, along with several theories of contingency learning. The relevant models from these theories will be modified to see if they can account for valence effects.

2.0 Background on Valence Effects

2.1 Valence Effects across Sub-Fields of Psychology

Previous work in several domains has either directly or indirectly examined the role of valence in illusory inferences. In the social psychology literature, several empirical studies have elicited illusory correlation inferences with both positive and negative outcomes (e.g., Hamilton & Gifford, 1976; Schaller & Maass, 1989). Mullen and Johnson (1990) conducted a meta-analysis and found that illusory correlation effects were larger when the rare outcome had a negative valence (e.g., John, a member of Group A, visited a friend in the hospital) than when it had positive valence (e.g., Tim, a member of Group B, is rarely late to work). The pattern whereby illusory correlations effects are larger when the rare outcome is negative is the valence effect of interest in the current work. However, none of the studies in Mullen and Johnson's review directly compared positive and negative outcomes to test for valence effects. Further, I am unaware of any illusory correlation study since their review that has directly examined valence effects.

Researchers in the fear learning literature have also examined patterns similar to valence effects in illusory correlation. For example, there is evidence that people infer stronger illusory correlations between a neutral stimulus and an aversive outcome when the outcome is more strongly aversive (e.g., Wiemer, Mühlberger, & Pauli, 2014). However, this finding with small vs. moderate negative outcomes is different than the ones above that compared positive vs. negative outcomes, and how this finding would map onto positive vs. negative outcomes is unclear. It is possible that for outcomes with a physiological component, more extreme positive (appetitive) or negative (aversive) outcomes trigger stronger illusory correlations symmetrically. Another piece

of evidence relevant to valence comes from studies of fear learning that compare illusory correlations among people with high vs. low fear (e.g., spider phobia). These studies have found that high fear individuals form stronger illusory correlations between fear-relevant stimuli (e.g., pictures of spiders) vs. non-fear-relevant pictures and aversive outcomes (e.g., a shock vs. no shock) (e.g., De Jong, Merckelbach, & Arntz, 1995; Tomarken, Mineka, & Cook, 1989; Tomarken, Sutton, & Minenka, 1995; Wiemer & Pauli, 2016). This is further indirect evidence of valence effects; presumably fear-relevant stimuli are more aversive for phobic individuals, which could explain why those individuals form stronger illusory correlations.

A third domain in which valence effects are relevant is the literature on placebo and nocebo effects. Whereas a placebo effect involves a belief that a benign stimulus is beneficial, a nocebo effect involves a belief that it is harmful. In a new paradigm for studying placebo/nocebo effects using a trial-by-trial learning format similar to the paradigms in social psychology and fear learning, participants first learn a real contingency between the cue and the outcome, and later the contingency turns to zero. For example, Au Yeung et al. (2014) induced a placebo effect using a sham device put on a patient's arm that is said to be able to moderate the intensity of pain resulting from an electric shock. Participants completed an acquisition phase during which the activation of the sham device was paired with weaker shocks, teaching them that the device reduced their pain. After a number of learning trials, the experiment transitioned into a test phase in which this contingency was secretly eliminated; the shock was always strong, regardless of whether or not the device was active. The placebo effect in studies such as this is measured in time to extinction: how long does it take participants to realize that the device no longer reduces the strength of the shock?

Colagiuiri et al. (2015) conducted a similar study regarding nocebo effects. The procedure was very similar, with the key difference being a reversed contingency between the device and the shock; in the acquisition phase the device activation was accompanied by more intense shocks. Colagiuiri et al. found that participants' nocebo beliefs did not extinguish by the end of the test phase of the study. Though these two effects have not been directly compared statistically, they provide indirect evidence that nocebo effects are more difficult to extinguish than placebo effects, which could be viewed as a type of valence effect. Furthermore, although placebo and nocebo effect sizes are typically similar in magnitude (Petersen et al., 2014), there is some tentative evidence that nocebo effects are easier to induce. Whereas placebo effects typically require an acquisition phase in which the placebo is initially correlated with reduced pain, simply telling participants that a cue amplifies pain is enough to instill a nocebo effect (e.g., Colloca et al., 2008). This pattern is similar to the valence effects found in the social psychology literature, as well as the findings in the fear learning literature; people form stronger illusory correlation inferences related to negative outcomes (nocebo) compared to positive outcomes (placebo).

Illusory correlations have also been studied in the causal learning domain (e.g., Matute et al., 2015). As in the other domains, valence effects have not been directly manipulated. In the field of causal learning, there are studies of relationships between variables that use positive outcomes (e.g., plants blooming or not; Spellman, Price, & Logan, 2001), negative outcomes (e.g., headache vs. no headache; Cheng 1997; Matute & Blanco, 2014), or neutral outcomes (e.g., levels of neurotransmitters in the brain; Rottman & Hastie, 2016), but there has been no investigation of the valence effect – whether illusory causal inferences are stronger when the rare outcome is negative.

Another piece of indirect evidence comes from an 'illusory control' paradigm, which is closely related to illusions of causality except in an illusion of control the subject believes that

their actions make a difference to an outcome. Aeschleman, Rosen, and Williams (2003) asked participants to press keys on a computer keyboard to change words on a computer screen. Participants in the positive condition were tasked with making the word "GOOD" appear on the screen and keeping it there for as long as possible. Participants in the negative condition were tasked with preventing the word "BAD" from appearing on the screen, and making it go away if it did appear. In reality the appearance of the words were not related to the buttons the participants pressed. Participants in one of the negative conditions gave higher ratings of control. This provides more indirect evidence of a valence effect; although Aeschleman et al.'s effect is very clear, valence was technically confounded with another aspect of the design; participants were asked to prevent the negative word but to produce the positive word.

2.2 Explanations for Valence Effects

2.2.1 Negativity bias – salience

It has been argued that people exhibit an overall negativity bias in which they attend to negative outcomes more than positive ones (e.g., Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001; Rozin & Royzman, 2001; Vaish, Grossman, & Woodward, 2008). Baumeister et al. argued that this negativity bias could stem from an evolutionary advantage for giving preferential attention to negative events in relation to positive ones. After all, the limitations of positive and negative events in the world are asymmetrical; no positive event can eliminate the possibility of all future negative events, but a sufficiently extreme negative event (i.e., death) can eliminate future positive events. Previous research in behavioral economics, specifically the loss aversion effect (Kahneman

& Tversky, 1979, 1984) also provides evidence for this kind of negativity bias. Specifically, people feel monetary losses more acutely than gains; imagine the joy of finding a $20 bill while walking down the street vs. the pain of realizing that you accidentally lost $20. In sum, if people attend more to negative outcomes, perhaps this attention exacerbates the illusory correlation effect.

2.2.2 Salience might not explain the valence effect

Though the increased salience of negative events feels like an intuitive explanation for the valence effect in illusory correlation, there are a number of reasons to be skeptical of this hypothesis. First, when learning about the relation between a cue and outcome and there actually is a genuine contingency, people learn the contingency faster when the outcome is negative compared to when it is positive (e.g., Fazio, Eiser, & Shook, 2004). According to associative learning theories, this finding could be explained by increased salience – a higher learning rate parameter (e.g., Rescorla & Wagner, 1972). The problem for the salience account of valence effects is that many models of associative learning predict that increased salience would produce less illusory correlation due to faster learning that there is not a relation. In Section 8 I attempt to explain the valence effect with computational models, especially exploring the role of salience.

There is some evidence that anxiety plays an important role in how we process negative events. Several studies the in the placebo/nocebo domain have found that high state anxiety is related to larger nocebo effects (e.g., Colloca, Petrovic, Wager, Ingvar, & Benedetti, 2010) and smaller placebo effects (e.g., Morton, Watson, El-Deredy, & Jones, 2009). Further, Benedetti and colleagues have found evidence that the link between expectations of pain and hyperalgesia is mediated by cholecystokinin (CCK), and that CCK antagonists can reduce or even eliminate nocebo effects (e.g., Benedetti, Amanzio, Casadio, Oliaro, & Maggi, 1997; Benedetti, Amanzio,

Vighetti, & Asteggiano, 2006; Colloca & Benedetti, 2007). Similarly, Johnston, Atlas, and Wager (2012) found that expectancy for pain (i.e., anxiety) enhanced the nocebo hyperalgesia effect. While anxiety is helpful in explaining some positive/negative asymmetries in placebo/nocebo studies, this is probably related to the relatively extreme nature of the stimuli used in these studies. Previous research suggests that people become anxious when contemplating imminent pain (e.g., Quartana, Campbell, & Edwards, 2009). However this level of negative valence is not present in the illusory correlation literature, nor in the experiments in the present work. It is difficult to imagine that participants are experiencing increased state anxiety in anticipation of reading sentences with negative information. While anxiety may play a role in how people reason about positive and negative events generally, it is probably not related to valence effects within the scope of the current work.

2.2.3 The importance of controlling outcome magnitude

One methodological difficulty when measuring and explaining the valence effect for illusory correlation is equalizing the objective magnitude of the positive vs. negative outcome. Kahneman and Tversky's loss aversion (1979) was demonstrated with gains and losses of monetary prospects, which made it easy to have the same monetary gain vs. loss. However, studies that have tested illusory correlation have used outcomes without such objective magnitudes and therefore make them more difficult to compare. In the case of social stereotyping studies, it is difficult to isolate the salience that comes from the valence of attributes and behaviors from the salience that comes from their extremity. There is no precise way to say that describing a person as "rarely late for work" (positive attribute) is more or less extreme compared to "always talks about himself and his problems" (negative attribute) (Hamilton & Gifford, 1976, p. 394). The same

problem applies in a different way to Andreatta and Pauli's (2015) comparison of conditioning with positive and negative outcomes. One of their findings was that participants' learned associations between a neutral cue and an outcome took longer to extinguish if the outcome was negative (a shock) than positive (a food reward); it is unclear whether a shock and a food reward can be compared on equal footing.

The present studies disentangle valence from magnitude using monetary gains and losses as well as other comparisons. By using monetary outcomes, the objective positive/negative magnitude can be held constant while the valence (positive or negative) varies with money being given/taken.

3.0 Open Questions

The current research addresses several empirical questions. First, do people infer a stronger illusory correlation for negative outcomes when the negative and positive outcomes are equal in magnitude? This question is addressed in Experiments 1 and 3.

Second, do illusory correlations broadly, and valence effects particularly, stem from overestimations of the rarest combination (distinctiveness) or the most common combination? If participants' illusory correlation inferences come from overweighting the rarest combination, and if this tendency is stronger in the ND condition than the PD condition, that would provide evidence that people differentially process rare events depending on valence. This pattern would have implications for the role of salience in valence effects, which will be discussed in Section 8. This question is addressed in Experiments 1-3.

Third, what counts as sufficiently negative or positive to produce a valence effect? Will valence effects be induced through stories about hypothetical negative vs. positive outcomes as well as through gains and losses of small amounts of money, and does combining the two produce larger valence effects? Although monetary gains and losses have been used in other fields such as behavioral economics, they are not typically employed in illusory correlation studies. This question is addressed in Experiment 1.

Fourth, are valence effects driven by absolute or relative gains vs. losses? In Table 1, the negative (antisocial) outcome is negative on an absolute scale relative to neutral whereas the positive (prosocial) outcome is positive on an absolute scale. However, in Table 1, the negative outcome is also more negative than the positive outcome in a relative comparison. For this reason,

when comparing negative vs. positive outcomes, it is unclear if the valence effect is driven merely by a relative comparison of better vs. worse or by an absolute comparison of good vs. bad.

Though valence effects have previously been discussed as due to the rare outcome being absolutely good vs. bad, the literature on reference dependence suggests that a relative comparison might actually drive valence effects. Reference dependence is one of the fundamental ideas in Kahneman and Tversky's Prospect Theory (1979). For example, if an individual receives some amount of money, their interpretation of whether that windfall is a gain or a loss depends on their expectations. If the money was a surprise (i.e., the person was not expecting to receive any money), or if the amount was higher than expected, the person interprets this as a gain. However, if the amount was lower than expected, even a positive gain can be interpreted as a loss. The combination of reference dependence and loss aversion predicts several now-famous psychological phenomena such as endowment effects by which people overvalue their own possessions in relation to others (e.g., Kahneman et al., 1990) and framing effects by which people respond differently to prospects framed as gains or losses (Tversky & Kahneman, 1981). For this reason, Experiment 2 will assess whether the valence effect is due to absolute or relative gains vs. losses.

Fifth, do stronger positive and negative outcomes lead to larger valence effects compared to weaker positive and negative outcomes? If salience is the source of valence effects (i.e., if anything that makes the rare outcome more salient will increase illusory correlations), then larger, more salient gains/losses would yield larger valence effects. This question will be studied in Experiment 3.

Sixth, are there also valence-related differences in the extinction of learned contingencies? Specifically, if participants learn a real contingency between a stimulus and an outcome, and the contingency then changes to zero, will it take longer for participants to learn the new contingency

if the outcome is negative than if it is positive? This question approaches the phenomenon of valence effects from the paradigm of placebo vs. nocebo effects; there is indirect evidence that it takes longer to extinguish nocebo effects than placebo effects (e.g., Au Yeung et al., 2014; Colagiuiri et al., 2015). This question is studied in Experiment 4.

The last goal of the current work is to examine the extent to which current theories can explain—or can be modified to explain—valence effects. After the four experiments, I describe several theories of causal learning that have been used to explain illusory correlation effects. None of the theories can account for valence effects in their original form. However, it is possible that some of them can be adapted to explain valence effects.

4.0 Experiment 1: Story Valence vs Monetary Valence

In Experiment 1, the comparative impacts of monetary and non-monetary outcomes in valence effects were examined. This was accomplished by manipulating the modality of valence presentation to form three groups: a Combined Valence group, a Story group, and a Monetary group. In previous illusory correlation experiments, researchers have often used cover story valence. This experiment may be the first illusory correlation study to examine this phenomenon using purely monetary outcomes.

4.1 Method

4.1.1 Participants

Participants (n = 234, 98 female) were recruited through MTurk, and their average age was 36.25 years (SD = 11.25). Participants were paid a base rate of $3.50 for their participation. In addition, they were paid bonuses for accuracy (M = $2.18, SD = $0.21).

4.1.2 Design and cover stories

The design was a mixed factorial, with valence (PD vs ND) manipulated within subjects and valence modality manipulated between subjects. Participants in the Combined group experienced both story and monetary outcomes. On each trial, they received positive/negative cover story information (e.g., they were told that their patient had a good outcome, and saw a

picture of a smiling face) and their bonus amount was adjusted upward or downward by six cents. Participants in the Story group viewed the same visual stimuli for positive/negative outcomes as participants in the Combined group, but these were not accompanied by monetary rewards or punishments. Finally, participants in the Monetary group gained and lost six cents for good and bad outcomes, respectively; however, they did not receive the cover story relating to patients in a hospital setting. They were told that the task was to find out if there were relations between shapes (instead of faces) and colors (instead of medications). Each shape was indicative of either a good or a bad outcome. For example, for one participant a square might always be accompanied by a six-cent increase in their bonus. In this condition, participants learned whether colors were associated with the "good" or "bad" shapes. (See Table 2.)

Table 2. Datasets from each condition in Experiment 1. Monetary payouts indicated by +/-.

		Negative-Distinctive			Positive-Distinctive	
Combined Valence		Good (+6¢)	Bad (-6¢)		Bad (-6¢)	Good (+6¢)
	Drug 1	24	12	*Drug 3*	24	12
	Drug 2	8	4	*Drug 4*	8	4
Story		Good (0¢)	Bad (0¢)		Bad (0¢)	Good (0¢)
	Drug 1	24	12	*Drug 3*	24	12
	Drug 2	8	4	*Drug 4*	8	4
Monetary		Star (+6¢)	Triangle (-6¢)		Oval (-6¢)	Square (+6¢)
	Red	24	12	*Purple*	24	12
	Blue	8	4	*Green*	8	4

4.1.3 Procedure

Participants were given different instructions depending on their valence modality. Participants in the Combined and Story conditions were told that the study was about how people learn about medications.

> For this task, imagine that you work in a hospital, and your job is to make sure that the patients have GOOD outcomes after treatment. You will see two scenarios, each containing information about patients who were treated for a fictional disease. You are responsible for finding out if the different medications are effective or not. Keep in mind that different diseases have different success rates for treatment. Before seeing data about each disease, you will be told which treatment is the most common, and whether the outcomes for the disease are typically GOOD or BAD.

Participants in the Monetary condition were told that the study was about learning the relationships between shapes and colors:

> For this task, you will learn about the relationships between shapes and colors. You will see two scenarios, each containing information about two shapes and two colors. You are responsible for finding out if there are relationships between the shapes and colors. (For example, when the shape is RECTANGLE, is the color more likely to be ORANGE or YELLOW?) Keep in mind that some shapes are more common than others, and some colors are more common than others. Before seeing the shape/color combinations, you will be told which shapes and colors are the most common.

Participants then completed a brief training session for the task. In the Combined and Monetary conditions, they were shown that good/bad outcomes were accompanied by monetary rewards/punishments. In the Story condition, the instructions were the same as the Combined condition, but the good/bad patient outcomes were not accompanied by monetary rewards/punishments.

Before each of the two scenarios, participants were shown a briefing screen which showed which value of each variable was more common. To reinforce this, participants were shown the

possible values of each variable (e.g., good and bad treatment outcomes) and had to click on the more common value of each variable before beginning the experiment.

Participants then completed a trial-by-trial learning phase and a test phase in the first scenario, followed by learning and test phases in the second scenario. On each trial in the learning phase, participants were given the value of the outcome variable (treatment outcome/shape), and then guessed about the value of the cue (drug/color) (Figure 1). To support accurate learning, participants could only advance after each trial by clicking on the correct value of the second variable.

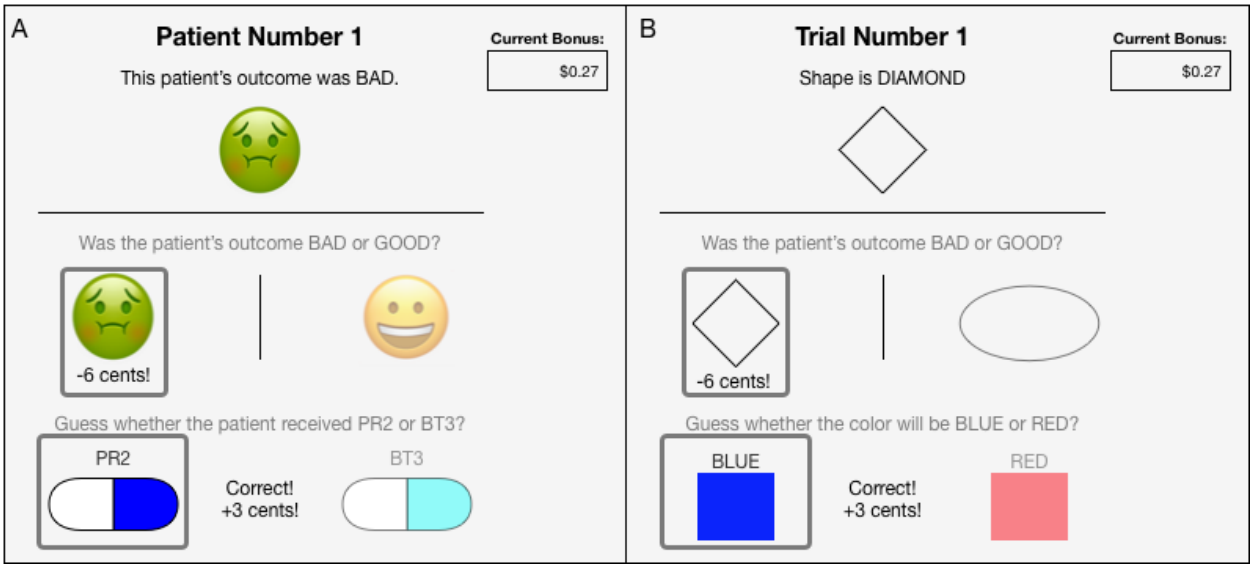

Figure 1. Feedback in trial-by-trial learning phase of Experiment 1, Combined (A) and Monetary (B) conditions. The Story condition was identical to the Combined condition, except for the monetary outcomes on each trial (e.g., "-6 cents").

The bonusing scheme was as follows. Participants began the first scenario with a $0.30 bonus. In the Monetary and Combined conditions, money was added to/taken from participants' bonuses on each trial, in accordance with the trial's outcome. Additionally, participants could earn

accuracy bonuses for correctly guessing the cue. If participants correctly guessed the cue, they were given an additional 3-cent accuracy bonus in the feedback phase of the trial. The bonus carried over between scenarios.

Participants were told that it was possible for their bonus amounts to be negative, but that most people finish the study with a positive bonus amount. They were also told that if they finished with a negative amount it would be rounded up to zero. In reality, it was not possible to finish the experiment with a negative bonus, because the PD and ND conditions were perfectly symmetrical; participants began the study with $0.30, and ended with $0.30 plus their combined accuracy bonuses.

After completing the learning phase, participants advanced to the test phase. The test phase contained three tasks for participants. Two of the tasks were memory-related, and one was a modified causal judgment. In previous studies (e.g., Eder et al., 2011) participants have been given the marginal distribution of the groups, and asked to fill in the distribution of outcomes within each group. Participants answered a similar question in the test phase of Experiment 1: "Of the patients you just saw, 12 received PR2. How many of them had BAD/GOOD outcomes?" (See Figure 2A.) Participants then entered their best estimate of BAD/GOOD outcomes for PR2, and did the same for the other drug (e.g., "Of the patients you just saw, 12 received PR2. How many of them had BAD/GOOD outcomes?"). The left/right locations of bad and good outcomes and medications were randomized at the participant level. This kind of memory judgment will subsequently be called an Outcome|Cue (or O|C) estimate, because participants are reasoning about the outcome (bad or good) with information about the cue (i.e., that the patient received PR2).

One potential problem with these O|C judgments is that in the learning phase participants were given the outcome and made predictions about the cue (Figure 1). If participants have difficulty with the mental switch in the direction of reasoning, that could presumably cause illusory correlations. For this reason, Cue|Outcome (C|O) memory estimates were also included, such as, "Of the patients you just saw, 16 had BAD outcomes. How many of them received PR2/BT3?" (Figure 2B). Whether participants were first asked O|C or C|O memory items was randomized at the subject level. The order within each question (e.g., whether they were asked about PR2 or BT3 first in the O|C judgment) was randomized at the scenario level.

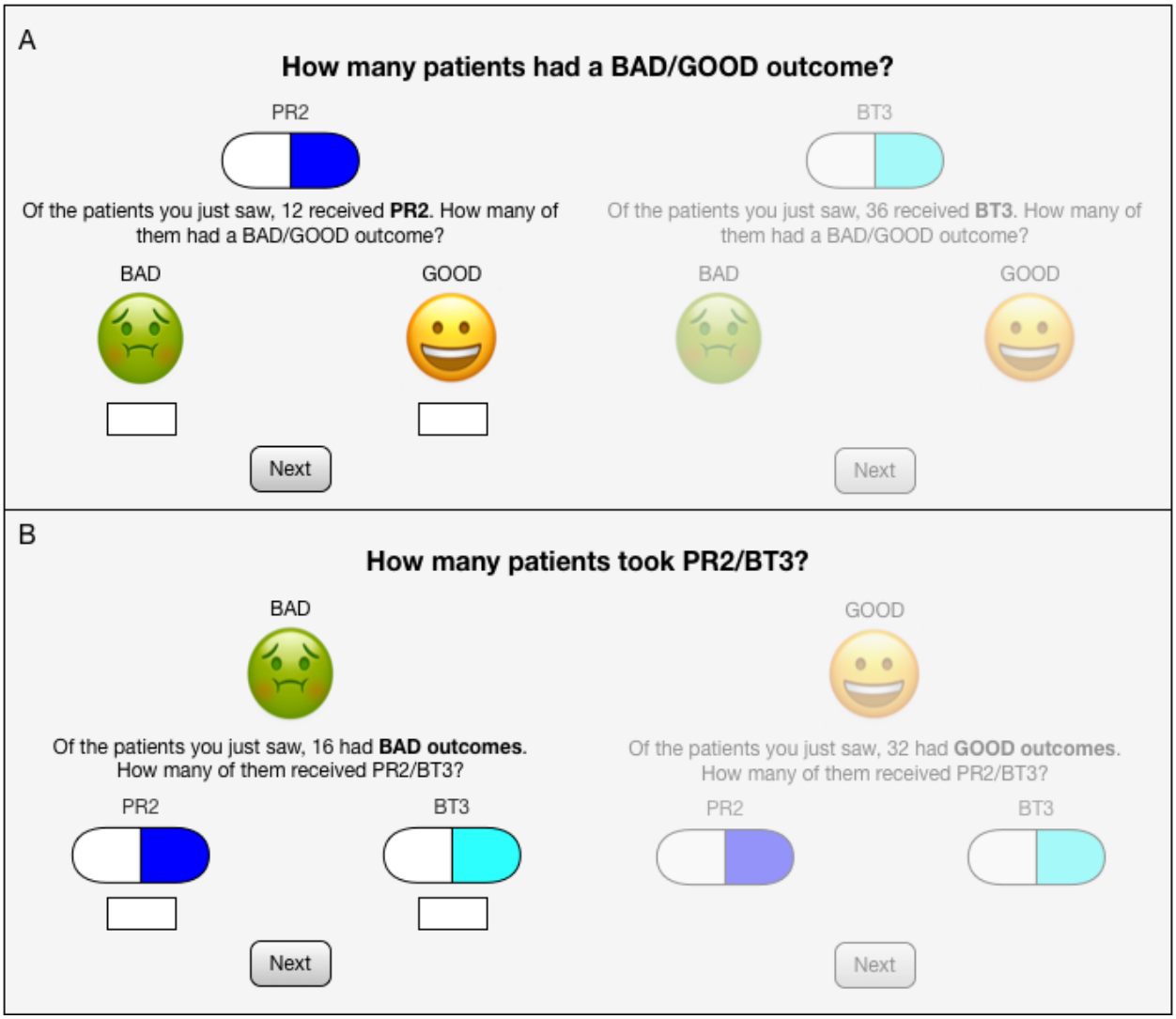

Figure 2. Outcome|Cue (A) and Cue|Outcome (B) memory judgments in the test phase of Experiment 1.

Consistent with previous studies, these C|O and O|C judgments were converted to phi correlations prior to analysis (Equation 1).

$$\phi = \frac{AD-BC}{\sqrt{(A+B)(C+D)(A+C)(B+D)}} \qquad \textbf{Equation 1}$$

Phi correlations are equivalent to Pearson's r for binary variables (Davenport & El-Sanhurry, 1991). Table 3 shows the a set of trials equivalent to half of a dataset in Experiment 1, with a phi coefficient of zero. Table 4 shows the same data in long form; Pearson's r for the data in Table 4 is zero.

Table 3. Example dataset in which phi = 0.

	Good	Bad
PR2	12 (A)	6 (B)
GS5	4 (C)	2 (D)

Table 4. Data from Table 3 in long format; r = 0. Shading corresponds to cells A-D.

	A											B						C				D	
Patient Number:	1	2	3	4	5	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Drug (1 = PR2):	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0
Outcome (1 = Good):	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	1	1	1	1	0	0

Illusory correlation studies typically rely on phi coefficients from these remembered cell frequencies (i.e., O|C estimates) rather than asking participants about their causal beliefs. This approach makes sense in the context of judging groups of people (e.g., Hamilton & Gifford, 1976), because a social desirability bias might make participants reluctant to endorse explicit judgments about groups. However, in the context of medications and patient outcomes there is no reason to avoid a more direct measure of illusory correlation than remembered O|C or C|O frequencies.

After completing O|C and C|O estimates, subjects indicated their beliefs about which drug was better for patient outcomes (Figure 3). Participants were given a forced choice task regarding which medication to prescribe to a new patient, and their bonus was adjusted depending on that

patient's outcome. This methodology was chosen because it incentivized learning and accurate judgments. The judgment was presented as a gambling task: participants would gain $1.00 if the patient had a good outcome and lose $1.00 if the patient had a bad outcome. Participants could then click to choose which drug to prescribe. After clicking a drug, they were given confirmation that they had chosen that drug, and asked to adjust a slider indicating their confidence about their choice. The anchors on each end were "I am very confident about [Drug]," and the text in the middle of the slider said "I am not sure at all." Participants were not given the option to select a slider value incompatible with their selection in the forced choice. A back button was available if they wanted to select the other drug before finalizing their slider judgment. After finalizing the sliding scale judgment, participants were given feedback about their patient's outcome. Despite what was implied in the instructions, the outcomes were deterministic in accordance with the outcome base rates. In the PD condition the outcome was always negative; in the ND condition it was always positive. Because valence was manipulated within subjects, these "gambles" cancelled out for every participant. Because the causal judgment task was essentially an additional learning trial, it was always presented after the memory judgments. Before analysis the causal strength judgments were recoded such that more positive numbers indicate a causal strength judgment in the predicted direction (i.e., that the rare cue causes the rare outcome or that the common cue causes the common outcome).

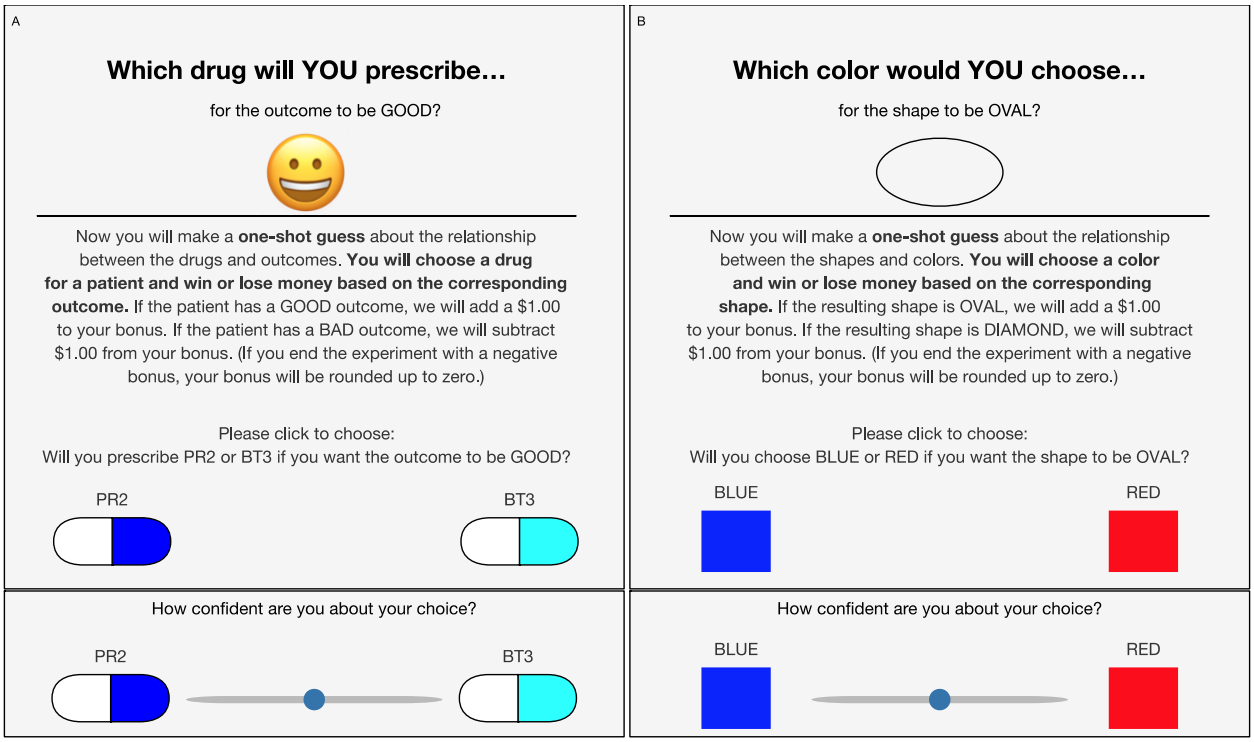

Figure 3. Causal judgment task in the Combined/Story (A) and Monetary (B) conditions. Sliding scale below replaced choice when participants chose a drug/color.

After two scenarios, participants completed a brief demographics questionnaire and were directed to a debriefing page. The entire procedure took approximately 25 minutes to complete.

4.2 Results

There were several predicted patterns. First, it was predicted that participants' illusory correlations in the ND condition would be significantly positive for all three outcomes. Second, it was predicted that these illusory correlations would be stronger in the ND condition than the PD condition (valence effects). Third, if participants form stronger illusory correlations in the ND condition, this should also be reflected in their trial-by-trial predictions. Finally, if illusory

correlations are driven by distinctiveness, and valence effects are the result of increased distinctiveness when the rarest combination involves a negative outcome, participants should tend to overweight the occurrence of the rarest outcome relative to the most common outcome in their O|C and C|O memory estimates.

4.2.1 Illusory correlations in memory and causal judgments

The expected pattern held very clearly; participants' illusory correlation inferences were all significantly above zero in the ND conditions (Figure 4). Single sample t-tests showed that participants' O|C estimates were significantly positive in the Combined ($t(78) = 6.54$, $p < .001$, $d = 0.74$), Monetary ($t(78) = 7.04$, $p < .001$, $d = 0.79$), and Story groups ($t(75) = 8.13$, $p < .001$, $d = 0.93$). Similarly, participants' C|O judgments were significantly positive in the Combined ($t(78) = 5.29$, $p < .001$, $d = 0.60$), Monetary ($t(79) = 3.56$, $p < .001$, $d = 0.40$), and Story groups ($t(75) = 6.94$, $p < .001$, $d = 0.80$). Finally, participants' causal strength judgments in the ND condition were significantly above zero in the Combined ($t(79) = 10.53$, $p < .001$, $d = 1.18$), Monetary ($t(80) = 8.48$, $p < .001$, $d = 0.94$), and Story groups ($t(78) = 10.25$, $p < .001$, $d = 1.15$).

4.2.2 Valence effects in memory and causal judgments

To test for a main effect of valence and examine whether valence effects interact with valence modality, a mixed ANOVA with Type III sums of squares was conducted for each outcome with valence as a within-subjects factor and valence modality as a between-subjects factor.

For the O|C memory judgments[1], there was a significant main effect of valence, $F(1, 231) = 52.57$, $p < .001$, $\eta^2_G = .09$. This valence effect was not modulated by a valence modality interaction, $F(2, 231) = 2.20$, $p = .11$, $\eta^2_G = .001$. Similar results were found in participants' C|O judgments. The main effect of valence was significant ($F(1, 232) = 21.32$, $p < .001$, $\eta^2_G = .04$), but the interaction was not, $F(2, 232) = 2.16$, $p = .12$, $\eta^2_G = .01$. The largest valence effects were found in participants' causal judgments, $F(1, 233) = 249.98$, $p < .001$, $\eta^2_G = .40$. The interaction was again nonsignificant, $F(1, 233) = 2.17$, $p = .11$, $\eta^2_G = .01$.

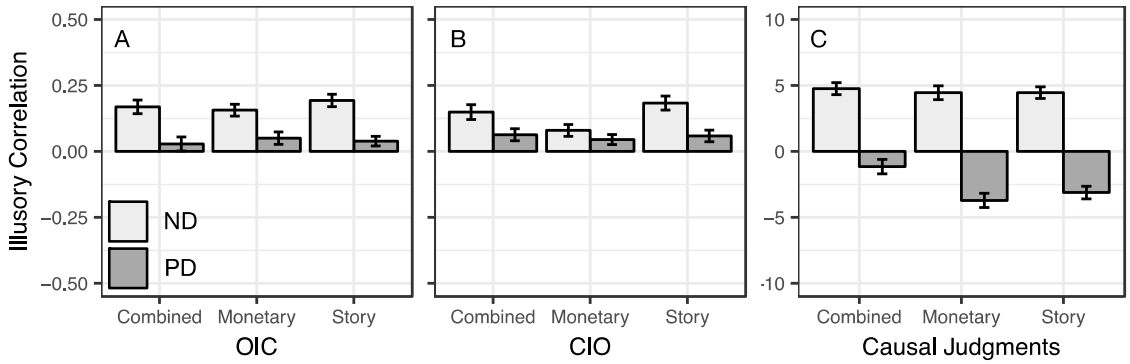

Figure 4. Illusory correlations for each condition and group, for each DV in Experiment 1. Error bars indicate standard error of the mean. Note that the y axes differ and are truncated in panels A and B to more clearly illustrate the effect.

[1] For both O|C and C|O memory judgments, there was an order effect corresponding to which memory question participants answered first. However, because the pattern of results collapsing across order the same as when only the first memory item was analyzed, I report the inferential tests collapsing across order.

4.2.3 Valence effects in predictions

Another way to measure illusory correlation is in participants' predictions during each trial. If participants believe that either drug is equally likely to cause a good/bad outcome, then they should choose the rare drug at roughly equivalent rates regardless of whether the patient had the common/rare outcome.

For each participant, the probability of choosing the rare drug was calculated for trials in which the patient had good and bad outcomes[2] (Figure 5). In this case, an illusory correlation would manifest as a difference between the probability of choosing the rare drug when the patient's outcome was common vs rare (i.e., the difference between the light and dark bars in Figure 5). A valence effect would be a larger difference in the ND condition than the PD condition. The differences were again compared using a mixed ANOVA with Type III sums of squares. The effect of valence was significant, $F(1, 234) = 21.19$, $p < .001$, $\eta^2_G = .06$, and the interaction with valence modality was again nonsignificant, $F(2, 234) = 1.77$, $p = .17$, $\eta^2_G = .01$.

[2] In Experiments 1-3, some of the trial-by-trial predictions were lost in the data collection process (approx. 0.2%). For participants missing two or fewer predictions (out of 96 total), the choice probability was simply calculated for the incomplete data, rather than discarding the data from these participants. Participants missing more than two predictions were excluded. For all three experiments, the results of the ANOVAs for trial-by-trial predictions were not changed by reporting this way rather than only including participants with complete data.

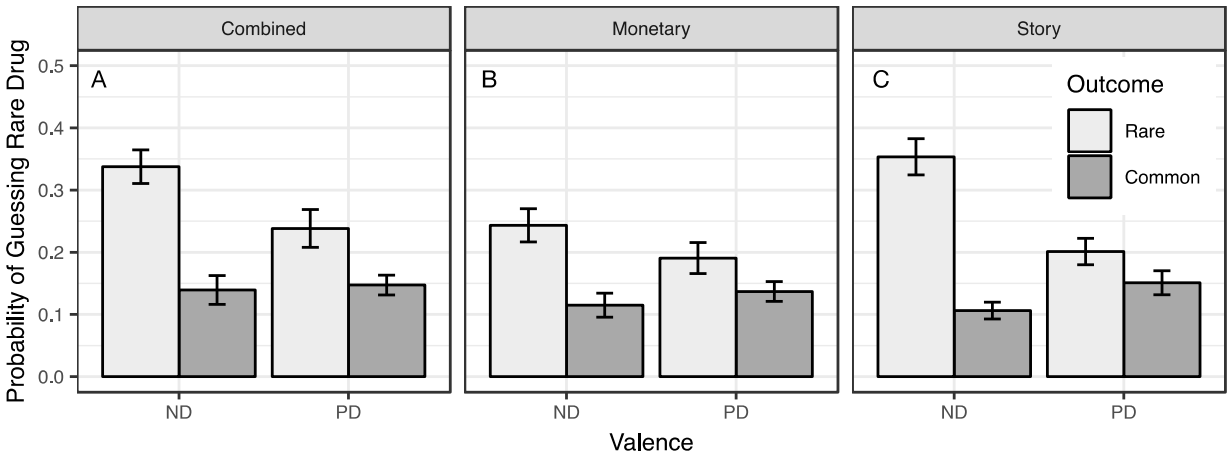

Figure 5. Probability of guessing rare drug when the patient experienced the common and rare outcomes in Experiment 1. Error bars indicate standard error of the mean.

4.2.4 Cell weighting

Finally, participants' reported frequencies for each combination were compared. The memory estimates above suggest that participants either over-estimate instances of the most common combination, the rarest combination, or both. If the distinctiveness of the rarest combination drives illusory correlations (and valence effects specifically), there should be a consistent pattern whereby participants overweight the occurrence of the rarest combination. This section first reports whether the rarest combination was overestimated generally, then examines the differences between ND and PD conditions. Then the same questions are examined for the most common combination.

First participants' overall C|O and O|C frequencies for the rarest combination were compared to the actual values using single sample t-tests. Participants significantly overestimated

the occurrence of the rarest combination in both their O|C ($t(467) = 15.75$, $p < .001$, $d = 0.73$) and C|O estimates ($t(469) = 18.15$, $p < .001$, $d = 0.84$).

Valence effects and interactions with modality were tested, again using a 2(Valence: ND vs PD, within subjects) x 3 (Modality: Combined vs Monetary vs Story) mixed factorial ANOVA with Type III sums of squares. For O|C judgments, there was a main effect of valence such that estimates of the rare combination (Figure 6C) were significantly higher in the ND condition than the PD condition ($F(1, 231) = 15.49$, $p < .001$, $\eta^2_G = .03$). The interaction was nonsignificant ($F(2, 231) = 1.23$, $p = .29$, $\eta^2_G = .01$). For C|O judgments, there was also a main effect of valence such that estimates of the rare combination (Figure 6D) were higher in the ND condition than the PD condition ($F(1, 232) = 24.13$, $p < .001$, $\eta^2_G = .04$). Here the interaction was significant ($F(2, 231) = 1.23$, $p = .29$, $\eta^2_G = .01$), although it is probably an anomaly given the nonsignificant interactions in the other analyses.

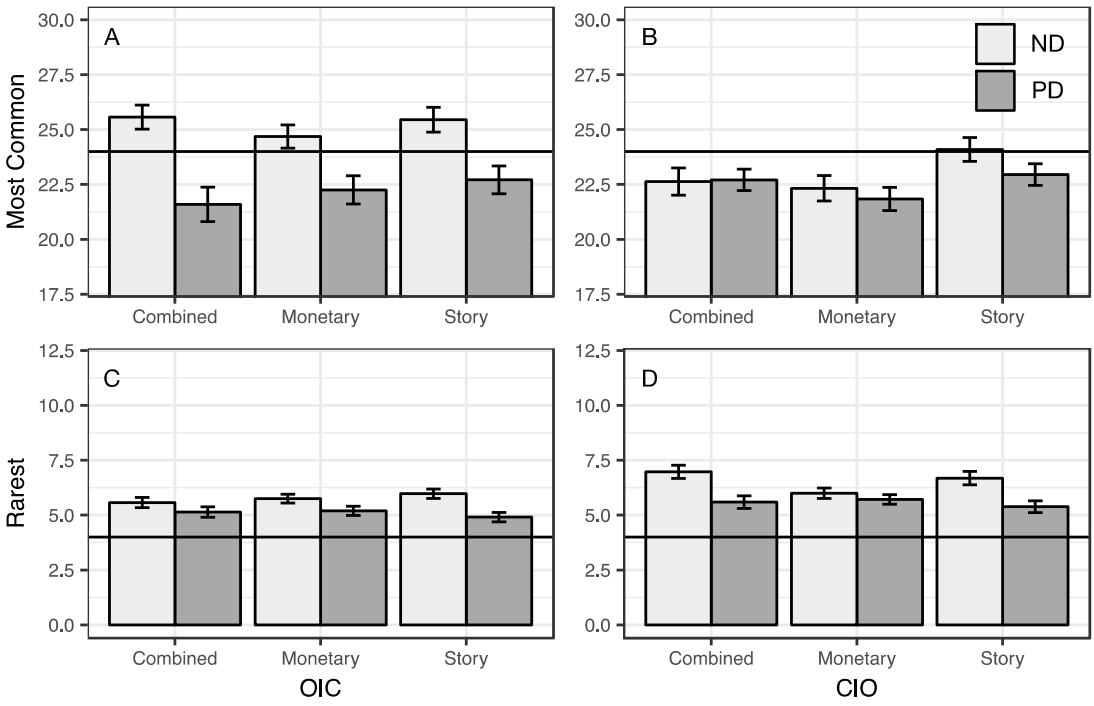

Figure 6. Average estimated frequencies of the most common (A,B) and rarest (C,D) cue/outcome combination, for O|C (A,C) and C|O (B,D) memory estimates in Experiment 1. Error bars indicate standard error of the mean. Horizontal lines indicate correct frequency.

Analyses of participants' estimates of the rare combination showed that participants consistently overestimated the rarest combination in their memory judgments, and that this tendency was stronger in the ND condition when that rare combination involved a negative outcome.

By contrast, participants did not overweight the most common combination in their O|C judgments (t(467) = -1.12, p = .26, d = 0.05), and actually underweighted the most common combination in their C|O judgments (t(469) = -5.62, p < .001, d = 0.26). Examining the ND/PD differences, with the same mixed ANOVA from above, participants had significantly higher O|C estimates of the most common combination (Figure 6A) in the ND condition than the PD condition (F(1, 231) = 39.41, p < .001, η^2_G = .07), and again the interaction was nonsignificant (F(2, 231) =

0.95, p = .39, η^2_G < .01). However, this pattern did not extend to participants' C|O judgments (Figure 6B), which did not differ between the ND and PD conditions (F(1, 232) = 1.90, p = .17, η^2_G < .01). The interaction with valence modality was nonsignificant (F(2, 232) = 0.87, p = .42, η^2_G < .01).

4.3 Discussion

In Experiment 1, I replicated the basic illusory correlation effect in nearly all conditions, and found evidence that this effect is driven by participants overestimating the occurrence of the rare cue/outcome combination. Additionally, I found a consistent valence effect whereby participants made stronger illusory correlation inferences when the rare outcome was negative than when it was positive.

Several findings from Experiment 1 were particularly notable. First, to my knowledge Experiment 1 is the first study to induce illusory correlations using purely monetary cue/outcome combinations. One might think that using real monetary outcomes that subjects receive as bonuses, rather than good or bad outcomes within a cover story, would lead to greater attention and better accuracy, and therefore an attenuated IC effect; however, the effect sizes for ICs in the monetary group were comparable to the other groups.

Second, participants' memory judgments suggest that their illusory correlation inferences were driven by an overestimation of the frequency of the rarest combination (Figure 6). The clear pattern is that participants tend to overestimate this rare combination, consistent with Hamilton and Gifford's (1976) original distinctiveness perspective. The general overestimation of rare outcomes is consistent with previous work (e.g., Arkes & Harkness, 1983; Kahneman & Tversky,

1979). However, the fact that participants' overestimation of the rare combination was stronger it the ND condition is not predicted by existing theories. This pattern is consistent with distinctiveness as a driver of illusory correlation, and valence as a marker of distinctiveness/salience. Implications of attention, distinctiveness, and negativity are explored further in Section 8.

Third, Experiment 1 yielded strong evidence for valence effects across all three valence modalities (Story, Monetary, and Combined groups), and with two different dependent measures (remembered frequencies and causal judgments). In some previous experiments it was possible to attribute valence effects to the wording of the memory question, because it drew attention to the negative outcomes explicitly. For example, Eder et al. (2011) told participants the total number of people in Group A and Group B, and asked how many statements from each group involved negative behaviors. Because participants in Experiment 1 did not answer these kinds of memory questions, but instead filled in contingency table information for positive and negative outcomes, it is clear that valence effects obtained in Experiment 1 cannot be an artifact of the test phase procedure; they must arise due to differences in learning or memory.

Finally, the patterns of valence effects in participants' memory and causal judgments were different. Participants gave IC judgments in the predicted direction for ND datasets. However, their judgments in PD datasets were actually in the opposite direction from what was predicted, which is why the valence effects for the causal strength judgments were so much larger than the valence effects in the frequency estimates. This pattern was found in Experiments 2-4 as well, and will be discussed in more detail in the General Discussion.

5.0 Experiment 2: Valence Effects and Loss Aversion

Hamilton and Gifford (1976) proposed the idea of distinctiveness in terms of frequencies. Rare events are said to be distinctive, and the co-occurrence of such rare events is even more distinctive. However, as previously discussed, a myriad of factors could conceivably contribute to how distinctive an experience is in memory. One such factor is whether an experienced outcome is positive or negative objectively; another conceivable factor is whether subjects encode the outcome as a gain or a loss.

The idea of reference dependence in behavioral economics can be traced at least as far back as Kahneman and Tversky's (1979) Prospect Theory. Kahneman and Tversky demonstrated not only that subjects switch from risk averse to risk seeking when facing the prospect of losses, they also showed that participants can become risk seeking in absolute gain scenarios that are framed as relative losses. The reverse has also been demonstrated; subjects can encode absolute monetary losses as relative gains, as when a purchase is less expensive than anticipated (e.g., Tereyağoğlu, Fader, & Veeraraghavan, 2017). Researchers have found that loss aversion is reference dependent in a wide variety of areas such as brand comparisons in a grocery store setting (e.g., Hardie, Johnson, & Fader, 1993), real estate prices (e.g., Genesove & Mayer, 2001), and professional sports (e.g., Pope & Schweitzer, 2011). If loss aversion broadly is reference dependent, is the same true for valence effects?

In the Combined condition of Experiment 1, absolute gain/loss and gain/loss framing were confounded; a positive outcome yielded a gain of 6 cents, and a negative outcome yielded a loss of 6 cents. The primary goal of Experiment 2 was to explore the relationship between absolute and relative gains and losses in the valence effects discussed thus far. Absolute gain/loss was

manipulated between subjects, and the relative gain/loss (PD vs ND distribution) was manipulated within subjects, as in Experiment 1. In the absolute Gain condition, participants were rewarded for good outcomes, but were not penalized for bad ones; in the Loss condition they were penalized for bad outcomes, but were not rewarded for good ones.

There is reason to think that valence effects would be particularly strong in the Loss condition. Experiment 1 provided evidence that people interpret negative outcomes as particularly salient, and it is not a stretch to imagine that trials in which a monetary reward/penalty occurs are more salient than trials in which no monetary change occurs. In other words, subjects may disproportionately attend to trials in which things happen, rather than those in which nothing changes (e.g., Kao & Wasserman, 1993). If this is the case, the rare-negative outcome in the ND condition would be much more salient in the Loss condition than in the Gain condition, resulting in an interaction. By contrast, if valence effects are driven by the experience of relative gains and losses, valence effects will be equivalent between the absolute Gain and Loss conditions.

5.1 Method

5.1.1 Participants

Participants (n = 157, 65 female, 91 male, 1 unreported) were recruited through MTurk. The average age was 35.46 years (SD = 9.69). Participants were paid a base rate of $3.50 for their participation, in addition to patient outcome bonuses ($2.88) and accuracy bonuses (M = $1.89, SD = $0.19).

5.1.2 Design

In order to test whether absolute or relative experience of gains/losses is the driving factor in valence effects, absolute Gain/Loss was manipulated between subjects (Table 5). Participants in the Gain condition received a bonus of 6 cents each time the patient outcome was good, and their bonus was not altered when the patient outcome was bad. By contrast, participants in the Loss condition lost 6 cents each time the patient outcome was bad, and their bonus was not altered when the outcome was good. As in Experiment 1, the base rates of good/bad outcomes was manipulated within subjects (Positive-Distinctive vs Negative-Distinctive). Thus Experiment 2 features a 2 (Absolute Bonus: Gain vs Loss, between subjects) x 2 (Valence: PD vs ND, within subjects) mixed factorial design. To make the total payout (minus accuracy bonuses) equal between the framing groups, participants in the Gain and Loss groups began the study with different patient outcome bonuses. Participants in the Gain group began with no money, whereas participants in the Loss group began with \$5.76. Both ended the study with \$2.88 in bonuses, in addition to trial-by-trial accuracy bonuses (same as Experiment 1).

Table 5. Datasets for Experiment 2.

	Negative-Distinctive			Positive-Distinctive		
Absolute Gain		+6¢	0¢		0¢	+6¢
(Start with \$0.00, end	*Drug1*	24	12	*Drug3*	24	12
with \$2.88.)	*Drug2*	8	4	*Drug4*	8	4
Absolute Loss		0¢	-6¢		-6¢	0¢
(Start with \$5.76, end	*Drug1*	24	12	*Drug3*	24	12
with \$2.88)	*Drug2*	8	4	*Drug4*	8	4

5.1.3 Procedure

Aside from the differences discussed above, the procedure was identical to the "Combined" cover story in Experiment 1. The entire procedure took approximately 25 minutes to complete.

5.2 Results

5.2.1 Valence effects in memory and causal judgments

Valence effects in memory and causal judgments were evaluated with a mixed factorial ANOVA with Type III sums of squares (Figure 7). For causal judgments, there was a significant valence effect ($F(1, 155) = 98.15$, $p < .001$, $\eta^2_G = .29$), but the main effect of absolute bonus ($F(1, 155) = 1.68$, $p = .20$, $\eta^2_G < .01$) and the interaction ($F(1, 155) = 0.34$, $p = .56$, $\eta^2_G < 01$) were both nonsignificant. Participants' O|C judgments yielded a similar result; the valence effect was significant ($F(1, 153) = 7.37$, $p < .01$, $\eta^2_G = .02$), but the main effect of absolute bonus ($F(1, 153) = 0.40$, $p = .53$, $\eta^2_G < .01$) and the interaction ($F(1, 153) = 2.62$, $p = .11$, $\eta^2_G < .01$) were not. Participants' C|O memory estimates[3] yielded no significant main effects for valence ($F(1, 69) =$

[3] There was a sizable order effect in participants' C|O memory estimates, in which the valence effect was present for participants who had already completed the O|C judgments, but not for those who completed the C|O judgments first. To eliminate the possibility of contamination from the O|C measurement, all relevant analyses exclude participants who completed the O|C judgments first.

0.01, p = .93, η^2_G < .01) or absolute bonus (F(1, 69) = 0.26, p = .61, η^2_G < .01), and no significant interaction (F(1, 69) = 0.02, p = .89, η^2_G < .01).

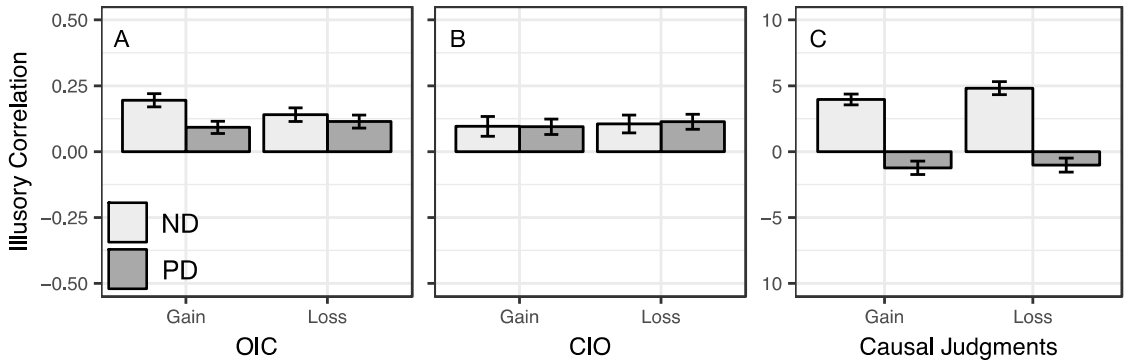

Figure 7. Illusory correlations for each level of Framing and Valence, for each DV. Note that the y axes differ and are truncated in panels A and B to more clearly illustrate the effect. Error bars indicate standard error of the mean.

5.2.2 Valence effects in predictions

As in Experiment 1, participants' trial-by-trial predictions of the drugs each patient received were used as another way to measure valence effects (Figure 8). The trial-by-trial illusory correlations were calculated in the same way as in Experiment 1, and were analyzed using the same 2x2 mixed ANOVA discussed above, with the trial-by-trial illusory correlations as the dependent variable. There was a significant valence effect (F(1, 141) = 5.34, p = .02, η^2_G = .02), but the main effect of absolute bonus (F(1,141) = 1.20, p = .29, η^2_G < .01) and the interaction (F(1,141) = 0.18, p = .67, η^2_G < .01) were both nonsignificant.

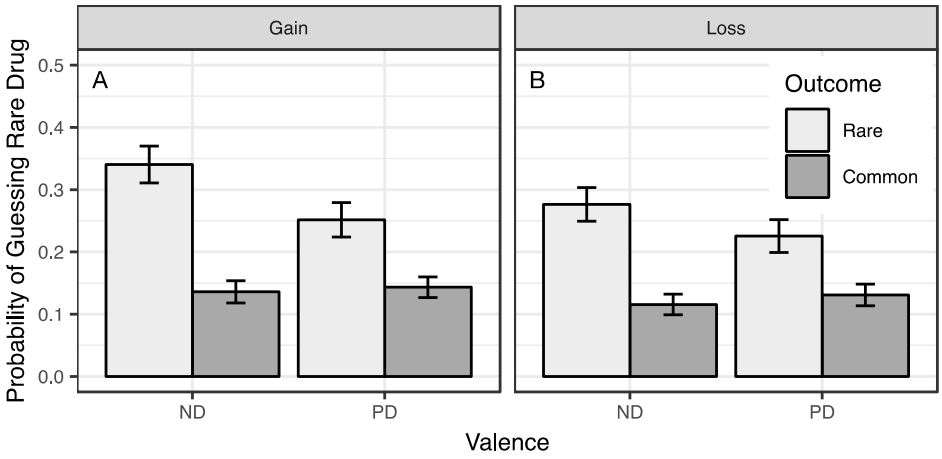

Figure 8. Probability of guessing rare drug when the patient experienced the common and rare outcomes in Experiment 2. Error bars indicate standard error of the mean.

5.2.3 Cell weighting

As in Experiment 1, participants' reported frequencies for the combination of the most common and rarest cue/outcome combinations were analyzed to determine whether valence effects were driven by participants overestimating the frequency of the rarest combination (i.e., distinctiveness).

Participant's estimates of the rarest combination were compared to the actual values using single-sample t-tests. Participants significantly overestimated the rarest combination in their O|C judgments ($t(309) = 14.83$, $p < .001$, $d = 0.84$) as well as their C|O judgments, $t(141) = 11.88$, $p < .001$, $d = 1.00$ (Figure 9).

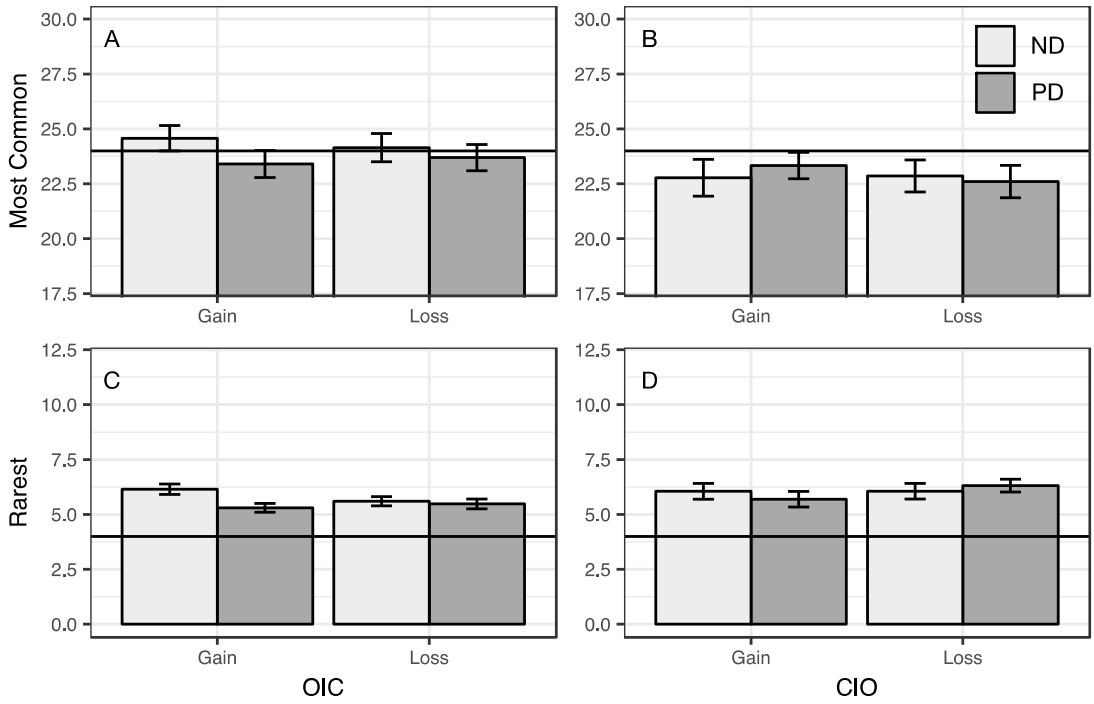

Figure 9. Average estimated frequencies of the most common (A,B) and rarest (C,D) cue/outcome combination, for O|C (A,C) and C|O (B,D) memory estimates in Experiment 2. Error bars indicate standard error of the mean. Horizontal lines indicate correct frequency.

The influence of valence and absolute bonus on these overestimations was then tested using the same mixed ANOVA from the previous Experiment 2 analyses, with rare combination O|C and C|O estimates as dependent variables. For the O|C judgments, there was a significant main effect of valence ($F(1, 153) = 4.62$, $p = .03$, $\eta^2_G = .02$), but the main effect of absolute bonus was nonsignificant ($F(1,153) = 0.76$, $p = .38$, $\eta^2_G < .01$), as was the interaction ($F(1, 153) = 2.62$, $p = .11$, $\eta^2_G = .01$). The analysis of C|O judgments yielded no significant main effect of valence ($F(1, 69) = 0.24$, $p = .88$, $\eta^2_G < .01$) or absolute bonus ($F(1, 69) = 0.79$, $p = .38$, $\eta^2_G = .01$), and the interaction was also nonsignificant ($F(1, 69) = 0.85$, $p = .36$, $\eta^2_G = .01$).

Next participants' estimates of the most common combination were analyzed in the same way. Participants' O|C estimates of the most common combination were not significantly different from the actual frequency ($t(309) = -0.15$, $p = .88$, $d = 0.01$), and their C|O estimates were significantly below the actual frequency ($t(141) = -3.05$, $p < .01$, $d = 0.26$).

Unlike participants' estimates of the rarest combination, participants' O|C estimates of the most common combination yielded no significant main effects of valence ($F(1, 153) = 2.11$, $p = .15$, $\eta^2_G < .01$) or absolute bonus ($F(1,153) = 0.01$, $p = .92$, $\eta^2_G < .01$), and no significant interaction ($F(1, 153) = 0.41$, $p = .52$, $\eta^2_G < .01$). Similarly, participants' C|O estimates yielded no significant main effects of valence ($F(1, 69) = 0.05$, $p = .82$, $\eta^2_G < .01$) or absolute bonus ($F(1,69) = 0.17$, $p = .68$, $\eta^2_G < .01$). The interaction was nonsignificant here as well ($F(1, 69) = 0.37$, $p = .54$, $\eta^2_G < .01$).

5.3 Discussion

Experiment 2 tested whether valence effects were sensitive to absolute or relative outcomes. Although the results from Experiment 1 were replicated, and they were once again fairly robust across all three dependent measures, there was no interaction between absolute Gain/Loss outcomes and Valence.

This pattern of results indicates that participants were more sensitive to the relative gains and losses than to the absolute amounts associated with those gains and losses; for participants in the Gain condition, failing to gain 6 cents was as distinctive as a loss. This is further evidence that some of the strongest illusory correlations occur when the common outcome is better and the rare outcome is worse. Unfortunately, the corollary is that most of us live in precisely the kind of world

in which people are most likely to exhibit illusory correlations (generally, good outcomes outnumber bad outcomes). Further, Experiment 2 shows that the valence effect appears to be pervasive; it can happen both for good and bad outcomes so long as the better outcomes are more common than the worse outcomes.

In Experiment 2 the congruency between absolute and relative gains/losses was manipulated, but the monetary difference between good and bad outcomes was constant (i.e., a good outcome was always six cents better than a bad outcome). In Experiment 3 the magnitude of this difference is manipulated in pursuit of a related question: are valence effects larger when bad outcomes are (more) bad and good outcomes are (more) good?

6.0 Experiment 3: Examining the Magnitude of the Outcomes in the Valence Effect

Previous research suggests that negative experiences can lead to illusory correlations, and that more intense negative stimuli generate stronger illusory correlations. For example, Wiemer et al. (2014) induced illusory correlations between neutral stimuli and startle sounds in a fear learning study. Their participants inferred stronger illusory correlations when the startle sounds were louder. Shook, Fazio, and Eiser (2006) found similar results in a category learning task in which participants classified fictional beneficial/harmful beans. Participants more readily generalized the features of harmful beans than beneficial ones, and this tendency was exaggerated for more extreme negative outcomes.

The primary research question in Experiment 3 was whether valence effects could be moderated by stronger positive/negative outcomes. This question has interesting implications for distinctiveness and salience perspectives. From a distinctiveness perspective, one might expect the trials with stronger outcomes to be misremembered as more frequent than they were (e.g., Jacoby & Craik, 1979). This would manifest as a main effect of outcome magnitude, and may (Figure 10A) or may not (Figure 10B) include an interaction with valence. By contrast, a gradual error-reduction theory (e.g., Rescorla & Wagner, 1972) might predict better learning for more salient outcomes, producing smaller illusory correlations, and perhaps even attenuating valence effects (Figure 10C).

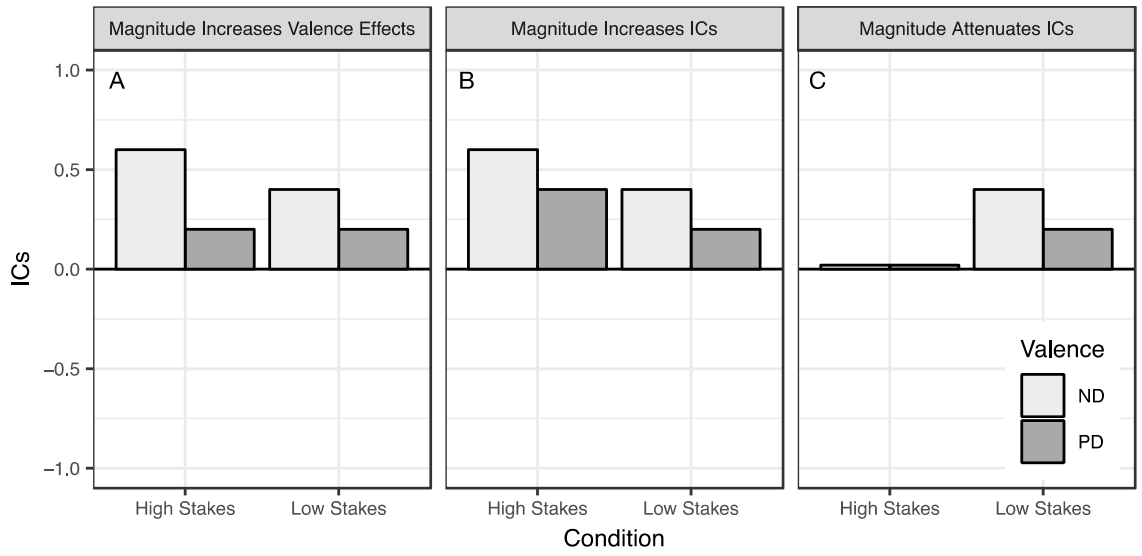

Figure 10. Possible patterns of results if higher stakes lead to larger valence effects (A), larger illusory correlations (B), or smaller illusory correlations (C).

6.1 Method

6.1.1 Participants

Participants (n = 159; 71 females, 88 males) were recruited through MTurk, with an average age of 36.89 years (SD = 11.10). Participants were paid a base rate of $3.50 for their participation, as well as accuracy bonuses (M = $2.15, SD = $0.20). As in Experiment 1, participants were given bonuses according to patient outcomes, but because ND vs PD valence was manipulated within subjects, these balanced out at 0 (Table 6).

Table 6. Datasets for Experiment 3.

	Negative-Distinctive			Positive-Distinctive		
		+18¢	-18¢		-18¢	+18¢
Higher Stakes	*Drug 1*	24	12	*Drug 3*	24	12
	Drug 2	8	4	*Drug 4*	8	4
		+6¢	-6¢		-6¢	+6¢
Lower Stakes	*Drug 1*	24	12	*Drug 3*	24	12
	Drug 2	8	4	*Drug 4*	8	4

6.1.2 Design and procedure

The design of Experiment 3 is very similar to Experiment 2: a 2x2 mixed factorial, with ND/PD Valence manipulated within subjects (Table 6). The between subjects factor was the magnitude of the outcome. Participants in the Higher Stakes condition were awarded 18 cents for good patient outcomes and penalized 18 cents for bad patient outcomes. Participants' bonus amounts were initialized at 30 cents, as in Experiment 1. Aside from these design differences, the procedure was identical to Experiment 2. The entire procedure took approximately 25 minutes to complete.

6.2 Results

6.2.1 Valence effects in memory and causal judgments

As in Experiments 1 and 2, illusory correlations for all three dependent measures were evaluated with a mixed factorial ANOVA. Results from Experiment 3 were analyzed with a

2(Valence: ND vs PD, within-subjects) x 2(Magnitude: Higher Stakes vs Lower Stakes) mixed ANOVA with Type III sums of squares. As in Experiments 1 and 2 there was a main effect of valence for the causal judgments, $F(1, 155) = 171.89$, $p < .001$, $\eta^2_G = .39$. However, there was not a significant main effect of magnitude ($F(1, 155) = 1.08$, $p = .30$, $\eta^2_G < .01$), nor was there an interaction, $F(1, 155) = 3.15$, $p = .08$, $\eta^2_G = .01$ (Figure 11). The same was true of participants' O|C judgments; the valence effect was significant ($F(1, 151) = 28.46$, $p < .001$, $\eta^2_G = .09$), but the main effect of magnitude ($F(1, 151) = 0.13$, $p = .72$, $\eta^2_G < .01$) and the interaction ($F(1, 151) = 1.60$, $p = .21$, $\eta^2_G = .01$) were both nonsignificant. Finally, there was also a significant valence effect in participants' C|O judgments[4] ($F(1, 151) = 22.23$, $p < .001$, $\eta^2_G = .08$). There was a marginal main effect of magnitude ($F(1, 151) = 2.90$, $p = .06$, $\eta^2_G < .01$), and the interaction was nonsignificant ($F(1, 155) = 0.13$, $p = .71$, $\eta^2_G < .01$).

[4] There was an order effect in the O|C judgments, as in Experiments 1 and 2. However, the results of the ANOVA were the same regardless of the order in which participants answered O|C and C|O judgments. Therefore, the tests reported here are collapsed across order.

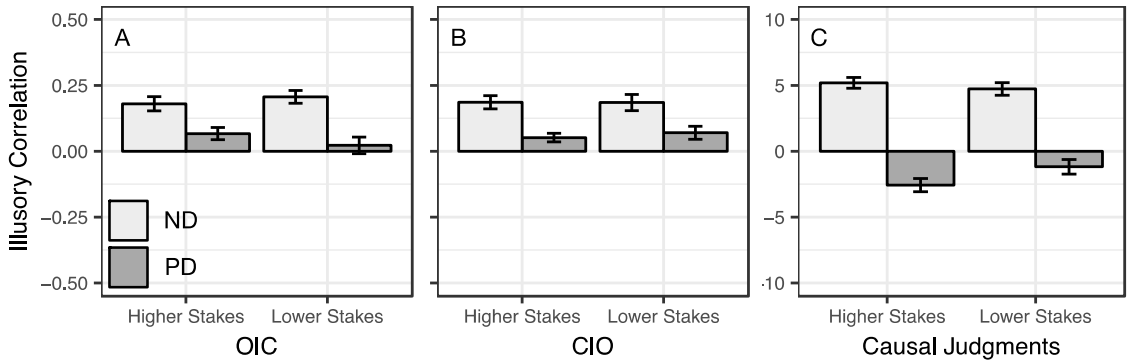

Figure 11. Illusory correlation judgments by valence and outcome magnitude (Higher Stakes; HS vs Lower Stakes; LS). Note that the y axes differ and are truncated in panels A and B to more clearly illustrate the effect. Error bars indicate standard error of the me

6.2.2 Valence effects in predictions

As in Experiments 1 and 2, valence effect were also calculated using participants' trial-by-trial predictions of the drugs each patients received (Figure 12). This was also analyzed using a 2(Valence: ND vs PD, within-subjects) x 2(Magnitude: Higher Stakes vs Lower Stakes) mixed ANOVA with Type III sums of squares. Again there was a main effect of valence ($F(1, 156) = 40.99$, $p < .001$, $\eta^2_G = .14$), but not magnitude ($F(1, 156) = 0.41$, $p = .52$, $\eta^2_G < .01$). There was no significant interaction ($F(1, 156) = 0.30$, $p = .59$, $\eta^2_G < .01$).

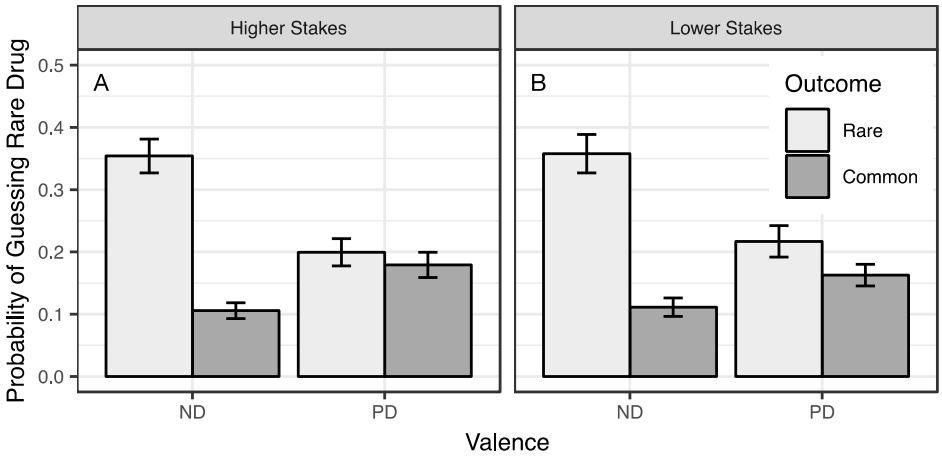

Figure 12. Probability of guessing rare drug when the patient experienced the common and rare outcomes in the Higher Stakes (A) and Lower Stakes (B) conditions. Error bars indicate standard error of the mean.

6.2.3 Cell weighting

As in Experiments 1 and 2, participants' estimates of the rarest and most common cue/outcome combinations were compared to confirm that illusory correlations in their memory estimates were driven by overweighting of rarest cue/outcome combination rather than the most common (Figure 13). Single sample t-tests showed that Participants' estimates of the rarest combination were significantly higher than the actual value in both their O|C ($t(305) = 12.31$, $p < .001$, $d = 0.70$) and C|O estimates ($t(305) = 15.34$, $p < .001$, $d = 0.88$).

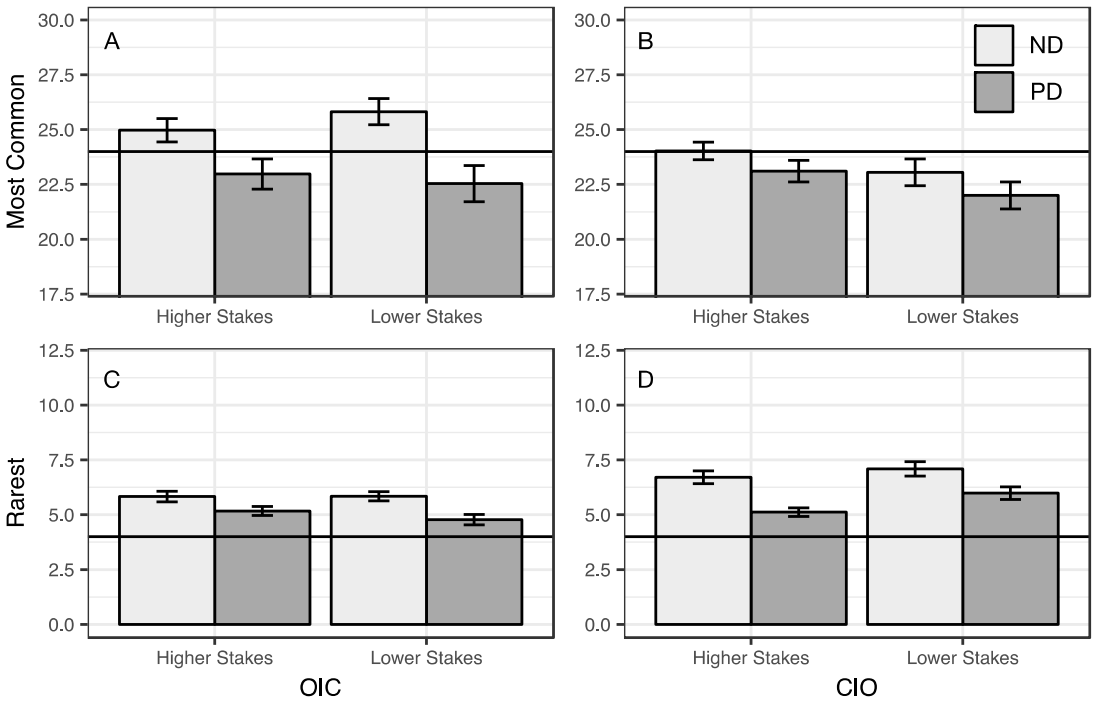

Figure 13. Average estimated frequencies of the most common (A,B) and rarest (C,D) cue/outcome combination, for O|C (A,C) and C|O (B,D) memory estimates in Experiment 3. Error bars indicate standard error of the mean. Horizontal lines indicate correct frequency.

Mixed ANOVAs with Type III sums of squares were conducted to examine valence effects in cell weighting. For O|C judgments, there was a significant main effect of valence ($F(1, 151) = 14.09$, $p < .001$, $\eta^2_G = .05$), but neither the effect of magnitude ($F(1, 151) = 0.78$, $p = .38$, $\eta^2_G < .01$) nor the interaction was significant ($F(1, 151) = 0.77$, $p = .38$, $\eta^2_G < .01$). There was also a significant main effect of valence in participants' C|O judgments ($F(1, 151) = 25.82$, $p < .001$, $\eta^2_G = .07$), as well as a main effect of magnitude ($F(1, 151) = 4.58$, $p = .03$, $\eta^2_G = .02$). The interaction was nonsignificant, $F(1, 151) = 0.85$, $p = .36$, $\eta^2_G < .01$.Thus, consistent with the idea that negative valence increases distinctiveness processing for rare events, participants consistently

overestimated the occurrence of the rare outcome, and this tendency was stronger in the ND condition.

Turning to participants' estimates of the common combination, single sample t-tests revealed that participants' O|C estimates were not significantly different from the actual value (t(305) = 0.22, p = .83, d = 0.01), and their C|O estimates significantly underestimated the frequency of the common combination (t(305) = -3.53, p < .001, d = 0.20). A mixed ANOVA with Type III sums of squares revealed a main effect of valence in the O|C judgments (F(1, 151) = 14.37, p < .001, η^2_G = .05). The main effect of magnitude (F(1, 151) = 0.10, p = .75, η^2_G < .01) and the interaction were nonsignificant (F(1, 151) = 0.84, p = .36, η^2_G < .01). A similar pattern emerged in the C|O judgments, which yielded a significant valence effect (F(1, 151) = 4.63, p = .03, η^2_G = .01), but not a significant main effect of magnitude (F(1, 151) = 2.93, p = .09, η^2_G = .01) or a significant interaction (F(1, 151) = 0.02, p = .89, η^2_G < .01).

Overall this suggests that people tend to overweight the frequency of the rare outcome (in line with the distinctiveness hypothesis), and that this tendency is stronger when the rare outcome is negative. Additionally, participants did not consistently overweight the occurrence of the common combination, which is consistent with the idea that the distinctiveness of the (negative) rare combination drives increased frequency estimates.

6.3 Discussion

In Experiment 3 the relationship between valence effects and outcome magnitude was examined. Valence effects from the previous experiments, were replicated but there was no main

effect of outcome magnitude. Finally, the interaction between valence and outcome magnitude was only significant in one of the three outcome variables, and the effect was very weak.

The absence of an interaction is difficult to interpret. It is possible that participants simply did not differentially encode gains/losses between the Higher and Lower Stakes conditions. Because outcome magnitude was manipulated between subjects, participants in the Higher Stakes condition were unaware that there was a Lower Stakes condition; perhaps 18 cents would only be encoded as "Higher Stakes" if participants knew that others were gaining/losing less money for the outcomes. Further, there is evidence from previous work that participants tend to discretize continuous variables (e.g., Marsh & Ahn, 2009). For example, if some continuous variable such as dosage of a drug were to increase by 18 mg from one trial to the next, participants may simply view that as "an increase" rather than encoding a representation of the magnitude. Similarly, it is possible that participants simply categorize losses of 6 cents and 18 cents as "losses," which are equally salient in memory.

It is also possible that the difference between 6-cent outcomes and 18-cent outcomes was simply too small to register as a meaningful difference for subjects; however, there are several reasons to be skeptical of this perspective. First, the difference between losing 6 cents and losing 18 cents is 12 cents. Valence effects were obtained in the previous experiments with a 12-cent difference between gain and loss (Experiment 1), and even with a 6-cent difference (Experiment 2). Additionally, the value weighting function in prospect theory is steepest for small gains and losses (Kahneman & Tversky, 1979). A difference of 12 cents will be felt the most when the prospects are close to zero (e.g., the difference between 6 cents and 18 cents should be viewed as much bigger than the difference between $1.06 and $1.18).

Experiments 1-3 extended the distinctiveness-based illusory correlation effect (e.g., Hamilton & Gifford, 1976) to causal scenarios and established a strong and reliable trend whereby negative-distinctive (ND) scenarios yield stronger illusory correlation inferences than positive-distinctive (PD) scenarios. Experiment 4 will examine a different question regarding valence: whether contingencies involving negative outcomes are more resistant to extinction.

7.0 Experiment 4: Valence Effects and Extinction

Experiment 4 investigates a different type of valence effect. Rather than examine the differential strength of illusory correlations according to valence, as in Experiments 1-3, Experiment 4, was related to differential extinction; are contingencies involving negative outcomes more difficult to extinguish compared to contingencies involving positive outcomes?

The approach in Experiment 4 was modeled after placebo/nocebo experiments (e.g., Au Yeung et al., 2014; Colagiuiri et al., 2015). While placebo/nocebo studies and studies of conditioning (e.g., Andreatta & Pauli, 2015) hint at a pattern of valence effects in extinction, I have not found an example in which positive (placebo) and negative outcomes (nocebo) are compared in the same study.

Experiment 4 was designed make a direct comparison between the rate of extinction with a positive vs. a negative outcome. Participants underwent a learning phase with 48 trials in which they learned a generative relationship; a medicine was either correlated with a good or a bad patient outcome. In in the next 48 trials there was zero correlation between the cue an outcome (an extinction phase). After the extinction phase, participants indicated their beliefs about the contingency. If contingencies involving negative outcomes are more difficult to extinguish, then participants in the Negative valence condition would have stronger causal strength beliefs than those in the Positive valence condition.

This approach could prove useful for two reasons. First, while it is important to understand how people acquire causal illusions, from a practical perspective it also important to understand how they are extinguished, and whether negative valence interferes with this process. There is some evidence from the fear learning literature that covariation biases (i.e., illusory correlations)

are stronger and more resistant to extinction when they involve aversive, fear-relevant stimuli (e.g., de Jong & Merckelbach, 2000; Pauli, Diedrich, & Müller, 2002). Additionally, people are sometimes required to detect changing contingencies over time in their everyday lives due to interactions with unobserved variables (e.g., Rottman & Ahn, 2011). Further, contingencies sometimes attenuate over time naturally. For example, the relationship between coffee intake and alertness diminishes as one habituates to caffeine (e.g., Rottman & Ahn, 2009).

Second, if correlations involving negative outcomes are more difficult to extinguish, it raises the possibility that the same underlying phenomenon could explain placebo/nocebo effects and the illusory correlations in the current work (i.e., a negativity bias).

7.1 Method

7.1.1 Participants

Participants (n = 160, 84 female, 75 male, 1 unreported) were recruited through MTurk. The average age was 34.84 (SD = 10.11). Participants were paid $3.50 plus accuracy bonuses (M = $1.53, SD = $0.13) to complete the task.

7.1.2 Design

Experiment 4 involved a simple between subjects manipulation. Subjects in the Positive condition learned about the relationship between a drug (vs. no drug) and a good (vs. neutral) outcome. Those in the Negative condition learned about the relationship between a drug (vs. no

drug) and a bad (vs. neutral) outcome (Table 7). Participants first learned an actual contingency between a the drug and the outcome in an acquisition phase. After 48 trials in the acquisition phase, participants transitioned to an extinction phase. The extinction phase consisted of the same kinds of trials, but with no contingency between the drug and the outcome, so subjects' prior beliefs about the contingency between the drug and the outcome should extinguish to some degree. Similarly to Colagiuiri et al.'s (2015) design, the base rate of the drug in the extinction phase was 50%.

Table 7. Acquisition and extinction data in Experiment 4.

		Negative Valence			Positive Valence	
		Bad Outcome	Normal Outcome		Good Outcome	Normal Outcome
Acquisition Phase	*Drug*	16	8	*Drug*	16	8
	No Drug	8	16	*No Drug*	8	16
Extinction Phase	*Drug*	12	12	*Drug*	12	12
	No Drug	12	12	*No Drug*	12	12

In Experiment 4, subjects were presented with only cover story valence in the outcomes; a positive/negative outcome did not involve monetary reward/punishment. Participants were still given bonuses when they accurately guessed whether or not each patient received the drug in the learning and extinction phases.

7.1.3 Procedure

Participants were told that they would be learning about the relationship between a medication and a patient outcome, and that their task was to determine whether the medication

caused the positive/negative outcome. After the initial training, participants completed an acquisition phase which they learned a generative contingency between a drug and the outcome (Table 7). After completing the acquisition trials, the procedure transitioned seamlessly to the extinction phase, in which the contingency between the drug and the outcome was zero. Finally, participants completed a test phase. In this final phase they filled in drug/outcome combinations from memory, just like in Experiments 1-3, and completed a causal strength slider[5] on a 21 point scale (participants could not see the numerical values). For example, if participants had learned about the drug XF7 in the Negative[Positive] condition, they were asked "How does XF7 affect patient outcomes?" The slider had text anchors on the ends; the left anchor would say, "XF7 strongly causes bad[good] outcomes," and the right anchor would say "XF7 strongly prevents bad[good] outcomes." The entire procedure took approximately 25 minutes to complete.

7.2 Results

If contingencies involving negative outcomes are more difficult to extinguish, then participants in the Negative valence condition would have stronger causal strength beliefs after the

[5] The causal strength judgment in Experiment 4 was different from the causal strength judgment in Experiments 1-3, for several reasons. If participants were given a forced choice about whether or not to prescribe the drug, they may have a bias toward prescribing in the positive condition and against prescribing in the negative condition, even if they believe the drug no longer has an effect. Participants in the Positive condition have never known the drug to harm a patient, and those in the Negative condition have never known it to benefit a patient.

extinction phase. This question was addressed using independent samples t-tests for each of the three outcomes.

Participants' causal judgments were significantly above zero in both the Negative ($t(78) = 6.95$, $p < .001$, $d = 0.78$) and Positive conditions ($t(80) = 3.75$, $p < .001$, $d = 0.41$) (Figure 14). Note however, that unlike in prior studies, this does not represent an illusory correlation because in the learning data there was a positive correlation in the first half. Participants in the Negative condition gave significantly stronger causal judgments compared to those in the Positive condition, $t(158) = 2.00$, $p = .047$, $d = 0.32$, which represents the valence effect. Though this effect was technically significant, it is very close to the threshold and the effect size is in the small range.

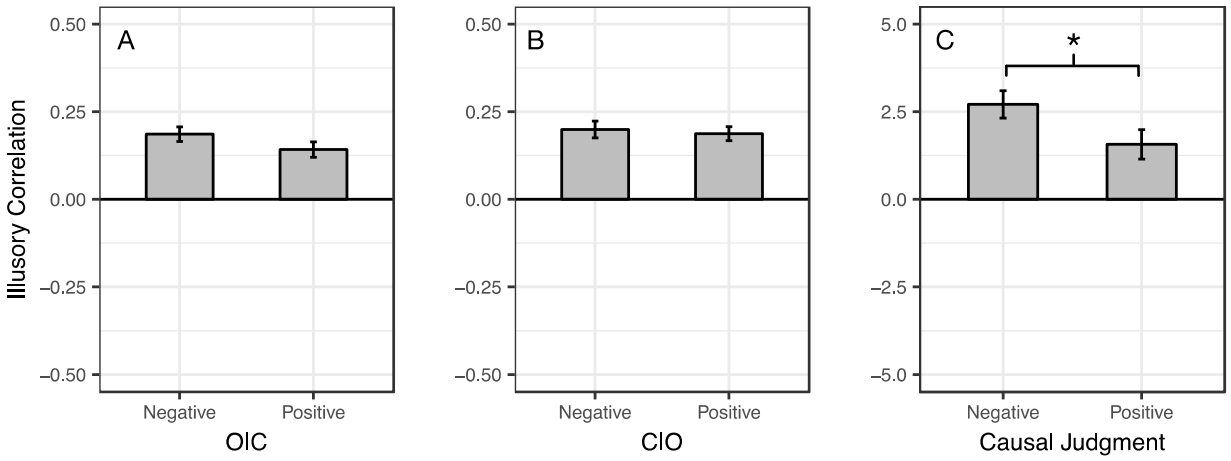

Figure 14. Illusory correlations from memory estimates and causal judgments. Note that the y axes differ and are truncated to more clearly illustrate the effect. Error bars indicate standard error of the mean. *p < .05

Participants' O|C judgments were also significantly above zero in both the Negative ($t(77) = 8.91$, $p < .001$, $d = 1.01$) and Positive conditions ($t(80) = 6.48$, $p < .001$, $d = 0.72$). However, the difference between them was nonsignificant, $t(157) = 1.45$, $p = .15$, $d = 0.23$.

The same pattern was present in the C|O judgments, which were significantly above zero in both Negative ($t(77) = 8.33$, $p < .001$, $d = 0.94$) and Positive conditions ($t(79) = 9.40$, $p < .001$, $d = 1.05$), but were not significantly different from each other, $t(156) = 0.39$, $p = .69$, $d = 0.06$.

7.3 Discussion

Experiment 4 provided some support for the idea that contingencies involving negative stimuli are more difficult to extinguish than those involving positive stimuli. Although the O|C and C|O memory estimates were nonsignificant, participants' causal judgments after the extinction phase were slightly stronger when the outcome was negative than when it was positive.

The current experiments have demonstrated that valence effects are robust across multiple valence modalities (Experiment 1), are driven by relative rather than absolute gains/losses (Experiment 2), and are insensitive to the magnitude of the gains/losses (Experiment 3). Further, correlations about negative outcomes are somewhat more resistant to extinction (Experiment 4). The remainder of the manuscript will focus on the implications of these findings, as well as whether the models introduced above can be modified to explain them.

8.0 Theoretical Accounts of Valence Effects

Current theories[6] that apply to causal learning and illusory correlation do not explicitly account for positive and negative valence, but it is possible that they could be modified to account for valence effects, whether through parameters relating to the salience of the outcome (e.g., Rescorla & Wagner, 1972) or the differential importance subjects place on different kinds of event combinations (e.g., Hamilton & Gifford, 1976; Kao & Wasserman, 1993). In this section I will review three theories and discuss whether they can be modified to explain valence effects.

8.1 Rescorla-Wagner

The Rescorla-Wagner model is a trial-by-trial error-correction algorithm for learning about the relations between cues and an outcome. On each trial the model uses its current weights of each cue to make a prediction about the state of an outcome. It then uses the difference between the prediction and the observed outcome to modify the weights in the correct direction. Equation 2 presents the form of the model when there is only one cue.

$$\Delta V = \alpha\beta(\lambda - \mathrm{V})\mathrm{C} \qquad \textbf{Equation 2}$$

[6] Another model was attempted. However, there were ambiguities in the explanations of multiple aspects of how the model works that made it impossible for me to reproduce the model (Kruschke, 1996).

On a given trial, the model makes a prediction by subtracting the cue's current strength (V, initialized at 0) from the observed outcome (λ, typically 1 or 0 to reflect the state of the effect) to form an error term. The error term is multiplied by two parameters (α and β, both bound at 0 and 1) which determine the model's rate of learning. The α term is the learning rate associated with the cue, whereas β indicates the learning rate for the outcome. Finally, C indicates the presence (C = 1) or absence (C = 0) of the cue; the model only learns about an individual cue when it is present.

The model makes one prediction that is particularly relevant to the current work. Even when a cue and outcome are uncorrelated, RW initially infers that they are correlated, and only after time learns that they are unrelated. This implies that, in the process of learning that two variables are independent, there will always be a temporary illusory contingency in the beginning, which will be extinguished given enough trials.

RW does not encode differences in valence explicitly; however, it is possible that the illusory correlations in RW could be altered by more or less salient outcomes (i.e., by changing the value of β. Generally, larger β values create a higher weight initially, with a faster asymptote to the correct weight[7].

Valence effects are simulated by assuming a higher learning rate for the rare than the common outcome. To model this, RW was calculated separately for the common and rare outcome, with different β values for the two. This generates four association weights between the common/rare cues and the common/rare outcomes. First the cues are subtracted within each

[7] All of the simulations reported in the current work use $\alpha = 1$. The main theoretical interest of the current work is the β parameter.

outcome to find the difference in the cue weights for each outcome. These two difference curves are themselves compared, and the difference of difference scores is analogous to an illusory correlation. To model the valence effect itself, this procedure is completed with β values that are higher for the rare outcome than the positive outcome (i.e., ND condition) and with β values that are lower for the rare outcome than the positive outcome (i.e., PD condition).

Simulations were conducted over varying levels of β for the Common and Rare outcome. The higher value of β was associated with the rare outcome (i.e., ND condition) and the common outcome (i.e., PD) in separate simulations. Two sets of β values were chosen, representing very small (e.g., .01 vs .05) or somewhat larger values (e.g., .1, vs .2). For the purposes of the present work, whether a simulated "valence effect" is robust to different scales of β is relevant; RW's ability to simulate a valence effect is less convincing to the extent that it depends on a very specific range of β values. The current simulations involved 1000 randomly ordered instances of the data in Table 1 with each β combination. Table 8 shows the final illusory correlation differences after 48 trials for both sets of β values.

Table 8. Simulated illusory correlations with higher β parameters for either the common or the rare outcome.

Analogous Condition	*Common Outcome* b	*Rare Outcome* b	*Simulated IC*	*Difference*
Negative-Distinctive	.1	.2	.08	.05
Positive-Distinctive	.2	.1	.03	
Negative-Distinctive	.01	.05	.02	-.02
Positive-Distinctive	.05	.01	.04	

Overall the simulation presents mixed evidence; although there is a small difference in the expected direction when the learning rates are larger, the difference goes away or even reverses when the learning rates are smaller. In short, RW can explain the valence effects found in

Experiments 1-3 under specific conditions, namely if negative outcomes are much more salient than positive ones, and the learning rates for both outcomes are not very low.

8.2 Rule-Based Models with Differential Cell Weighting (A-Cell Bias)

Another theoretical perspective that could explain valence effects is a modification of the Δp rule (e.g., Allan, 1980). The Δp rule was developed as a normative measure of causal strength for present/absent causes and effects. According to Δp, the normative contingency between two binary variables can be calculated from the frequencies in cells A-D of a contingency table (Figure 15). The Δp measure is calculated by subtracting the probability of the effect occurring when a cause is present from the probability of the effect occurring if the cause is absent (Equation 3). In the case of the data in Table 1, there is no contingency; Δp = 0.

		Outcome	
		Present	Absent
Cue	Present	*A*	*B*
	Absent	*C*	*D*

Figure 15. A contingency table between a binary cue and outcome.

$$\Delta p = \frac{A}{A+B} - \frac{C}{C+D} \qquad \textbf{Equation 3}$$

Human contingency judgments often deviate from this normative standard. One such consistent deviation is in the weights participants typically assign to the cells in the contingency tables. Wasserman, Dorner, and Kao (1990) found that participants' judgments were consistent

with higher weights for A Cell trials, then B, C, and D Cell trials respectively. Further, Wasserman et al. found that participants explicitly endorsed the view that A > B > C > D (see also Kao & Wasserman, 1993; Schustack & Sternberg, 1981). Kao and Wasserman proposed a version of the Δp rule that accounts for these differing weights (Equation 4).

$$KW = \frac{A(W_A)}{A(W_A) + B(W_B)} - \frac{C(W_C)}{C(W_C) + D(W_D)} \qquad \textbf{Equation 4}$$

Kao and Wasserman's (1993) formulation (hereafter KW) preserves the structure of the Δp rule while incorporating the weights (W terms in Equation 3) that subjects place on each of the four outcomes in the contingency table. Because KW places the greatest weight on the A Cell, in which both variables are present, it predicts that participants will infer illusory correlations. Although the formula was developed for present/absent variables, it can still be used to model the data in the current studies; it seems intuitive that varying the weights attributed to the cells could potentially explain valence effects if participants weight the negative outcomes more strongly. A modified version of KW with the frequencies from Table 1 is presented in Equation 5, where W_C is the weight attached to the common outcome and W_R is the weight attached to the rare outcome.

$$KW_M = \frac{24(W_C)}{24(W_C) + 12(W_R)} - \frac{8(W_C)}{8(W_C) + 4(W_R)} \qquad \textbf{Equation 5}$$

The problem with only manipulating the weights of the outcome is that for noncontingent data, KW_M will never yield a nonzero estimate. This can be seen more easily when the term on the right is multiplied by 3/3 (Eq. 6).

$$KW_M = \frac{24(W_C)}{24(W_C) + 12(W_R)} - \frac{24(W_C)}{24(W_C) + 12(W_R)} = 0 \qquad \textbf{Equation 6}$$

By multiplying the value on the right by 3/3, we can see that no matter what values are assigned to wC and wR, both terms on the right side of the equation will always be equal (i.e., the model will always return 0 for the data in Experiments 1-3). Weighting for outcome valence alone is not enough to recreate valence effects in a version of Δp. Of course, the model could be further modified to produce valence effects if the cell frequencies are also weighted by the probability of the outcome.

8.3 Pseudocontingencies

According to the Pseudocontingencies view, subjects use information other than joint observations of two variables to form contingency judgments. Specifically, subjects use base rate information either in addition to or in place of relevant state combinations when inferring the contingency between two variables (e.g., Fiedler, Freytag, & Meiser, 2009). Fiedler, Kutzner, and Vogel (2013) described illusory correlations as the result of this PC heuristic that people use in addition to contingency information, or when contingency information is unavailable. The principle prediction of the PC perspective is that subjects will infer correlations between variables

with base rates skewed in the same direction. For example, if two variables both have a very common state and a very rare state, subjects will infer that the common states are positively correlated. This maps directly onto the traditional illusory correlation paradigm: most people are in Group A, and most of the behaviors are good, so the two must be correlated. In terms of the current work, subjects might reason that because most patients receive Drug 1 and most patients have good outcomes, Drug 1 must be the better treatment.

Although the PC literature contains many descriptions of this phenomenon, recent papers do not offer a formal mathematical model[8]. For the purposes of the current work, I developed a simple model to explain how subjects infer correlation from base rates of the cue and the outcome. The model in Equation 7 approximates the qualitative PC predictions regarding basic illusory correlations without valence effects. The model makes the following prediction, in line with Fiedler and colleagues (e.g., Fiedler & Freytag, 2004; Fiedler et al., 2013; Kutzner & Fiedler, 2017). When the variables are more skewed the model outputs a stronger judgment.

$$PC = \frac{P(Common\ Cue) - .5}{.5} \times \frac{P(Common\ Outcome) - .5}{.5} \qquad \textbf{Equation 7}$$

The model's only inputs are the base rates of each variable. It generates estimates by specifying the skew for each variable (subtracting .5 from the base rate P), and dividing that difference by .5. Because both the left and right sides are bounded at 0 and 1, the resulting estimates

[8] An information loss model of illusory correlation was presented by Fiedler (1996). However, PC theorists in recent years do not typically discuss pseudocontingencies in terms of information loss.

are also bounded at 0 and 1. For example, the model predicts a positive estimate (.17) for the data in Experiments 1-3, in which P(Common Cue) = .75 and P(Common Outcome) = .67.

While the PC model can explain a standard illusory correlation as in Table 1, the PC theory does not predict the valence effect. One possibility for modifying the PC theory so that it could explain valence effects is to assume that people interpret valence of the outcome as more skewed than it really is when the outcome is negative (i.e., that people under-estimate the probability that negative outcomes will occur in the ND condition and overestimate the probability that they will occur in the PD condition), which could yield a more skewed impression of the base rate in the ND condition. For example, a positive outcome with a base rate of .8 might seem to have a base rate of .95. Equation 8 represents one possible model,

$$PC_M = \frac{P(Common\ Cue) - .5}{.5} \times Sfun(\frac{P(Common\ Outcome) - .5}{.5}, S) \qquad \textbf{Equation 8}$$

in which *S* is a salience parameter bounded at 0 and 1, and *Sfun* is defined in Equation 9.

$$Sfun(x,S) = \begin{cases} x + (1-x)S & \text{if x} > 0 \\ x + (-1-x)S & \text{if x} < 0 \end{cases} \qquad \textbf{Equation 9}$$

This modified PC model (PCM) yields stronger illusory correlation estimates when the S parameter is larger, so a larger vs. smaller S could be used to model ND and PD conditions respectively. When S=0, PCM estimates the correlation between the variables in Experiments 1-3 as .17, same as in the original PC model. When S = .5, the modified PC model produces an

estimate of .33, and when S = 1, it is .5. This model would therefore predict that more salient rare outcomes lead to stronger illusory correlations.

Although this modification may produce valence effects, there is some reason to doubt its psychological plausibility. The mechanism by which the model produces valence effects is essentially by underweighting the frequency of the rare outcome in the ND condition. In Experiments 1-3 in the current work, participants overestimated the occurrence of the rare combination. (This pattern does not directly contradict the idea that participants underestimate the base rates of negative outcomes, because in Experiments 1-3 participants were explicitly told the overall base rates before making their judgments.) Additionally, underestimating the occurrence of rare negative outcomes would be a peculiar departure from a well-known phenomenon by which subjects tend to overweight the probability of rare events and underweight the probability of common events (e.g., Kahneman & Tversky, 1979).

8.4 Summary of Models

So far, straightforward adaptations of several models revealed the following. First, increasing the learning rate parameter for negative outcomes for RW may explain the valence effect, but only in some ranges of . Second, a weighted version of Δp with higher weights for negative outcomes has no impact on the predicted strength of illusory correlation, and therefore cannot explain valence effects. Third, there is a fairly straightforward way to modify the pseudocontingencies model such that more salient (negative) outcomes produce larger illusory correlations. However, this modification requires some unlikely assumptions about the way

participants reason about the base rates of positive and negative outcomes. In short, none of these modifications can robustly account for valence effects.

9.0 General Discussion

The present work yielded several empirical findings which together suggest that people process negative outcomes as particularly distinctive events, leading to valence effects in illusory correlation. First and most importantly, all of the experiments in the current work found reliable valence effects. In Experiments 1-3, which were modeled after the original illusory correlation paradigm (e.g., Hamilton & Gifford, 1976; Mullen & Johnson, 1990), participants showed an overall illusory correlation between the rare cue and the rare outcome. Further, the current work has demonstrated for the first time that illusory correlations are larger when the rare combination involves a negative outcome (negative-distinctive; ND) compared to when it involves a positive outcome (positive-distinctive; PD).

Second, valence effects are present with and without monetary outcomes. Experiment 1 involved three valence presentation modalities: a Story condition in which the good/bad outcomes did not affect participants' bonuses, a Monetary condition in which the good/bad outcomes were not reflected in the cover stories, and a Combined condition in which both cover story and monetary valence were present. Valence effects were present across all three conditions.

Third, valence effects are driven by relative rather that absolute gains and losses. Participants in Experiment 2 were randomized to either an absolute gain condition (better outcomes involved 6-cent gains, and worse outcomes involved no gain) or absolute loss condition (better outcomes involved no loss, and worse outcomes involved 6-cent losses). Valence effects were significant in both conditions, not just the loss condition, implying that they are driven by relative better vs. worse outcomes rather than absolute loss.

Fourth, in Experiment 3, participants in both a lower stakes (good and bad outcomes of +/- 6 cents) and higher stakes condition (good and bad outcomes of +/- 18 cents) formed stronger illusory correlations in the ND condition, and there was no difference between conditions. This raises the possibility that, at least within the range tested, valence effects are driven by relative losses but are not obviously influenced by the magnitude of the loss. Experiments 1-3 showed that valence effects are robust to a variety of manipulations.

Finally, valence effects also manifest in an extinction paradigm similar to previous research on placebo/nocebo effects (e.g., Au Yeung et al., 2014; Colagiuiri et al., 2015). In Experiment 4, participants experienced 48 trials with a generative contingency between a drug and a positive/negative outcome, followed by 48 noncontingent trials. Participants had stronger causal beliefs after the extinction phase if the initial contingency they learned involved a negative outcome. Although this finding is interesting because of the parallels to the placebo/nocebo domain of research, it was the weakest effect of the four experiments, and was only present for some of the dependent variables, unlike the very robust valence effects in Experiments 1-3.

Simulations showed that none of the models discussed in the current work (KW, RW, PCs) could reasonably account for the valence effect. KW would require extensive modification to account for valence effects, and the PC modification proposed in the current work relies on assumptions that are contradicted by the data in Experiments 1-3. However, a simple modification to RW could recreate the effect under some specific learning rate parameters. The modification is meant to augment the salience of trials in which the outcome is negative, reflecting a negativity bias by which people disproportionately attend to negative stimuli (e.g., Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001; Rozin & Royzman, 2001; Vaish, Grossman, & Woodward, 2008).

9.1 Valence Effects, Distinctiveness, and Negativity Bias

The current research also provides some new evidence about the reason for illusory correlation, specifically, that the illusory correlation effect itself (regardless of valence) is driven by distinctiveness (e.g., Hamilton & Gifford, 1976; Hamilton, Dugan, & Trolier, 1985) rather than purely by information loss (e.g., Fiedler, 1996; 2000) or the more recent pseudocontingencies perspective (e.g., Fiedler et al., 2009; Fiedler et al., 2013). In Experiments 1-3, participants were prompted to estimate the frequency of each combination of the cue and outcome. While the pseudocontingencies model discussed previously depends on participants underestimating the rare-negative outcome, the distinctiveness perspective predicts overestimation for the rare combination. In all three of these experiments, participants disproportionately overestimated the rare cue/outcome combination, consistent with distinctiveness-based illusory correlation.

At first glance the distinctiveness pattern obtained in the current work may appear to contradict a well-known finding in causal learning, namely that people differentially weight cells in a contingency table in the form of A-Cell bias: A > B > C > D (Figure 15) (e.g., Kao & Wasserman, 1993). However, the A/B/C/D cell labeling in causal learning studies is not to be interpreted in the same way as the labeling in Experiments 1-3 of the present work. Causal learning studies often use present/absent cues and outcomes rather than cues and outcomes with multiple levels (e.g., Kao & Wasserman, 1993; Schustack & Sternberg, 1981; Spellman et al., 1996; Wasserman et al., 1990). Disproportionate cell weighting in studies with these kinds of stimuli comes from participants giving disproportionate weight to trials in which cues are present. The presence of a cue is more salient than its absence, so much so that some learning models only update on trials in which the cue is present (e.g., Rescorla & Wagner, 1972).

Rather than contradicting the distinctiveness account, it is possible that the distinctiveness of A-Cells is what makes them stand out in memory. Generally it is easier to remember examples of things happening than examples of nothing happening (e.g., Hearst, 1991), and models such as RW take this assumption into account. However, there are some situations in which the absence of a cue is distinctive (e.g., Van Hamme & Wasserman, 1994; Wasserman and Castro, 2005). For example, if the base rate of the cue occurring is very high, one might expect reduced or eliminated A-Cell bias.

In all three of the relevant experiments, participants' tendency to overweight the rare cue/outcome combination was stronger when the rare combination involved a negative outcome than when it involved a positive outcome. Thus, participants seem to process negative outcomes as distinctive events, increasing their estimate of the rare cue/outcome frequency.

The pattern of results in the present work is consistent with the idea of an overall negativity bias by which people attend to negative events more closely than positive events (e.g., Baumeister et al., 2001; Vaish et al., 2003). It is possible that negativity bias is an evolutionary adaptation for minimizing the potential harm that comes from experiencing negative events. Rozin and Royzman (2001) found that not only do people disproportionately attend to negative information, but they interpret outcomes that involve a mixture of positive and negative components in overly negative ways. Rozin and Royzman explained this phenomenon through the metaphor of contagion. A drop of poison in a gallon of water renders the whole gallon poisonous. At a less extreme level, the sighting of a single cockroach is often enough to ruin a delicious meal. Yet, as Rozin and Royzman point out, there is no "anti-cockroach." In other words, no food is so delicious that a small amount of it can overwhelm the presence of a large number of cockroaches. The point may seem obvious,

but in some ways the obviousness is the point. Cockroaches carry diseases, and so we should have a healthy fear of contamination if they are near our food.

There is also evidence that negativity bias informs the explanations we create in our daily lives. Hindsight bias is a tendency to apply post-hoc explanations as if they were predictive (i.e., "I knew it all along") (e.g., Hoffrage & Pohl, 2003). Previous work suggests that hindsight bias is stronger when explaining negative outcomes (e.g., Pezzo, 2003; Schkade & Kilbourne, 1991). In other words, participants are more likely to form post-hoc explanations to explain negative events than positive ones, as if the need to explain negative events is greater than the need to explain positive events. Participants may import this desire to explain negative outcomes into the current experiments, particularly because they involve a causal strength judgment at the end. A causal strength judgment represents an invitation to explain the distribution of outcomes in terms of the common or rare drug, not merely describe the frequencies of the four combinations. Thus the link between hindsight bias and valence effects may be particularly strong for the experiments in the current work.

9.2 Pattern of Illusory Correlations in the PD Condition

In Experiments 1-3, participants' causal judgments in the PD condition were usually in the unexpected direction. While the predicted pattern included stronger illusory correlations in the ND condition than the PD condition, previous research suggests that participants would still form illusory correlations between the rare cue and outcome in the PD condition. However, participants in the PD condition gave "negative" answers on the causal scale (indicating an IC between the rare outcome and the common cue). More puzzling, this pattern did not appear in participants' memory

estimates in the PD condition, raising the question of how participants were interpreting the causal strength question. This section will review the causal task and explain why this pattern of results, while surprising, does not detract from the overall pattern of valence effects.

Participants completed a causal strength task at the end of each scenario, after the 48 learning trials and the two memory estimates. The task was framed as a gamble, with the bolded heading "Which drug will YOU prescribe…" Participants were told that they would now choose a drug to prescribe to a new patient, and their bonus would be adjusted based on the patient's outcome (Figure 3).

One concern regarding the pattern of participants' judgments is that they were simply selecting the most common drug without consideration. Because the question was always framed around predicting a good outcome, in the ND condition (in which the good outcome was common) selecting the most common drug would amplify the illusory correlation in the expected direction. In the PD condition (in which the good outcome was rare), selecting the most common drug would push the illusory correlation away from the expected direction. It is possible that the causal strength measure is, to some extent, tainted by participants' preference for the common drug over the rare drug. However, it is clear from the task instructions and the screenshots (Figure 3) that participants must have given some level of consideration to choosing the most common outcome. They stood to lose most (in some cases all) of their bonus if they were incorrect, they were given the option to change their choice if they desired, and there was no time limit on the task itself.

If they were purposefully choosing the most common drug, this might reflect a preference beyond participants' pure causal strength beliefs. For instance, they may be more comfortable betting on the option that they know more about (i.e., a kind of "Devil you know" preference). Alternatively, they may be importing their beliefs about the real world into the task; in real life it

is reasonable to assume that the most commonly prescribed drug for a given disease is effective (although cover story explanations are less persuasive given that the effect was also found in the Monetary condition of Experiment 1, in which participants were not given information about medications). In any case, the reversals in the PD condition are to some extent concerning.

However, there are still several strong indicators of valence effects in the present work. First, analyses of participants' trial-by-trial predictions in Experiments 1-3 showed the same pattern; when given the rare outcome, participants were more likely to choose the rare outcome than the common one (an illusory correlation). This difference was consistently larger in the ND condition than the PD condition. Second, participants also exhibited valence effects in their memory estimates. Third, Experiment 4 used a more traditional causal strength measure, as well as cues and outcomes with 50% base rates (meaning that participants could not default to choosing the most common value of the cue). Even here there was a small but significant valence effect. Additionally, a pilot study was conducted that used a different wording for the causal strength question. Participants were asked which drug leads to worse patient outcomes. With the question framed this way, it would no longer make sense for participants to select the common drug because it is better, since the task was to select the drug that was worse. Participants exhibited valence effects in this study as well, and the average illusory correlation in the PD condition in the pilot study was slightly negative (but not significantly different from zero).

9.3 Conclusions

Our ability to detect related variables is more useful if we also possess the ability to detect when such relationships do not exist. The current work presented evidence from four experiments

suggesting that negative valence interferes with our ability to detect such non-contingent relationships.

Future empirical work could focus on several open questions raised by the current work. One of the current experiments manipulated the magnitude of the monetary gains and losses, but participants in each condition did not have a context in which to place the magnitude of their outcomes. Perhaps using larger outcomes such as 50 cents, manipulating the absolute gains and losses within subjects instead of between, or adding a condition in which the gains and losses are asymmetrical (e.g., gain 6 cents, lose 36) would moderate the valence effect. This would allow for a related question to be addressed: at what point are gains large/salient enough to cancel out or even reverse valence effects?

The present work has also put forth several possibilities for how existing models may be adjusted to account for increased salience that comes from negative valence. Although these possibilities are far from exhaustive, they represent a possible jumping-off point for future modeling work on the valence effect, and on illusory correlation more generally.

Regardless of the specific directions of future research in illusory correlation and illusory causal inference, one hopes that it will account for the most consistent finding from the present work; negative valence augments illusory correlations.

Illusory correlations are an important and widely studied paradigm in cognitive psychology; however, the role of valence in how people form illusory correlations has not previously been explored. The current work demonstrates that valence is an important factor in how people think about non-contingent variables, including how they might form conclusions in their everyday lives. If we live in a world in which people are generally good, and in which we encounter some kinds of people more frequently than others, even the most fair-minded person is

already set up to form unfair stereotypes about already-marginalized groups. Further study of valence effects in illusory correlation may help us overcome these structures that tilt our perceptions toward negative bias.

Bibliography

Acorn, D. A., Hamilton, D. L., & Sherman, S. J. (1988). Generalization of biased perceptions of groups based on illusory correlations. *Social Cognition, 6*(4), 345–372. doi: 10.1521/soco.1988.6.4.345

Aeschleman, S. R., Rosen, C. C., & Williams, M. R. (2003). The effect of non-contingent negative and positive reinforcement operations on the acquisition of superstitious behaviors. *Behavioural Processes, 61*(1-2), 37–45. doi: 10.1016/s0376-6357(02)00158-4

Allan, L. G. (1980). A note on measurement of contingency between two binary variables in judgment tasks. *Bulletin of the Psychonomic Society, 15*(3), 147–149. doi: 10.3758/bf03334492

Andreatta, M., & Pauli, P. (2015). Appetitive vs. aversive conditioning in humans. *Frontiers in Behavioral Neuroscience, 9*. doi: 10.3389/fnbeh.2015.00128

Arkes, H. R., & Harkness, A. R. (1983). Estimates of contingency between two dichotomous variables. *Journal of Experimental Psychology: General, 112*(1), 117–135. doi:10.1037/0096-3445.112.1.117

Au Yeung, S. T., Colagiuri, B., Lovibond, P. F., & Colloca, L. (2014). Partial reinforcement, extinction, and placebo analgesia. *Pain, 155*(6), 1110–1117. doi: 10.1016/j.pain.2014.02.022

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology, 5*(4), 323–370. doi: 10.1037/1089-2680.5.4.323

Benedetti, F., Amanzio, M., Casadio, C., Oliaro, A., & Maggi, G. (1997). Blockade of nocebo hyperalgesia by the cholecystokinin antagonist proglumide. *Pain. 71*. 135-40. doi: 10.1016/S0304-3959(97)03346-0.

Benedetti, F., Amanzio, M., Vighetti, S., & Asteggiano, G. (2006). The biochemical and neuroendocrine bases of the hyperalgesic nocebo effect. *Journal of Neuroscience, 26*(46), 12014–12022. doi: 10.1523/jneurosci.2947-06.2006

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review, 104*(2), 367–405. doi: 10.1037//0033-295X.104.2.367

Colagiuri, B., Quinn, V. F., & Colloca, L. (2015). Nocebo hyperalgesia, partial reinforcement, and extinction. *Pain, 16*(10), 995–1004. doi: 10.1016/j.jpain.2015.06.012

Colloca, L., & Benedetti, F. (2007). Nocebo hyperalgesia: How anxiety is turned into pain. *Current Opinion in Anaesthesiology, 20*(5), 435–439. doi: 10.1097/aco.0b013e3282b972fb

Colloca, L., Petrovic, P., Wager, T. D., Ingvar, M., & Benedetti, F. (2010). How the number of learning trials affects placebo and nocebo responses. *Pain, 151*(2), 430–439. doi: 10.1016/j.pain.2010.08.007

Colloca, L., Sigaudo, M., & Benedetti, F. (2008). The role of learning in nocebo and placebo effects. *Pain, 136*(1–2), 211–218. doi: 10.1016/j.pain.2008.02.006

Davenport, E. C., & El-Sanhurry, N. A. (1991). Phi/Phimax: Review and synthesis. *Educational and Psychological Measurement, 51*(4), 821–828. doi:10.1177/001316449105100403

De Jong, P. J., & Merckelbach, H. (2000). Phobia-relevant illusory correlations: The role of phobic responsivity. *Journal of Abnormal Psychology, 109*(4), 597–601. doi:10.1037/0021-843x.109.4.597

De Jong, P. J., Merckelbach, H., & Arntz, A. (1995). Covariation bias in phobic women: The relationship between a priori expectancy, on-line expectancy, autonomic responding, and a posteriori contingency judgment. *Journal of Abnormal Psychology, 104*(1), 55–62. doi: 10.1037/0021-843x.104.1.55

Eder, A. B., Fiedler, K., & Hamm-Eder, S. (2011). Illusory correlations revisited: The role of pseudocontingencies and working-memory capacity. *The Quarterly Journal of Experimental Psychology, 64*(3), 517-532. doi: 10.1080/17470218.2010.509917

Fazio, R. H., Eiser, J. R., & Shook, N. J. (2004). Attitude formation through exploration: Valence asymmetries. *Journal of Personality and Social Psychology, 87*(3), 293–311. doi: 10.1037/0022-3514.87.3.293

Fiedler, K. (1996). Explaining and simulating judgment biases as an aggregation phenomenon in probabilistic, multiple-cue environments. *Psychological Review, 103*(1), 193–214. doi:10.1037/0033-295x.103.1.193

Fiedler, K. (2000). Illusory correlations: A simple associative algorithm provides a convergent account of seemingly divergent paradigms. *Review of General Psychology, 4*(1), 25–58. doi:10.1037/1089-2680.4.1.25

Fiedler, K., & Freytag, P. (2004). Pseudocontingencies. *Journal of Personality and Social Psychology, 87*(4), 453–467. doi: 10.1037/0022-3514.87.4.453

Fiedler, K., Freytag, P., & Meiser, T. (2009). Pseudocontingencies: An integrative account of an intriguing cognitive illusion. *Psychological Review, 116*(1), 187–206. doi: 10.1037/a0014480

Fiedler, K., Kutzner, F., & Vogel, T. (2013). Pseudocontingencies: Logically unwarranted but smart inferences. *Current Directions in Psychological Science, 22*(4), 324-329. doi: 10.1177/0963721413480171

Genesove, D., & Mayer, C. (2001). Loss aversion and seller behavior: Evidence from the housing market. *The Quarterly Journal of Economics, 116*(4), 1233–1260. doi: 10.1162/003355301753265561

Hamilton, D. L., Dugan, P. M., & Trolier, T. K. (1985). The formation of stereotypic beliefs: Further evidence for distinctiveness-based illusory correlations. *Journal of Personality and Social Psychology, 48*(1), 5-17. doi: 10.1037/0022-3514.48.1.5

Hamilton, D. L., & Gifford, R. K. (1976). Illusory correlation in interpersonal perception: A cognitive basis of stereotypic judgments. *Journal of Experimental Social Psychology, 12*(4), 392-407. doi: 10.1016/s0022-1031(76)80006-6

Hardie, B. G. S., Johnson, E. J., & Fader, P. S. (1993). Modeling loss aversion and reference dependence effects on brand choice. *Marketing Science, 12*(4), 378–394. doi:10.1287/mksc.12.4.378

Jacoby, L. L., & Craik, E I. M. (1979). Effects of Elaboration of Processing at Encoding and Retrieval: Trace Distinctiveness and Recovery of Initial Context. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of Processing in Human Memory* (pp. 1-22). Hillsdale, NJ: Erlbaum.

Johnston, N. E., Atlas, L. Y., & Wager, T. D. (2012). Opposing effects of expectancy and somatic focus on pain. *PLoS ONE, 7*(6), e38854. doi: 10.1371/journal.pone.0038854

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1990). Experimental tests of the endowment effect and the coase theorem. *Journal of Political Economy, 98*(6), 1325–1348. doi:10.1086/261737

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*(2), 263. doi: 10.2307/1914185

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist, 39*(4), 341–350. doi: 10.1037/0003-066x.39.4.341

Kao, S.-F., & Wasserman, E. A. (1993). Assessment of an information integration account of contingency judgment with examination of subjective cell importance and method of information presentation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(6), 1363–1386. doi: 10.1037/0278-7393.19.6.1363

Köszegi, B., & Rabin, M. (2006). A model of reference-dependent preferences. *The Quarterly Journal of Economics, 121*(4), 1133–1165. doi: 10.1093/qje/121.4.1133

Kruschke, J. K. (1996). Base rates in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(1), 3–26. doi: 10.1037/0278-7393.22.1.3

Kruschke, J. K. (2003). Attention in learning. *Current Directions in Psychological Science, 12*(5), 171–175. doi: 10.1111/1467-8721.01254

Kutzner, F., & Fiedler, K. (2017) Stereotypes as pseudocontingencies, *European Review of Social Psychology, 28*(1): 1-49. doi: 10.1080/10463283.2016.1260238

Kutzner, F., Vogel, T., Freytag, P., & Fiedler, K. (2011). A robust classic. *Experimental Psychology, 58*(6), 443–453. doi: 10.1027/1618-3169/a000112

Marsh, J. K., & Ahn, W.-K. (2009). Spontaneous assimilation of continuous values and temporal information in causal induction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(2), 334– 352. doi: 10.1037/a0014929

Matute, H., & Blanco, F. (2014). Reducing the illusion of control when an action is followed by an undesired outcome. *Psychonomic Bulletin & Review, 21*(4), 1087–1093. doi: 10.3758/s13423-014-0584-7

Matute, H., Blanco, F., Yarritu, I., Díaz-Lago, M., Vadillo, M. A., & Barberia, I. (2015). Illusions of causality: How they bias our everyday thinking and how they could be reduced. *Frontiers in Psychology, 6*. doi: 10.3389/fpsyg.2015.00888

Matute, H., Yarritu, I., & Vadillo, M. A. (2011). Illusions of causality at the heart of pseudoscience. *British Journal of Psychology, 102*(3), 392-405. doi: 10.1348/000712610x532210

Morton, D. L., Watson, A., El-Deredy, W., & Jones, A. K. (2009). Reproducibility of placebo analgesia: effect of dispositional optimism. *Pain, 146*(1-2), 194-198. doi: 10.1016/j.pain.2009.07.026

Mullen, B. and Johnson, C. (1990), Distinctiveness-based illusory correlations and stereotyping: A meta-analytic integration. *British Journal of Social Psychology, 29*, 11–28. doi: 10.1111/j.2044-8309.1990.tb00883.x

Pauli, P., Diedrich, O., & Müller, A. (2002). Covariation bias in the affect-modulated startle paradigm. *Journal of Behavior Therapy and Experimental Psychiatry, 33*(3-4), 191–202. doi:10.1016/s0005-7916(02)00051-4

Pauli, P., Montoya, P., & Martz, G.-E. (1996). Covariation bias in panic-prone individuals. *Journal of Abnormal Psychology, 105*(4), 658–662. doi: 10.1037/0021-843x.105.4.658

Pauli, P., Montoya, P., & Martz, G. E. (2001). On-line and a posteriori covariation estimates in panic-prone individuals: Effects of a high contingency of shocks following fear-irrelevant stimuli. *Cognitive Therapy and Research*, *25*(1), 23-36.

Petersen, G. L., Finnerup, N. B., Colloca, L., Amanzio, M., Price, D. D., Jensen, T. S., & Vase, L. (2014). The magnitude of nocebo effects in pain: A meta-analysis. *Pain, 155*(8), 1426–1434. doi: 10.1016/j.pain.2014.04.016

Pope, D. G., & Schweitzer, M. E. (2011). Is Tiger Woods loss averse? Persistent bias in the face of experience, competition, and high stakes. *American Economic Review, 101*(1), 129–157. doi:10.1257/aer.101.1.129

Quartana, P. J., Campbell, C. M., & Edwards, R. R. (2009). Pain catastrophizing: a critical review. *Expert Review of Neurotherapeutics, 9*(5), 745–758. doi:10.1586/ern.09.34

Rescorla, R., & Wagner, A. R. (1972). A Theory on Pavlovian 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 York, NY: Appleton-Century-Crofts.

Rottman, B. M., & Ahn, W. (2009). Causal learning about tolerance and sensitization. *Psychonomic Bulletin & Review, 16*(6), 1043–1049. doi: 10.3758/pbr.16.6.1043

Rottman, B. M., & Ahn, W. (2011). Effect of grouping of evidence types on learning about interactions between observed and unobserved causes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(6), 1432–1448. doi: 10.1037/a0024829

Rottman, B. M., & Hastie, R. (2016). Do people reason rationally about causally related events? Markov violations, weak inferences, and failures of explaining away. *Cognitive Psychology, 87*, 88-134. doi: 10.1016/j.cogpsych.2016.05.002

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review, 5*(4), 296–320. doi: 10.1207/s15327957pspr0504_2

Schaller, M., & Maass, A. (1989). Illusory correlation and social categorization: Toward an integration of motivational and cognitive factors in stereotype formation. *Journal of Personality and Social Psychology, 56*(5), 709–721. doi: 10.1037/0022-3514.56.5.709

Schustack, M. W., & Sternberg, R. J. (1981). Evaluation of evidence in causal inference. *Journal of Experimental Psychology: General, 110*(1), 101–120. doi: 10.1037/0096-3445.110.1.101

Shook, N. J., Fazio, R. H., & Eiser, J. R. (2006). Attitude generalization: Similarity, valence, and extremity. *Journal of Experimental Social Psychology, 43*, 641-647. doi: 10.1016/j.jesp.2006.06.005

Spellman, B. A., Price, C. M., & Logan, J. M. (2001). How two causes are different from one: The use of (un)conditional information in Simpson's paradox. *Memory & Cognition, 29*(2), 193-208. doi: 10.3758/bf03194913

Tereyağoğlu, N., Fader, P. S., & Veeraraghavan, S. (2018). Multiattribute loss aversion and reference dependence: Evidence from the performing arts industry. *Management Science, 64*(1), 421–436. doi:10.1287/mnsc.2016.2605

Tomarken, A. J., Mineka, S., & Cook, M. (1989). Fear-relevant selective associations and covariation bias. *Journal of Abnormal Psychology, 98*(4), 381–394. doi: 10.1037/0021-843x.98.4.381

Tomarken, A. J., Sutton, S. K., & Mineka, S. (1995). Fear-relevant illusory correlations: What types of associations promote judgmental bias? *Journal of Abnormal Psychology, 104*(2), 312–326. doi: 10.1037/0021-843x.104.2.312

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*(4481), 453–458. doi: 10.1126/science.7455683

Vaish, A., Grossmann, T., & Woodward, A. (2008). Not all emotions are created equal: The negativity bias in social-emotional development. *Psychological Bulletin, 134*(3), 383–403. doi: 10.1037/0033-2909.134.3.383

Wasserman, E. A., Dorner, W. W., & Kao, S. F. (1990). Contributions of specific cell information to judgments of interevent contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(3), 509–521. doi: 10.1037/0278-7393.16.3.509

Wiemer, J., Mühlberger, A., & Pauli, P. (2014). Illusory correlations between neutral and aversive stimuli can be induced by outcome aversiveness. *Cognition & Emotion, 28*(2), 193-207. doi: 10.1080/02699931.2013.809699

Wiemer, J., & Pauli, P. (2016). Fear-relevant illusory correlations in different fears and anxiety disorders: A review of the literature. *Journal of Anxiety Disorders, 42*, 113–128. doi: 10.1016/j.janxdis.2016.07.003