Summer 7-31-2017

Cognitive Demand and the Outcome Density Effect

Ciara Louise Willett
ciarawillett@pitt.edu

COGNITIVE DEMAND AND THE OUTCOME DENSITY EFFECT

by

Ciara Louise Willett

A Thesis Submitted In Partial Fulfillment of the Requirements for the Master of Science in Experimental Psychology, Thesis with a Concentration in Cognitive Neuroscience

In

The Department of Psychology
Seton Hall University
August, 2017

SETON HALL UNIVERSITY
College of Arts & Sciences

APPROVAL FOR SUCCESSFUL DEFENSE

Masters Candidate, Ciara Willett, has successfully defended and made the required modifications to the text of the master's thesis for the M.S. during this summer 2017.

THESIS COMMITTEE

Mentor:
Kelly M. Goedert, PhD: _____

Committee Member:
Michael Vigorito, PhD: _____

Committee Member:
Bob Rehder, PhD: _____

Acknowledgments

I would like to thank my thesis advisor, Dr. Kelly Goedert, for all of her help and encouragement throughout the completion of my thesis. Without her support, this thesis would not be possible. I am immeasurably grateful for her guidance, which has allowed me to develop as both a student and a researcher.

Thank you to the members of my thesis committee, Dr. Michael Vigorito and Dr. Bob Rehder, who offered instrumental insight during the development of this project.

Finally, I would like to thank Raymond Blattner, Amanda Austin, Adil Yurekli, and Laura Mangus from the Cognition, Perception, and Embodied Thinking Lab at Seton Hall. I am incredibly appreciative for their contributions to the project, help with data collection, and assistance with coding.

Table of Contents

List of Figures

List of Tables

Abstract

Judgments regarding the strength of a cause to produce an outcome do not always follow predictions of normative causal reasoning models (Kao & Wasserman, 1993). In the case of the outcome density effect, individuals' ratings of the strength of a putative cause tend to be greater when the number of observed outcomes is high than when it is low (e.g. Jenkins & Ward, 1965). In the current experiment, I investigated the outcome density effect as a possible heuristic. Participants made causal judgments based on information about the prevalence of headaches in a sample of individuals who did or did not receive a mineral. To manipulate cognitive load, stimuli differed in sample size (n = 24 or 72) and presentation format (scrambled or organized information). Although each stimulus depicted a non-contingent relationship, there were pervasive outcome density effects for causal judgments in each condition. However, the probability of the outcome had no effect on estimates of causal power, suggesting the importance of how causal questions are worded. Manipulations of cognitive load did not influence the magnitude of the outcome density effect for causal judgments or affect causal power estimates. Thus, the outcome density effect does not appear to be used as a heuristic in tasks that vary in cognitive demand, at least as manipulated by sample size and the organization of information.

Introduction

The ability to accurately assess the relationship between a cause and an outcome enables individuals to respond appropriately to events in the world. However, individuals' judgments regarding the strength of a causal relationship do not always coincide with normative measures of covariation and causation. Instead, judgments sometimes reflect either less sophisticated information integration strategies or the use of heuristics (e.g., Fielder, 2009; Fleig, Meiser, Ettlin, & Rummel, 2017; Kao & Wasserman, 1993).

Heuristics are simple strategies or shortcuts that can be used to make judgments or decisions in cases of uncertainty (see Tversky & Kahneman, 1974), potentially leading to non-normative outcomes. For example, the availability heuristic describes instances when individuals make inferences about the frequencies of an event based on easily or quickly recallable (i.e., available) instances of the event. Heuristics can be beneficial in cognitively demanding situations, such as when an individual must complete a task in a limited amount of time. Alternatively, the use of heuristics can be disadvantageous, such as when a doctor incorrectly makes a diagnosis based on easily recallable instances of a disease despite its true rate of occurrence.

One example of non-normative reasoning in causal inference is the outcome density effect. The outcome density effect is observed when individuals increase their judgments of causal strength as the number of outcomes increases, regardless of the actual contingency between the putative cause and outcome. To illustrate, imagine a gardener wants to determine the extent to which two brands of fertilizer generate plant growth. He applies Fertilizer A to 6 of 12 plants in one plot and Fertilizer B to 6 of 12 plants in another. In plot A, 4 of 6 fertilized and 4 of 6 unfertilized plants grow. In plot B, 2 of 6 fertilized and 2 of 6 unfertilized plants grow. An

example of an outcome density effect would be if the gardener assumes Fertilizer A to be stronger because he observed a larger amount of plant growth, although neither fertilizer had an actual influence on plant growth.

In one of the first demonstrations of the outcome density effect, participants were instructed to illuminate a light by choosing to press or not press one of two buttons on each of 60 trials (Jenkins & Ward, 1965, Experiment 1). Participants then judged the degree to which their actions caused the light to turn on using a rating scale from 0 (*no control*) to 100 (*total control*). In three conditions, participants' actions had no influence on the outcome (i.e., the light turning on), but their ratings of control increased as the probability of the outcome increased from 0.13 to 0.50 to 0.80 (see Figure 1).

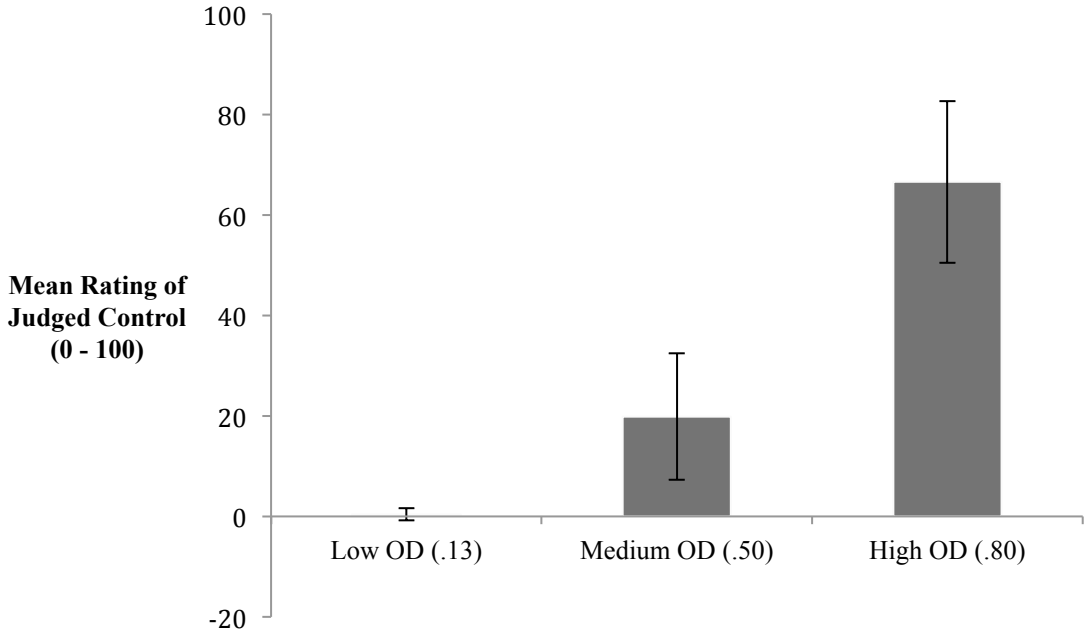


Figure 1. Data from Jenkins and Ward (1965, Experiment 1). When ΔP was held constant at 0, participants' ratings of control over the outcome increased as outcome density (OD) increased, despite no existing relationship between the cause and the outcome. Error bars indicate SD.

Normative models of causal judgment (e.g., Allan, 1980; Cheng, 1997) cannot fully account for outcome density effects, such as the one observed by Jenkins and Ward (1965), in

which there was no relationship between the cause and the outcome. The goal of this paper is to investigate whether outcome density may be a heuristic that people use when under cognitive load. Dual process models of cognition suggest that individuals may employ one of two separate systems – or sets of cognitive processes – for reasoning and decision making (e.g., Sloman, 1996). System 1 is automatic and effortless, resulting in quick, intuitive responses, whereas System 2 involves slower and more effortful reasoning (Kahneman & Frederick, 2001). Thus, heuristics are a component of system 1, which may be overridden by system 2 reasoning. If cognitive demand is too great or there is insufficient time, however, system 2 may fail and individuals will rely upon heuristics (e.g., Finucane, Alhakami, Slovic, & Johnson, 2000). Because the use of heuristics increases with increases in cognitive load, we may be able to test the hypothesis that the outcome density effect is a heuristic by observing whether outcome density effects increase under increasing cognitive load. The remainder of this introduction will review normative models of causal judgment, review evidence for the prevalence of outcome density effects in causal judgment, and discuss manipulations that may affect the cognitive demands of the causal learning task.

In Jenkins and Ward's (1965) early demonstration of outcome density, they used a free-operant trial-by-trial design, in which participants were to determine whether their action (i.e., pressing a button) determined the presence or absence of the outcome. However, outcome density effects in free-operant designs may be influenced by alternative variables such as the temporal contiguity between the participant's action and the outcome (see Vallée-Tourangeau, Murphy, & Baker, 2005). In contrast, many studies on causal learning utilize either a passive trial-by-trial design in which the cause-outcome contingencies are revealed over a series of trials, or a summary design, in which the frequency information is supplied to participants in a

summary format. Because of possible temporal contiguity effects in free-operant designs, this paper focuses on discrete trial-by-trial and summarized designs.

Normative Models of Covariation and Causal Judgment

If the outcome density effect does not reflect normative causal judgments, then what constitutes normative causal inference? Most normative theories of causal inference rely on an estimate of contingency, that is, the relationship between the cause and the outcome. Some normative models, such as the ΔP rule (Allan, 1980) and the power PC model (Cheng, 1997), are rule-based and posit that individuals make causal judgments by calculating statistical probabilities from the evidence. Other normative models, such as the Rescorla-Wagner (1972) model, are associative and propose that causal judgments are based on learning associations between the putative cause and the outcome.

ΔP Rule. According to the ΔP rule, normative causal judgments are based on the probability of an outcome in the presence versus the absence of a putative cause (Allan, 1980). In causal learning tasks regarding binary events, participants typically make causal judgments after they observe the presence or absence of an outcome in the presence or absence of a target cause. This information can be organized into a contingency table so that each cell totals the number of observations for each cause-outcome combination (see Figure 2).

		Outcome		
	Cause	Present	Absent	
	Present	*a*	*b*	$p(\text{o}\vert\text{c}) = \frac{\text{a}}{\text{a}+b}$
	Absent	*c*	*d*	$p(\text{o}\vert\sim\text{c}) = \frac{c}{c+d}$
				$p(\text{o}) = \frac{a+c}{a+b+c+d}$

Figure 2. Sample contingency table displaying the four possible combinations of a cause and outcome. Equations on the right refer to the probability of the outcome occurring in the presence of the cause, p(o|c), the probability of the outcome occurring in the absence of the cause, p(o|~c), and the overall probability of the outcome, *p*(o).

The value of ΔP is equal to the difference between the probability of the outcome occurring in the presence, $p(o|c)$, and the absence, $p(o|\sim c)$, of the cause (Equation 1). When a causal relationship is generative, ΔP is positive, suggesting the cause produces the outcome. Recalling the previous gardener example, ΔP is positive and the fertilizer generates plant growth if the probability of bloomed plants is greater in the presence than the absence of the fertilizer. When a causal relationship is preventive, ΔP is negative, suggesting the cause inhibits the outcome. If the probability of bloomed plants is larger in the fertilizer's absence than its presence, ΔP is negative and the fertilizer prevents plants from blooming. When a relationship is non-contingent, ΔP is equal to zero, suggesting the cause has no effect on the outcome. Thus, if the probability of bloomed plants is the same in the presence and the absence of the fertilizer, ΔP is equal to zero and the fertilizer has no effect on plant growth.

$$\Delta P = p(o|c) - p(o|\sim c) \quad (1)$$

Power PC Model. The ΔP rule suggests that individuals base causal judgments on the extent to which a relationship exists between the cause and the outcome. A correlation between two events, however, does not necessarily imply causation, as alternative causes may be responsible for an effect. The Power PC model (Cheng, 1997) is a newer, more theoretically complex model that proposes a number of boundary conditions and assumptions that individuals make when judging causation from covariation. Of most importance here, Cheng theorized that individuals scale their causal judgments with the base rate of the outcome, $p(o|\sim c)$. For generative causes, causal power is calculated as:

$$q = \frac{\Delta P}{1 - p(o|\sim c)} \quad (2)$$

As $p(o|\sim c)$ increases, the value of the denominator, $1 - p(o|\sim c)$, decreases. Thus, the generative power of the cause, q, will increase as the base rate approaches 1. When $p(o|\sim c)$ is

equal to 1, the value of the denominator is equal to 0 and q is undefined. To use the previous example, if plants always grow in the absence of a fertilizer (i.e., $p(o|{\sim}c) = 1$), a gardener cannot make assumptions regarding the fertilizer's influence on plant growth.

In preventive relationships, however, $p(o|{\sim}c)$ is the critical scalar (see Equation 3). For a preventive cause with constant ΔP, ratings of the strength of the target cause increase as $p(o|{\sim}c)$ approaches 0, but is undefined at 0 (e.g., a gardener cannot make any judgments about the strength of a weed-killer in a plot without weeds).

$$p = \frac{-\Delta P}{p(o|{\sim}c)} \tag{3}$$

Both the generative (Equation 2) and preventive (Equation 3) equations can be used to model the strength of non-contingent relationships. If ΔP is equal to zero, the numerators of p and q will be equal to zero, so that causal strength is also zero. Therefore, for non-contingent relations, both ΔP and the power PC model predict causal estimates to be zero.

Rescorla-Wagner Model (RWM). The Rescorla-Wagner (1972) model, initially designed to describe animal learning in classical conditioning, models how animals learn associations between two stimuli over time. When adapted to explain human causal reasoning, the RWM predicts causal judgments to change with new information over a series of trials. Individuals' causal judgments are described as the change in associative strength (ΔV) between the putative cause (C) and an outcome (E) on the current trial (i), or $\Delta V_{C\text{-}Ei}$ (see Equation 4).

$$\Delta V_{C\text{-}Ei} = \alpha \ \beta \ (\lambda - \Sigma V_i) \tag{4}$$

During a trial, individuals observe the presence or absence of a cause in the presence or absence of an outcome, which alters the associative strength between the two. ΣV_i represents the sum of these associative strengths in that trial (i). The rate of learning for the associative strengths is modulated by two parameters for the outcome (α) and the cause (β). As new

information is introduced, the value of ΣV_i becomes closer to the learning asymptote (λ), the maximum value of the association. Over time, the learner will update previous beliefs in the associative strengths with new information, so that the RWM is an error-correcting algorithm. When there is only one potential cause, as described above, the asymptotic result of the RWM is mathematically equal to ΔP and the model predicts causal judgments to reflect the contingency of the cause-outcome relationship (see Chapman & Robbins, 1990).

Outcome Density Effect and Models of Causal Judgment. Despite evidence that the ΔP rule (e.g., Lober & Shanks, 2000) and the power PC model (e.g., Buehner & Cheng, 1997) can accurately predict causal judgments in many circumstances, neither model fully accounts for the outcome density effect. When ΔP is held constant, the ΔP rule does not predict changes in causal judgments based on increased outcome density as $p(o|c)$ and $p(o|\sim c)$ will change proportionately for generative, preventive, and non-contingent relationships (Equation 1).

Power PC predicts outcome density effects in contingent, but not in non-contingent relations. For non-contingent relations, ΔP is equal to zero, and the numerator of the power PC equations for both generative (Equation 2) and preventive (Equation 3) relations will be equal to zero. Thus, $p(o)$ does not influence predicted causal power estimates. The model does predict, however, that $p(o)$ will affect causal judgments for generative and preventive relationships. If $p(o)$ increases, $p(o|\sim c)$ will also increase, causing the denominator of the power PC model to decrease for generative (see Equation 2) and increase for preventive (see Equation 3) relationships. As the denominator decreases, power PC predicts that judgments of causal strength will increase for generative and decrease for preventive relationships.

In contrast to the predictions of the power PC and ΔP models, the RWM suggests that outcome density effects may occur early on in learning for non-contingent causes, depending on

the sequence of the training trials (Musca, Vadillo, Blanco, & Matute, 2010). As individuals review more trials, the sum of associative strengths on the current trial (ΣV_i) will approach the learning asymptote (λ), which is the maximum learned strength of the association. That is, this model predicts that outcome density effects in pre-asymptotic causal judgments will dissipate as individuals update their beliefs with new information and ΣV_i approaches λ. The RWM, however, only makes predictions for trial-by-trial learning – the design utilized in most studies of outcome density. In the current experiment, I implemented a summarized design. Therefore, the RWM does not apply.

Demonstrations of the Outcome Density Effect

No models of causal learning can fully account for the outcome density effect. For example, ΔP does not predict a role of outcome density in causal judgments and the power PC model does not predict outcome density effects for non-contingent relationships. Still, outcome density effects are well documented in the literature for non-contingent, generative, and preventive relationships.

In causal learning tasks, participants judge the extent to which a relationship exists between the cause and the outcome at the end of a series of trials or after reviewing the summary information either using a bi-directional scale from -100 (*the cause always prevents the outcome*) to +100 (*the cause always produces the outcome*) or a unidirectional scale from 0 (*no relationship*) to +100 (*cause always has an effect on the outcome*).

Outcome Density Effects for Non-contingent Relationships. The majority of outcome density research looks at the effect when ΔP is equal to zero. As can be seen across studies in Table 1, causal judgments for non-contingent relationships tend to increase as the probability of the outcome, $p(o)$, increases (e.g., Allan, Siegel, & Tangen, 2005).

Table 1

Outcome Density Effects for Non-contingent Relationships

Study	Presentation Type	Sample Size	Findings
Allan et al. (2005)	TBT	60	OD effect. Causal ratings increased as *p*(o) increased from low (.2) to medium (.5) to high (.8) when rated after 20, 40, and 60 trials.
Allan et al. (2008, Exp. 3)	TBT	60	OD effect. More likely to classify a relationship as 'strong' as opposed to 'weak' in the high (.7) than in the low (.3) outcome density condition.
Blanco et al. (2013, Exp. 1)	TBT	100	OD effect. Causal ratings were greater in the high (.8) than in the low (.2) condition.
Buehner & Cheng (1997, Exp. 1a, 1b)	TBT	16	OD effect. As the probability of the outcome increased, causal ratings decreased in the preventive scenario (1a) and increased in the generative (1b) scenario.
Buehner & Cheng (1997, Exp. 2a, 2b)	Summary (Pie Chart)	100	OD effect. There was a negative and positive linear trend as *p*(o) increased for relationships framed as preventive (2a) and generative (2b), respectively. Further analysis suggests that the effect is driven by an OD effect at extreme values of *p*(o), specifically when *p*(o) = 0.
Buehner et al. (2003, Exp. 1)	TBT	16	OD effect. In a between-subjects design, researchers saw a positive and a negative linear OD trend for causal ratings of participants in the generative scenario and participants in the preventive scenario, respectively.
Buehner et al. (2003, Exp. 2)	Summary (Countable Images)	72	No OD effect. Researchers specifically asked if the cause influenced the outcome and 47/50 participants said the cause had no effect.
Buehner et al. (2003, Exp. 3)	TBT	24	OD effect. Although most (20/31) participants said there was no relationship, there was an OD effect for those who believed there was a relationship.
Crump et al. (2007)	TBT	60	OD effect. Causal judgments were greater in the high (.8) than in the low (.2) condition.
Musca et al. (2010)	TBT	50	OD effect. Causal judgments were greater in the high (.8) than in the low (.2) condition.
Perales & Shanks (2003, Exp. 2)	TBT	n/a[1]	OD effect. Causal judgments were greater in the high (.8) than in the low (.2) condition.

Note. OD = outcome density effect. TBT = trial-by-trial design; Sample size = number of trials in a TBT design or total number of instances in a summary chart.

[1] Participants studied as many trials as needed to make a causal judgment that was 100% reliable, for an average of 35.6 and 36.6 trials when *p*(o) = .2 and .8, respectively.

To walk through a typical study, Allan et al. (2005) gave participants information about the rate of bacteria survival (outcome) when a chemical (cause) was or was not added to the bacteria. In each trial, participants learned whether the cause was present or absent and then made a prediction about whether the bacteria would survive (outcome present) or not (outcome absent). After 60 trials, they rated the effect of the chemical on a scale of -100 (*negative effect on survival*) to +100 (*positive effect on survival*). As seen in Figure 3, causal judgments increased when the probability of bacteria survival was high even though the relationship was non-contingent.

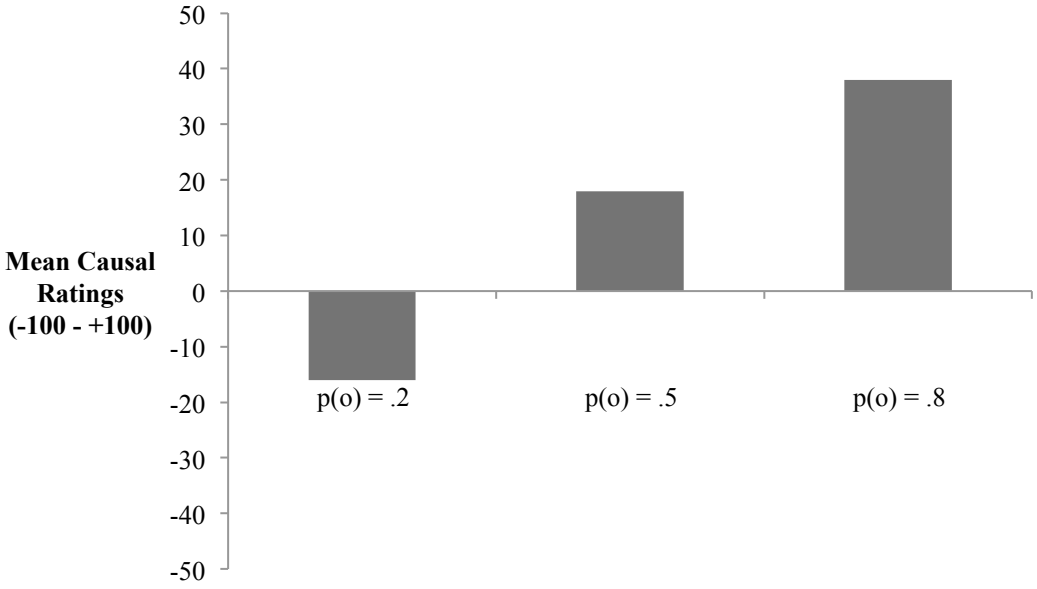


Figure 3. Estimated mean causal ratings for non-contingent causes with low, medium, or high levels of outcome density (OD) after reviewing 60 trials (causal ratings were approximated based on the data presented in Figure 3 of Allan et al., 2005). Causal judgments increased as the probability of the outcome, *p*(o), increased.

Although participants' causal judgments revealed an outcome density effect, *p*(o) had no effect on participants' predictive judgments. Using participants' predictions, Allan et al. (2005) calculated *p*(o|c) and *p*(o|~c) to determine an 'observed' value of ΔP. These observed ΔP values closely reflected normative expectations set by the ΔP rule (i.e., a judgment of zero), suggesting

that while participants' final causal estimates were influenced by the probability of the outcome, they were able to detect the actual contingency. Therefore, the outcome density effect cannot simply be explained by an incorrect perception of cause-outcome combinations. Although the outcome density effect for causal judgments is pervasive in studies of non-contingent relationships, it is incompatible with normative models of causal judgment and its nature is uncertain.

Outcome Density Effects for Generative Relationships. In Allan et al.'s (2005) study, participants also made causal judgments about a generative relationship. As they found with non-contingent relationships, mean causal judgments increased as $p(o)$ increased when ΔP was equal to .467 (see Figure 4).

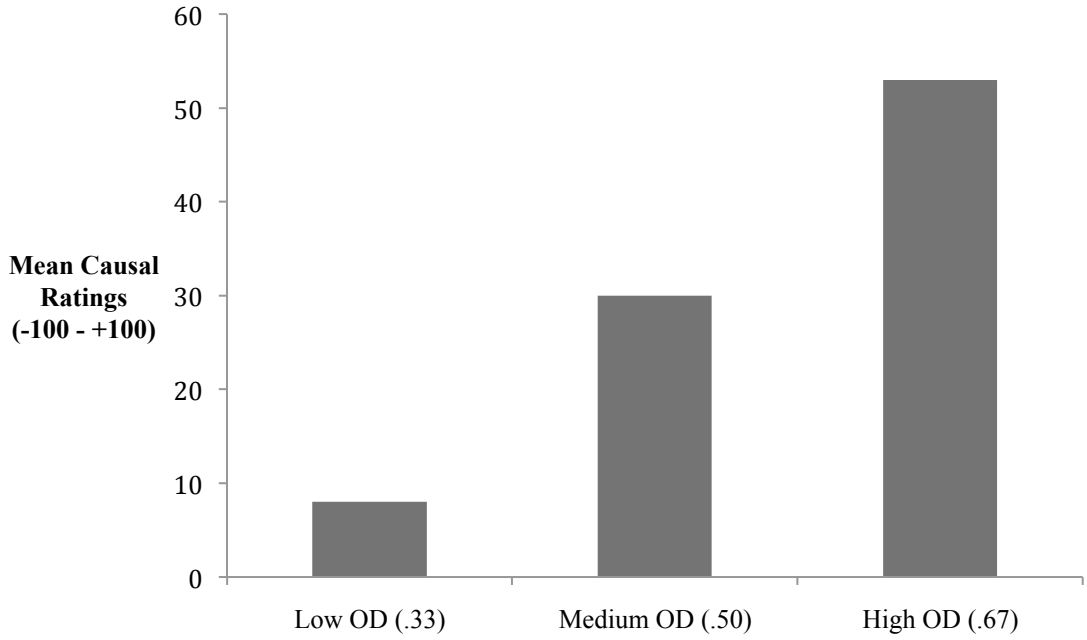


Figure 4. Approximates of participants' mean causal ratings after reviewing 60 trials (estimates based on Figure 3 in Allan et al., 2005). Mean causal ratings increased as the probability of the outcome, $p(o)$, increased from low (.33) to medium (.50) to high (.67).

Table 2 summarizes the studies investigating outcome density in generative relationships, in which causal judgments frequently increase as $p(o)$ increases. Although several studies

document outcome density effects for generative relationships, the presence of an outcome density effect is not consistently supported at each value of ΔP (e.g., Buehner, Cheng, & Clifford, 2003). In addition, outcome density effects for generative relationships may be sensitive to presentation format (e.g., Wasserman, Elek, Chatlosh, & Baker, 1993), as will be discussed later in this paper.

Table 2

Outcome Density Effects for Generative Relationships

Study	Presentation Format	Sample Size	ΔP	Findings
Allan et al. (2005)	TBT	60	.467	Mixed results. OD effect as *p*(o) increased from low (.333) to medium (.467) to high (.667). Found for causal judgments made after 40 and 60 but not 20 trials.
Allan et al. (2008, Exp. 3)	TBT	80	.1, .2, .3, .4, .5, .6	OD effect. The probability that a participant would classify a relationship as "strong" as opposed to "weak" increased as *p*(o) increased.
Buehner & Cheng (1997, Exp. 1b)	TBT	16	.25, .5, .75	Mixed results. Significant OD effect for ΔP = .25 and .50, approached significance for ΔP = .75 (p = .052).
Buehner & Cheng (1997, Exp. 2b)	Summary (Pie Chart)	100	.25, .5, .75	OD effect for each value of ΔP.
Buehner et al. (2003, Exp. 1)	TBT	16	.25, .5, .75	Mixed results. OD effect for ΔP = .25 and .5 but not .75.
Buehner et al. (2003, Exp. 2)	Summary (Countable Images)	72	.5	OD effect as *p*(o) increased from low (.25) to medium (.58) to high (.75).
Buehner et al. (2003, Exp. 3)	TBT	24	.5	OD effect as *p*(o) increased from low (.25) to medium (.58) to high (.75).
Crump et al. (2007)	TBT	60	.467	OD effect as *p*(o) increased from .33 to .67.
Lober & Shanks (2000, Exp. 3)	TBT	60	.4	OD effect as *p*(o) increased from .2 to .6 to .8. Evident OD effect each time participants gave a causal judgment (after viewing 20, 40, or 60 trials).
Lober & Shanks (2000, Exp. 6)	Summary (Pie Chart)	100	.4	No OD effect. There was a nonlinear trend for causal ratings of *p*(o) = .2, .6, and .8.

Note. OD = outcome density effect. TBT = trial-by-trial design; Sample size = number of trials in a TBT design or total number of instances in a summary chart.

Outcome Density Effects for Preventive Relationships. In contrast to generative relationships, outcome density effects occur in preventive relationships when causal judgments become increasingly positive as *p*(o) increases. Strong preventive causes should inhibit the outcome; therefore, outcome density effects in preventive causes will be shown by decreases in preventive strength as the probability of the outcome increases.

For example, in Buehner et al. (2003), participants rated the extent to which a medicine prevented headaches based on summarized information about the prevalence of headaches in a sample of 72 individuals, half of whom received the medicine. When ΔP was held constant at -.5, causal judgments were increasingly negative (i.e., judged as increasingly more preventive) as the number of headaches decreased (see Figure 5).

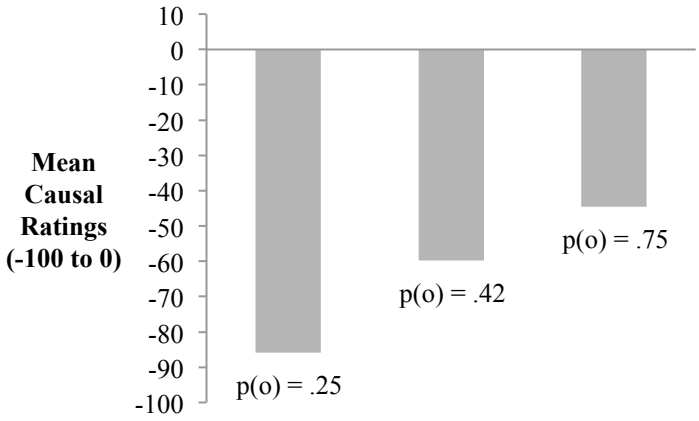


Figure 5. Participants' mean causal ratings after reviewing summarized information about a preventive relationship of ΔP = -.5 (Buehner et al., 2002). The absolute value of participants' ratings decreased and became increasingly less negative as *p*(o) increased, demonstrating an outcome density effect for preventive relations.

There is a considerable lack of research regarding the influence of *p*(o) for preventive causes while holding ΔP constant, but the few extant studies are summarized in Table 3. The results of these studies support the idea that the outcome density effects are not limited to generative or non-contingent relationships, but also occur for preventive causes.

Table 3

Outcome Density Effects for Preventive Relationships

Study	Presentation Format	Sample Size	|ΔP|	Findings
Buehner & Cheng (1997, Exp. 1a)	TBT	16	.25, .5, .75	OD effect for each value of ΔP.
Buehner & Cheng (1997, Exp. 2a)	Summary (Pie Chart)	100	.25, .5, .75	Mixed results. Causal ratings significantly decreased as *p*(o) increased for ΔP = .25 and .75 but not ΔP = .5.
Buehner et al. (2003, Exp. 1)	TBT	16	.25, .5, .75	OD effect for each value of ΔP.
Buehner et al. (2003, Exp. 2)	Summary (Countable Images)	72	.5	OD effect for each value of ΔP as *p*(o) increased from low (.25) to medium (.58) to high (.75).
Buehner et al. (2003, Exp. 3)	TBT	24	.5	OD effect for each value of ΔP as *p*(o) increased from low (.25) to medium (.58) to high (.75).

Note. OD = outcome density. TBT = trial-by-trial design; Sample size = the number of trials in a TBT design or total number of instances in a summary chart. ΔP values refer to the absolute value of a negative (preventive) ΔP. An OD effect for a preventive relationship indicates that causal judgments decrease as *p*(o) increases.

Outcome Density Effects in Other Dependent Measures. As previously described, causal judgments are sensitive to changes in outcome density for non-contingent, generative, and preventive causes. Other studies have demonstrated the pervasiveness of the outcome density effect using other dependent measures. For example, individuals are more likely to classify a non-contingent or generative relationship as "strong" as opposed to "weak" if the probability of the outcome is high than if it is low (Allan, Hannah, Crump, & Siegel, 2008). In another study, researchers found individuals' actual behavior was sensitive to changes in outcome density (Matute, Steegen, & Vadillo, 2004). Participants were more likely to exhibit preparatory behavior during a video game when the probability of the outcome was high as opposed to moderate. This suggests that outcome density affects not only numerical causal judgments but also the overall perception of a relationship and how individuals use outcome density information to interact in the world.

Potential Manipulations for Influencing Cognitive Demand

Some normative models of causal learning, such as the power PC model (Cheng, 1997), account for outcome density effects in generative and preventive relationships. However, no model can account for outcome density effects when a causal relationship is non-contingent. Still, individuals *are* sensitive to changes in the probability of the outcome, as outcome density effects are pervasive in the literature. If normative models expect individuals to detect non-contingent relationships as non-causal, then why do outcome density effects occur? The current experiment explores the possibility that individuals use outcome density as a heuristic during cognitively demanding causal learning tasks.

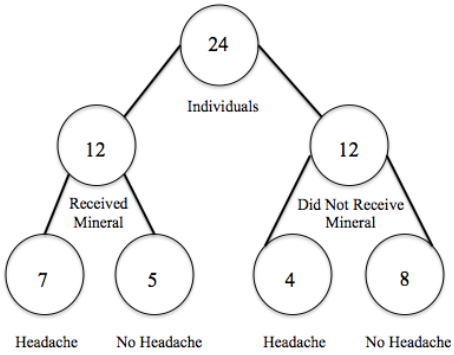
Presentation Format. One way to manipulate cognitive demand is through the format for presenting the contingency information. In a trial-by-trial design, participants are shown one cause-outcome combination per trial. In a summarized design, the cause-outcome information is displayed simultaneously in a variety of possible formats (see Figure 6). For both trial-by-trial and summarized designs, the different cause-outcome combinations represent the four cells of the contingency table (see Figure 2).

A. Countable Images

B. Simple Phrases

7 of 12 individuals who received the mineral got a headache.
5 of 12 individuals who received the mineral did not get a headache.
4 of 12 individuals who did not receive the mineral got a headache.
8 of 12 individuals who did not receive the mineral did not get a headache.

C. Frequency Tree



D. Pie Chart

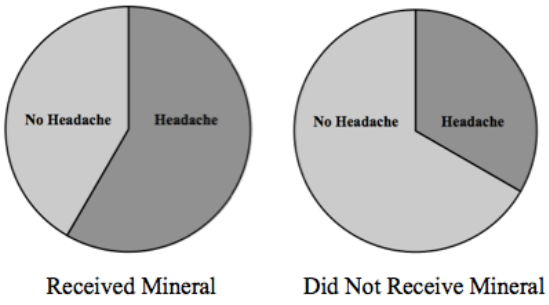


Figure 6. Four methods for organizing information in a summary design.

Individuals may rely on different strategies to make causal judgments depending on the presentation format. In summarized designs, individuals appear to make causal judgments based on the ΔP rule, which assumes equal weighting of the contingency cells (Kao & Wasserman, 1993; Ward & Jenkins, 1965). In trial-by-trial designs, however, individuals appear to implement less sophisticated strategies and place greater weight on different cells. Causal judgments in trial-by-trial designs may reflect use of the cell *A* strategy, in which individuals rely more heavily on cell *A* (i.e., the frequency of the outcome in the presence of the cause; Kao & Wasserman, 1993). An alternative strategy is the confirming cases heuristic, in which individuals rely on cell *A* and cell *D* (i.e., the frequency of the outcome not occurring in the absence of the cause; Ward & Jenkins, 1965).

Could increased use of the ΔP rule in summarized designs be due to the organization of information? As seen in Figure 6, summarized designs present cause-outcome information in a format that is similar to how the frequencies would be presented in a contingency table.

Therefore, the organization of information may make it easier to calculate and compare conditional probabilities. Ward and Jenkins (1965) actually used a contingency table to disseminate cause-outcome information to participants in the summarized format condition. In another condition, participants reviewed trial-by-trial information and then saw the same information in a summarized contingency table. Although participants in the summarized-only condition made causal judgments based on the ΔP rule, participants that saw both formats appeared to rely on the confirming cases heuristic, the same strategy implemented by participants who only saw trial-by-trial information.

Trial-by-trial designs place a greater demand on working memory, as participants must keep track of the different cause-outcome combinations before making a final causal judgment. In the Ward and Jenkins (1965) study, participants who only saw trial-by-trial information and participants who saw both trial-by-trial and summarized information may have relied on the confirming cases heuristic because of the increased cognitive load. Participants in the summarized-only condition, however, may have more easily implemented the ΔP rule due to decreased cognitive demand and the organization of contingency information.

Presentation format may also modulate the extent to which outcome density effects are observed. If trial-by-trial designs are more cognitively demanding, then individuals may rely on outcome density as a strategy to make causal judgments. In contrast, summarized designs may not elicit strong outcome density effects because the organization of information leads individuals to make normative judgments based on the ΔP rule. However, only three studies have specifically manipulated the probability of the outcome for constant values of ΔP using both trial-by-trial and summary designs (see Table 4).

Table 4

Outcome Density Effects in Summary vs. Trial-by-Trial Designs

Study	ΔP	Presentation Type	Sample Size	Findings
Buehner & Cheng (1997)	Generative	TBT (Exp. 1b)	16	Mixed results. OD effect for ΔP = .25, .50. OD trend approached significance for ΔP = .75.
		Summary (Pie Chart, Exp. 2b)	100	Mixed results. OD effect for ΔP = .25, .75, but not ΔP = .50.
	Preventive	TBT (Exp. 1a)	16	OD effect for ΔP = -.25, -.50, -.75.
		Summary (Pie Chart, Exp. 2a)	100	OD effect for ΔP = - .25, -.50, -.75.
	Non-contingent	TBT (Exp. 1)	16	OD effect. Causal ratings increased or decreased as *p*(o) increased for participants who saw the relationship framed as generative (Exp. 1b) or preventive (Exp. 1a), respectively.
		Summary (Pie Chart, Exp. 2)	100	OD effect. Causal ratings increased for participants who saw the relationship framed as generative (Exp. 2b); the trend seemed to be due to close-to-zero ratings for *p*(o) = 0 as ratings were not different for *p*(o) = .25, .50, and .75. Causal ratings decreased for participants who saw the relationship framed as preventive (Exp. 2a); the trend seemed to be due to close-to-zero ratings for *p*(o) = 1 as ratings were not different for *p*(o) = .25, .50, and .75.
Buehner et al. (2003)	Generative	TBT (Exp. 1)	16	Mixed results. Observed OD effect for ΔP = .25 and .50, but not ΔP = .75.
		TBT (Exp. 3)	24	OD effect for ΔP = .50.
		Summary (Countable Images, Exp. 2)	72	OD effect for ΔP = .50.
	Preventive	TBT (Exp. 1)	16	OD effect for ΔP = -.25, -.50, -.75.
		TBT (Exp. 3)	24	OD effect for ΔP = -.50.
		Summary (Countable Images, Exp. 2)	72	OD effect for ΔP = -.50.
	Non-contingent	TBT (Exp. 1)	16	OD effect. Causal ratings increased or decreased if the relationship

				framed as generative or preventive, respectively.
		TBT (Exp. 3)	24	OD effect. Most (20/31) participants said there was no relationship, but there was an OD effect for the 11 who believed there was a generative relationship.
		TBT (Exp. 4)	24	No OD effect. No relationship between causal ratings and *p*(o); participants were given randomly "spot-checks" between trials and asked to state whether the relationship was generative, preventive, or non-contingent.
		Summary (Countable Images, Exp. 2)	72	No relationship. 47/50 participants said there was no relationship.
Lober & Shanks (2000)	Generative	TBT (Exp. 3)	60	OD effect for ΔP = .40.
		Summary (Pie Chart, Exp. 6)	100	No OD effect for ΔP = .40.

Note. TBT = trial-by-trial design; sample size = number of trials in a TBT design or total number of instances in a summary chart; OD effect = increases in causal judgments for generative and non-contingent causes or decreases in causal judgments for preventive causes as the probability of the outcome, *p*(o), increases.

In the trial-by-trial conditions, the studies in Table 4 found fairly reliable outcome density effects for generative, preventive, and non-contingent relationships. In the summarized conditions, two of the studies revealed outcome density effects for generative and preventive conditions at most levels of ΔP (Buehner & Cheng, 1997; Buehner et al., 2003). Lober and Shanks (2000), however, did not find evidence of an outcome density effect when individuals reviewed information about a generative relationship in a pie chart. For the two studies that examined non-contingent relationships, Buehner and Cheng (1997) found evidence of an outcome density effect. In contrast, a strong majority (94%) of participants in Buehner et al.'s (2003) study explicitly stated there was no relationship at either level of outcome density.

From the limited research on outcome density in summarized vs. trial-by-trial designs, it appears that the use of outcome density to guide causal judgments is somewhat dependent on

presentation format. As previously discussed, this may be due to how the information is organized within a summarized design. One study examined how outcome density affected causal judgments about non-contingent and generative relationships within multiple summarized designs (see Figures 6 A, B, and C), finding evidence of outcome density effects in each condition (Vallée-Tourangeau, Payton, & Murphy, 2008). However, individuals' ability to distinguish between the non-contingent and contingent relationships (i.e., give higher causal ratings in the contingent condition) varied depending on the type of summarized design.

Vallée-Tourangeau, Payton, and Murphy (2008) gave participants information about non-contingent and generative relationships using countable images, simple phrases, or a frequency tree (see Figure 6 A, B, and C). Causal judgments increased as the probability of the outcome increased, producing an outcome density effect in each of the three display conditions. However, the ability to distinguish between a non-contingent relationship and contingent relationship (i.e. give higher causal ratings in the contingent condition) varied depending upon the summary display. When participants saw causal information in simple sentences, they could not distinguish between contingency values at either $p(o)$ whereas they could do so for both high and low outcome density levels if the information was organized in frequency trees. When participants saw countable objects, ratings of contingent relationships were only higher than non-contingent relationships if $p(o)$ was high.

Although Vallée-Tourangeau et al. (2008) found outcome density effects in multiple types of summarized designs, the information presented in Table 4 suggests that outcome density effects may not be as prominent in summarized designs as they are in trial-by-trial presentations. For example, to maintain a constant value of $\Delta P = 0$, $p(o|c)$ and $p(o|\sim c)$ must be equal such that the number of outcomes is the same in the presence and the absence of the cause. Therefore, it

may have been easy to identify a non-contingent relationship in the summarized formats used previously.

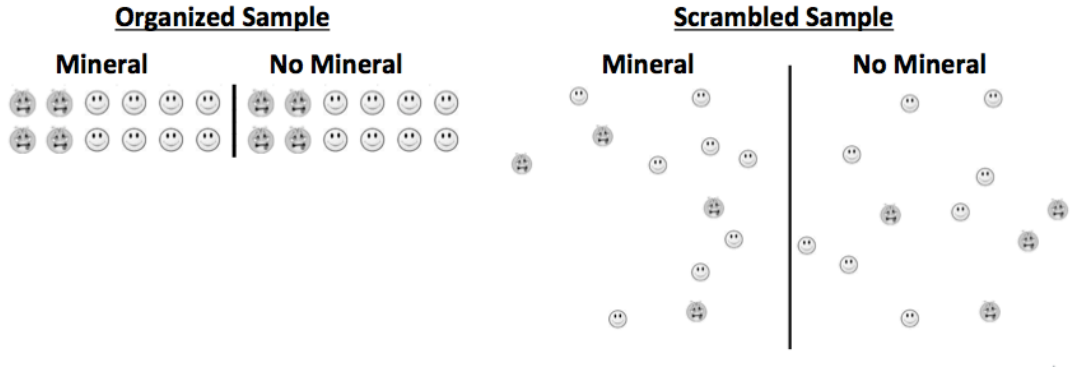


Figure 7. Example of stimuli to be used in the present experiment, similar to the stimuli used in Buehner et al. (2003). In the organized sample, participants may more easily identify the relationship as non-contingent because the proportion of outcomes is clearly the same in the presence and the absence of the cause. The same information is presented in the scrambled sample, but in a less identifiable manner.

To address this concern, participants in the current experiment reviewed countable images in summarized designs that were either organized or scrambled (see Figure 7). In the organized condition, the prevalence of headaches was clearly the same in the presence and absence of the mineral (i.e., the putative cause), similar to that of Buehner et al.'s (2003) non-contingent example. In the scrambled condition, participants reviewed a random organization of the same information. Whereas the organized design may obviate the need for attending to outcome density information, the scrambled design may prevent participants from easily recognizing the non-contingent relationship. Because the scrambled information is less discernible (i.e., more cognitively demanding), participants may be more likely to rely on a heuristic to guide causal judgments. If individuals demonstrate greater outcome density effects in the scrambled than the organized condition, this would suggest that individuals rely on outcome density effect as a heuristic to guide causal judgments during more cognitively demanding tasks.

Reasoning with Large Numbers. A second potential manipulation for influencing cognitive load is asking participants to make casual judgments about larger versus smaller samples of data. Larger numbers are represented less distinctly in the brain and prove to be more difficult compared to working with smaller numbers (Göbel et al., 2001; Nieder & Merten, 2007). If larger numbers are more difficult to work with, perhaps people are more likely to employ a heuristic, such as relying on *p*(o), as a cue to causality when making inferences from larger numbers.

The importance of numerical size is illustrated in magnitude comparison tasks. In a magnitude comparison task, participants determine whether a target number is greater or less than a previously established reference number (e.g., Moyer & Landauer, 1976). To illustrate, Göbel et al. (2001) asked participants if target numbers were greater or less than the reference numbers 5 and 65. Participants were quicker and more accurate at this task as the distance between the target and the reference number increased. For example, they would be quicker to determine that 1 is less than 5 than to determine that 4 is less than 5. This effect was more pronounced when participants compared target numbers with the reference number 65. The difference in reaction times between numbers close (e.g., 64) and far (e.g., 20) from 65 was greater than the difference in reaction times between numbers close (e.g., 4) and far (e.g., 1) from 5, producing what is called a size effect. These results suggest that larger numbers are less distinctly represented than smaller numbers. Single-cell recording work supports this interpretation. When rhesus monkeys were shown sets of black dots on a screen of varying magnitude, recordings of single unit neuron activity suggested firing specificity for small, but not large, numerical magnitudes (Nieder & Merten, 2007).

Therefore, during a causal reasoning task, perhaps it is more difficult to reason with larger sample sizes because larger numbers have less distinct mental representations. Both the ΔP rule and power PC model suggest that individuals base causal ratings on comparisons of frequency information (i.e., conditional probabilities). As the magnitude of these frequencies increase, individuals may find it more difficult to make causal judgments using these statistical rules and may rely upon heuristics to guide their causal judgments during tasks. Thus, the outcome density effect, as a potential heuristic, may increase when individuals make causal judgments about larger samples.

Current Experiment

The primary purpose of the current experiment was to investigate the possible use of outcome density as a heuristic in causal learning. I did so by manipulating the cognitive demands of the task. If individuals rely more on outcome density over normative models (e.g., ΔP rule) to guide causal judgments in a cognitively demanding task, then this would suggest outcome density is a heuristic.

Participants made causal judgments about the extent to which a relationship existed between various minerals (i.e., putative causes) and headaches (i.e., outcome). Their job was to determine whether each mineral produced headaches as a side effect, prevented headaches, or had no effect on headaches. In a summarized design, participants learned about the prevalence of headaches in a sample of individuals, half of whom received a mineral. Each mineral was non-contingent such that the probability of the outcome (i.e., headaches) was the same in the presence and absence of the putative cause (i.e., mineral). After reviewing each stimulus, participants made a causal judgment about the extent to which the mineral generated, prevented, or had no effect on headaches. If participants gave higher causal judgments for a mineral when the

probability of headaches was high [p(o) = .667] than when the probability of headaches was low [p(o) = .333], this would be evidence of an outcome density effect.

To examine the use of outcome density as a possible heuristic, I manipulated cognitive demand in two ways. First, I modified the sample size of the stimuli such that participants learned about a small or large sample of either 24 or 72 individuals, respectively. With the exception of streamed-trial designs (e.g., Crump et al., 2007), the majority of trial-by-trial outcome density research uses small sample sizes of 16 to 24. I chose 24 as the small sample condition not only to compare with the widely used trial-by-trial design, but also because 24 is large enough to allow for an outcome density manipulation. Additionally, I chose 72 for the more cognitively demanding, larger sample condition, because magnitude comparison tasks suggest numbers this large should be represented less distinctly (Göbel et al. 2001).

Second, because it may be easier to detect non-contingent relationships in an organized image where the probability of headaches is clearly the same in the presence and absence of the mineral, I altered the presentation format so that the cause-outcome information was either organized or scrambled (see Figure 7). If the cause-outcome information is less distinct, individuals may be more likely to rely on heuristics when making causal judgments.

Finally, others have speculated that individuals may only rely on outcome density when they fail to completely understand random assignment and the independence of alternative causes (Buehner et al., 2003). If participants understand that individuals are randomly assigned to receive or not receive a mineral, they should understand that the probability of headaches prior to the study is the same in both groups. If participants understand independence of alternative causes, they should understand that the putative strength of the mineral to affect headaches is independent of alternative background causes. Therefore, when the proportion of headaches is

the same in the group that receives a mineral and the group that does not receive a mineral, headaches must be due to an alternative cause. I assessed whether participants who could accurately answer questions about random assignment and the independence of alternative causes would be less reliant on outcome density as a heuristic when making causal judgments. If they have a strong understanding of experimental design, they may be less likely to use heuristics and more likely to provide normative causal judgments.

Methods

Participants

One hundred and seventy undergraduate students participated in exchange for course credit. An a-priori power analysis (with G*Power 3.1) yielded a sample size of 70 as sufficient to detect a small ($\eta_p^2 = .03$) between-within interaction at a power of 0.95 in a repeated-measures design using an alpha level of .05. After collecting data from 70 participants, initial data analyses suggested that a larger sample was necessary to determine whether individual differences in the understanding of random assignment and alternative causes interacted with outcome density. After which, I aimed to double the sample size. Data from nine participants were not analyzed because, contrary to instructions, they used scratch paper during the experiment. The final sample of 161 participants (age: $M = 19.30$, $SD = 2.86$) consisted mostly of women (118 women, 41 men, 2 no responses) and undergraduate freshmen (74 freshmen, 64 sophomores, 11 juniors, 12 seniors).

Design

The experiment was a 2 (outcome density: high, low) x 2 (sample size: small, large) x 2 (presentation format: scrambled, organized) x 2 (order of presentation format: scrambled first, organized first) mixed-design. Outcome density, sample size, and presentation format were manipulated within-groups so that all participants learned about the same eight minerals. The order of presentation format was counterbalanced between participants so that participants first reviewed either the block of four organized or the block of four scrambled minerals.

Materials

Cover story. Participants read a cover story (adapted from Liljeholm & Cheng, 2009) in which they imagined working at a pharmaceutical company that was developing an allergy

medicine comprised of several minerals, each of which could cause or prevent headaches as a side effect. The participants evaluated the results of studies done with eight different fictional minerals to determine the effect each mineral had on headaches. Each mineral had a unique alphanumeric label and was investigated by a different fictional laboratory to emphasize that each mineral was distinct.

Comprehension check questions. To evaluate participants' understanding of random assignment and the independence of alternative background causes, they answered two comprehension check questions (from Buehner et al., 2003) prior to evaluating the minerals. For the comprehension check of random assignment, participants were told to assume individuals were randomly assigned to one of two groups by a coin toss. If the coin landed on heads, the individual was placed in the group that received the mineral. If the coin landed on tails, the individual was placed in the group that did not receive the mineral. Participants then indicated whether they expected the proportion of headaches in the group that received the mineral to be greater than, less than, or about the same as the proportion of headaches in the group that did not receive the mineral before the study began. An "about the same" response suggested an understanding of random assignment.

For assessing participants' knowledge of the independence of alternative causes, participants were told that 50% of the individuals who received the mineral and 50% of the individuals who did not receive the mineral had a headache. Using this information, participants responded yes or no as to whether the headaches in the group of individuals who received the mineral could be attributed to the mineral. A "no" response suggested an understanding of the independence of alternative background causes.

Stimuli. On each trial, participants learned summarized information (i.e., countable images) about one mineral. I created eight different stimuli to represent each combination of the outcome density (high, low), sample size (small, large), and presentation format (scrambled, organized) conditions (see Table 5, see also Figures 8 and 9).

Table 5

Experimental Design.

				Cell Frequencies			
Sample Size	*p*(o)	*p*(o|c)	*p*(o|~c)	*A*	*B*	*C*	*D*
24	0.333	0.333	0.333	4	8	4	8
24	0.667	0.667	0.667	8	4	8	4
72	0.333	0.333	0.333	12	24	12	24
72	0.667	0.667	0.667	24	12	24	12

For each stimulus, half of the sample received the mineral (left side) and half of the sample did not receive the mineral (right side), such that the probability of receiving the mineral (i.e., probability of the cause) was $p(c) = .5$. The number of headaches present (i.e., cells *a* and *c*, sick emoticons) and headaches absent (i.e., cells *b* and *d*, healthy emoticons) varied for each stimulus depending on outcome density [$p(o) = .333, .667$] and sample size ($n = 24, 72$). Because the probability of the outcome was the same in the presence and the absence of the putative cause, however, ΔP was equal to zero for all eight minerals.

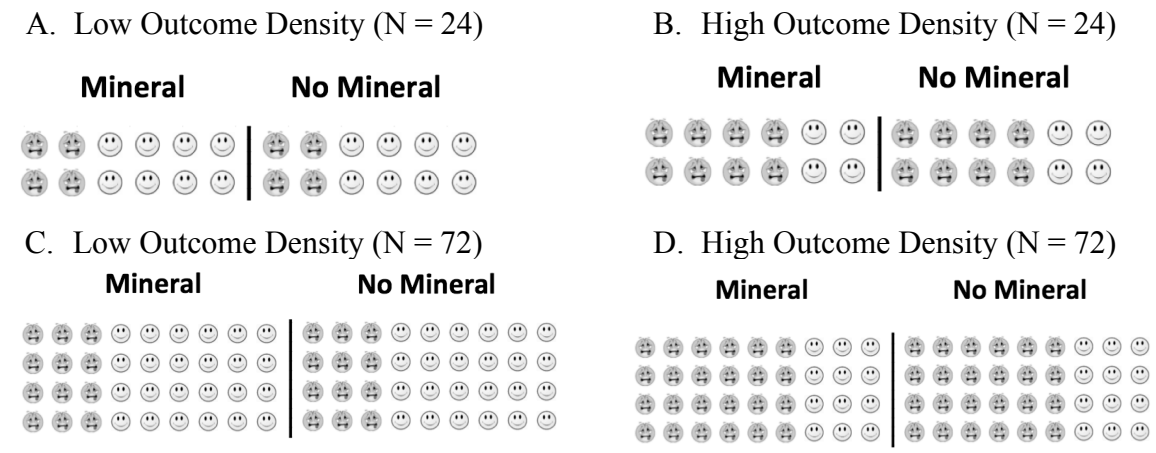


Figure 8. Organized presentation format. Half of the sample received the mineral (left side of the image) and half of the sample did not (right side).

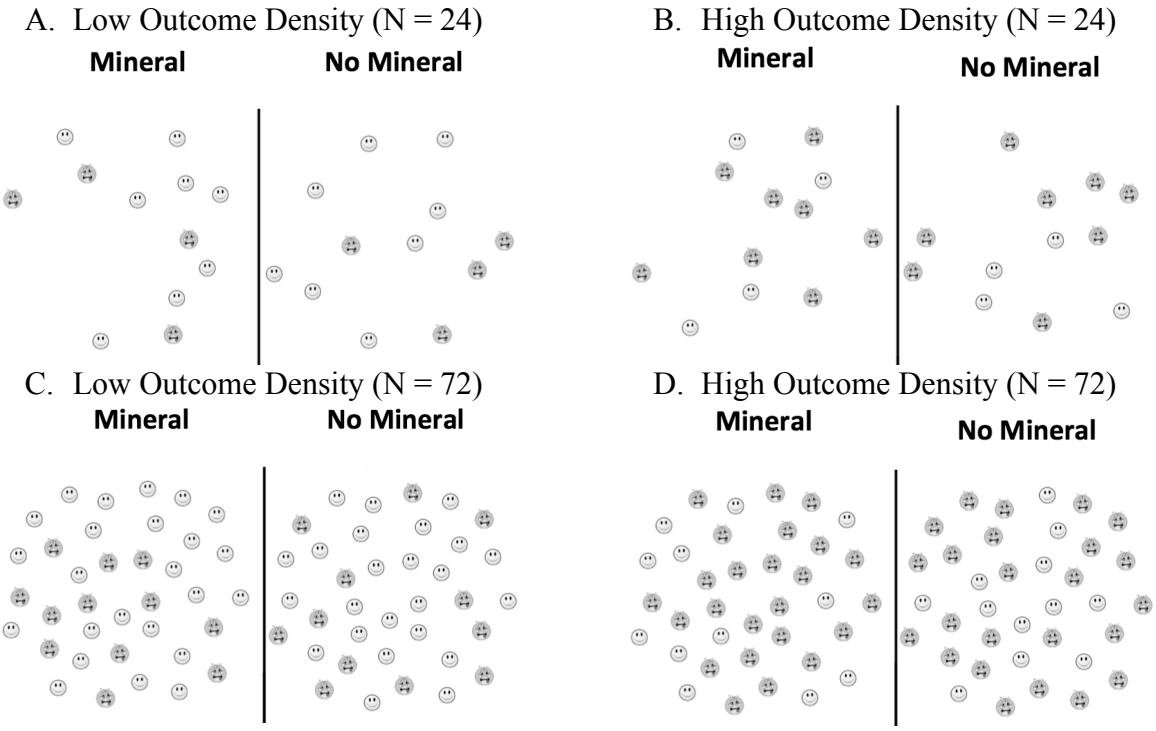


Figure 9. Scrambled presentation format. Half of the sample received the mineral (left side of the image) and half of the sample did not (right side).

The scrambled stimuli were created using a standardized procedure. To randomize the placement of emoticons on both sides of the image, I segmented a pie chart into nine slices. Each slice contained four placeholders for a total of 36 placeholders with a unique number (see Figure

10). I randomly assigned (www.random.org) each sick or healthy emoticon to a placeholder and repeated this procedure so that both sides of the image depicted a different scrambled organization of sick and healthy emoticons.

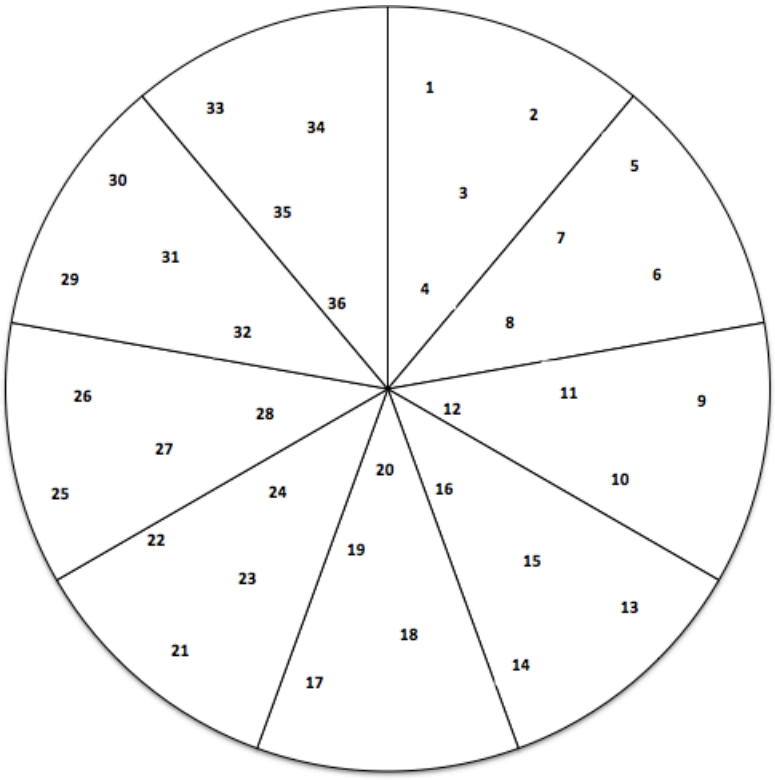


Figure 10. Pie chart with placeholders used to generate scrambled faces.

Procedure

Participants completed the experiment on a computer using E-Prime 2.0 software. First, participants read the cover story (see Appendix A) and answered the two comprehension check questions regarding random assignment and the independence of alternative causes. Next, participants reviewed information about each mineral for as long as they wanted before forwarding to the next screen, on which they made their causal judgment regarding that mineral. For each mineral, they made a causal judgment about the extent to which it influenced headaches

on a scale of -100 (*strong influence on preventing headaches*) to +100 (*strong influence on producing headaches*), where a value of 0 meant that the mineral had no influence on headaches.

Participants then answered a series of questions depending on their responses to the causal judgment question (see Figure 11). If participants made a judgment of 0, they then made a judgment about their confidence in the results from the lab on a scale of 0 (*not at all confident*) to 10 (*extremely confident*) and proceeded to the next trial (from Liljeholm & Cheng, 2009).

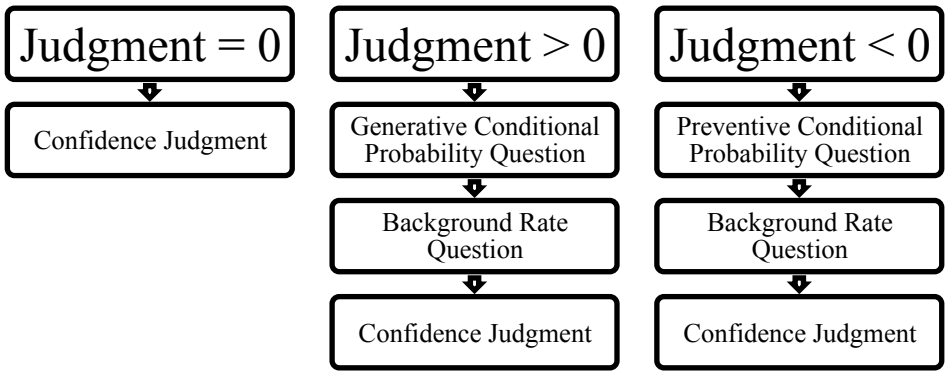


Figure 11. The sequence of questions for each mineral depended on whether the participant made a causal judgment of zero (non-contingent relationship), greater than zero (generative relationship), or less than zero (preventive relationship).

If participants made a positive judgment (the mineral produced headaches), they then answered a conditional probability question about the mineral's generative strength: "Suppose that Mineral X is given to 100 people who are not currently suffering from a headache. How many of the 100 would develop a headache?" Next, they evaluated the background rate of headaches: "Imagine a group of 100 people who have not been given the mineral. How many of the 100 would have a headache?" Finally, they made a confidence judgment and proceeded to the next trial. If participants made a negative judgment (the mineral prevented headaches), they answered a conditional probability question about the mineral's preventive strength: "Suppose that Mineral X is given to 100 people who are currently suffering from a headache. How many of

the 100 would no longer have a headache?" Next, they answered the background rate question, made a confidence judgment, and proceeded to the next trial.

Participants viewed the four scrambled and four organized minerals in two uninterrupted blocks and were randomly assigned to first review either the block of scrambled (n = 78) or organized (n = 83) stimuli. I randomized the four minerals within each block to prevent possible order effects of the outcome density or sample size conditions.

A number of participants spontaneously indicated that they noticed the left and right sides of the stimuli were exactly the same. When doubling the sample size, I included an additional open-ended question after participants evaluated all eight minerals: "Did you notice anything about the experiment?" Finally, participants answered demographic questions about their age, gender, and year in school.

Dependent Measures

Causal judgments. The primary variable of interest was the causal judgment that participants made on a scale from -100 to +100. From these judgments, I calculated difference scores between causal judgments for the high and low outcome density conditions. The difference score reflects the magnitude of the outcome density effect for each participant within each condition. Positive values mean that causal judgments were greater for the high outcome density condition, indicating an outcome density effect. Negative values mean that causal judgments were greater for the low outcome density condition, indicating the opposite of an outcome density effect.

Causal power. I calculated causal power (Cheng, 1997) using participants' responses to the generative or preventive conditional probability question and the background rate question. If

participants gave a positive causal judgment, I used the generative equation (see Equation 5) to calculate causal strength:

$$\frac{(\#\text{ of headaches in 100 people given the mineral}) - (\#\text{ of headaches in 100 people not given the mineral})}{100 - (\#\text{ of headaches in 100 not given the mineral})} \quad (5)$$

If participants gave a negative causal judgment, I used the preventive equation (see Equation 6) to calculate causal strength:

$$\frac{(\#\text{ of headaches in 100 people given the mineral}) - (\#\text{ of headaches in 100 people not given the mineral})}{\#\text{ of headaches in 100 not given the mineral}} \quad (6)$$

If participants gave a causal judgment of zero, their causal power was zero. As with causal judgments, the magnitude of the outcome density effect was the difference in mean causal power estimates for the high and low outcome density conditions.

Confidence judgments and response time data. Additional variables of interest included confidence judgments, the amount of time participants reviewed each stimulus, and the amount of time spent answering questions (i.e., causal judgments, generative power, preventive power). This data was not transformed into difference scores, as there are no predetermined normative standards for this information.

Results

Data Analyses

I conducted all data analyses using R (Version 3.3.2). Primary data analyses of causal judgments and causal power estimates included all participants ($N = 161$). Causal power analyses excluded 17 cases, in which causal power could not be calculated: some participants made off-scale responses (over 100) to the background rate question ($n = 1$); for some, the denominator of the causal power equation was equal to zero and causal power was thus undefined ($n = 6$); and for others, estimates could not be calculated due to missing responses from the background rate ($n = 9$), preventive power ($n = 1$), or generative power ($n = 1$) questions. Secondary analyses used a subset of the data based on participants' responses to the comprehension check questions. I excluded participants from these analyses if they did not answer the random assignment ($n = 6$) or independence of alternative causes ($n = 2$) question.

Median causal judgments were equal to zero in each condition. Thus, to determine the effect of outcome density, I conducted a Wilcoxon-ranked sum test for each condition (24-organized, 72-organized, 24-scrambled, 72-scrambled). To determine the effect of cognitive demand, all dependent variables were analyzed with separate linear mixed models, using participants' intercepts as the sole random factor and modelling as fixed factors the full 2 (sample size: small, large) x 2 (presentation format: organized, scrambled) x 2 (presentation format order: organized first, scrambled first) factorial. Denominator degrees of freedom for the fixed effects are Kenward-Rogers degrees of freedom. Measures of effect sizes for the outcome density effect and cognitive demand factors are based on pairwise comparisons of interest and are reported as Cohen's *d*.

All Participants

Causal judgments. As can be seen in Table 6, on average participants demonstrated an outcome density effect across all conditions, giving significantly higher ratings in the high versus low outcome density conditions (see also Figure 12). Seventy (43.48%) of the 161 participants gave causal judgments of zero for all conditions, the expected normative causal judgment.

Table 6

Mean Causal Judgments for All Participants.

Design	Sample Size	Outcome Density	*M*	*Mdiff*	*Wilcoxon Rank Sum*
Organized	24	Low	-5.15 *(34.57)*	13.14 *(55.72)*	z = -2.87 p = .002
		High	7.99 *(35.42)*	d = 0.24	
	72	Low	-4.01 *(34.45)*	13.20 *(51.04)*	z = 3.59 p < .001
		High	9.18 *(33.05)*	d = 0.26	
Scrambled	24	Low	-6.31 *(32.01)*	19.04 *(50.61)*	z = 4.53 p < .001
		High	12.73 *(36.60)*	d = 0.38	
	72	Low	-4.50 *(37.27)*	18.89 *(56.39)*	z = 4.19 p < .001
		High	14.39 *(39.52)*	d = 0.33	

Note. Causal judgments were made on a scale from -100 to +100. SDs in parentheses.

Because median causal judgments were equal to zero across all conditions, I conducted four Wilcoxon-rank sum tests to non-parametrically test judgments for high versus low outcome density minerals (see Table 6). Results revealed significant differences in ranked values for the high and low outcome density minerals in all conditions[2]. As predicted, causal judgments were greater for high than low outcome density minerals, suggesting a pervasive outcome density effect in each condition despite median causal judgments being equal to zero.

[2] Parametric one-sample t-tests also revealed that mean differences in causal judgments were significantly different from zero across all conditions.

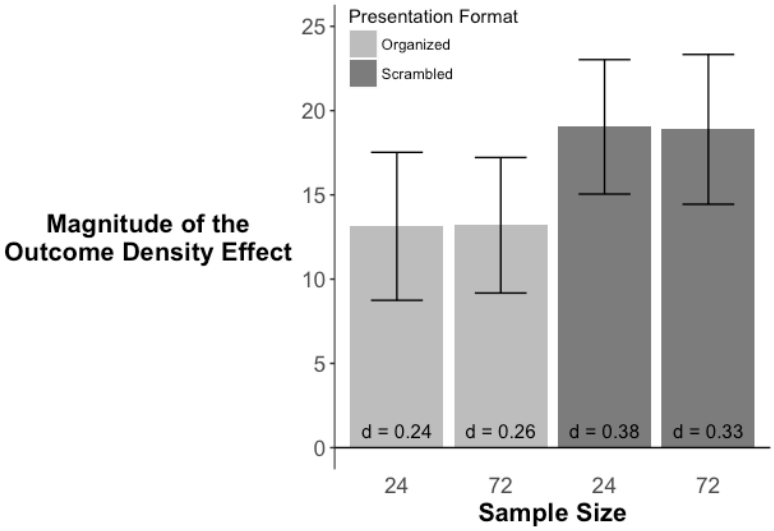


Figure 12. The magnitude of the outcome density effect represents the mean differences in causal judgments between high and low outcome density minerals for each of the four conditions. Error bars indicate standard error and *d* indicates Cohen's *dz* (i.e., size of the outcome density effect).

Effect of cognitive demand on causal judgments. As shown in Figure 13, there was an outcome density effect in each condition such that mean differences in causal judgments were positive (i.e., causal judgments were greater for high than low outcome density minerals). Contrary to my predictions, however, the magnitude of this effect was independent of cognitive demand manipulations.

The magnitude of the outcome density effect was slightly greater in the scrambled condition (M = 18.96, SD = 53.50) than in the organized condition (M = 13.17, SD = 53.35). However, the effect of presentation format was only marginally significant, $F(1, 477) = 3.56$, p = .060, d = 0.11. Thus, increased cognitive load as a function of presentation format does not appear to impact the outcome density effect. There were no effects of sample size, $F(1, 477) <$ 1.0, p = .979, dz < .01, order of presentation format, $F(1, 159) = 0.11$, p = .737, dz = 0.04, nor significant interactions (all p's > .213). Because the magnitude of the outcome density effect was

independent of cognitive load manipulations, individuals do not appear to use outcome density as a heuristic when making causal judgments during cognitively demanding tasks.

Causal power. Mean causal power estimates were close to zero (i.e., a normative causal power estimate) in each condition, as seen in Table 7. Of the 146 participants included in analyses, 75 (51.37%) had causal power estimates equal to zero for all eight minerals either because all initial causal judgments were equal to zero (n = 70) or all calculated causal power judgments were equal to zero (n = 5). As such, median causal power estimates were equal to zero for each condition.

Table 7

Mean Causal Power Estimates

Design	Sample Size	Outcome Density	*M*	*Mdiff*	*Wilcoxon Rank Sum*
Organized	24	Low	0.05 *(0.27)*	-0.03 *(0.34)*	z = -0.54 p = .293
		High	0.02 *(0.22)*	d = 0.09	
	72	Low	0.06 *(0.26)*	-0.04 *(0.32)*	z = -1.27 p = .103
		High	0.02 *(0.24)*	d = 0.13	
Scrambled	24	Low	0.05 *(0.26)*	-0.04 *(0.32)*	z = -2.29 p = .011
		High	0.01 *(0.24)*	d = 0.15	
	72	Low	0.05 *(0.29)*	0.01 *(0.33)*	z = 1.42 p = .922
		High	0.06 *(0.24)*	d = 0.03	

Note. SDs in parentheses.

Because mean causal power estimates were close to zero, there did not initially appear to be an effect of outcome density on causal power estimates in any condition. However, a Wilcoxon-rank sum test revealed that rankings of mean causal power estimates were significantly different between the high and low outcome density conditions when participants

responded to questions in the scrambled-small sample condition, $z = -2.29$, $p = .011$[34] (see Figure 13). There was an unexpected reversal of an outcome density effect, such that causal power estimates were greater for the low than the high outcome density mineral in the scrambled-small sample condition.

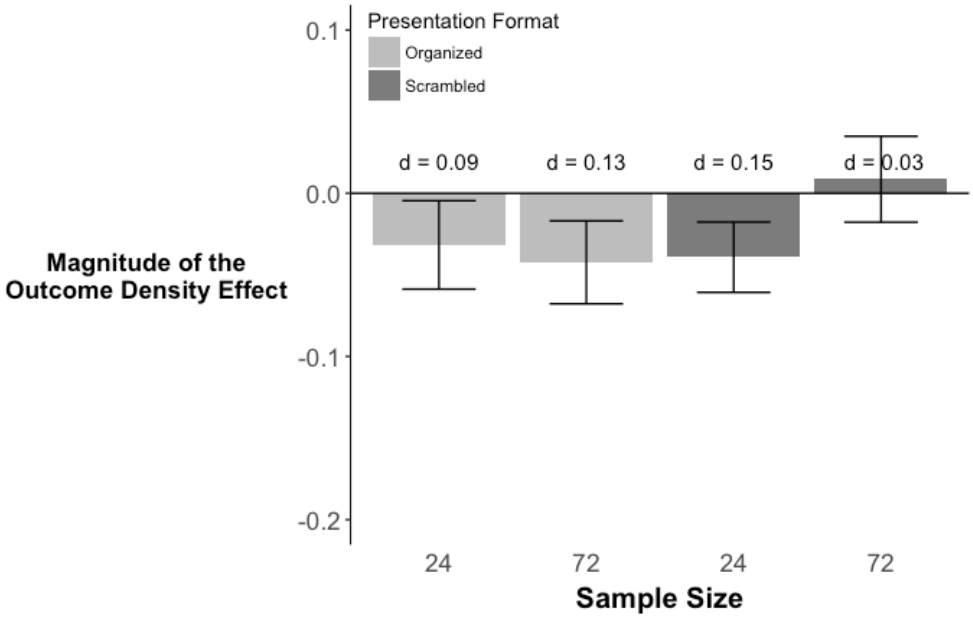


Figure 13. The magnitude of the outcome density effect represents the mean differences in causal power estimates between high and low outcome density minerals for each of the four conditions. Error bars indicate standard error and d indicates Cohen's dz (i.e., size of the outcome density effect).

I predicted an effect of outcome density in each condition, such that causal power estimates would be greater in the high than the low outcome density minerals within each condition. Instead, causal power estimates appeared relatively close to normative estimates of zero in the scrambled-large sample, organized-small sample, and organized-large sample

[3] Because of tied rank values, exact p-values could not be determined. Parametric one-sample t-tests did not show mean differences in causal judgments to be significantly different from zero in the scrambled-small sample condition.

[4] Parametric one-sample t-tests supported this interpretation, as mean differences in causal power estimates were not significantly different from zero in the scrambled-large sample, organized-small sample, and organized-large sample conditions.

conditions[5]. Although individuals demonstrated significant outcome density effects for causal judgments, this pattern was not replicated when evaluating estimates of causal power.

Effect of cognitive demand. As can be seen in Figure 14, participants demonstrated a difference between the high and low outcome density conditions, but in a direction opposite that of an outcome density effect when they first viewed the organized block than if participants first viewed the scrambled block of minerals. In contrast, there was no effect of outcome density for participants who first reviewed the scrambled condition. This interpretation was supported by a main effect of the order of presentation format on mean differences in causal power estimates, $F(1, 157.71) = 8.46$, $p = .004$, $d = 0.25$. While the direction of the effect in the organized-first condition was unanticipated, it does not lend support to the theory that outcome density could be used as a heuristic in cognitively demanding scenarios.

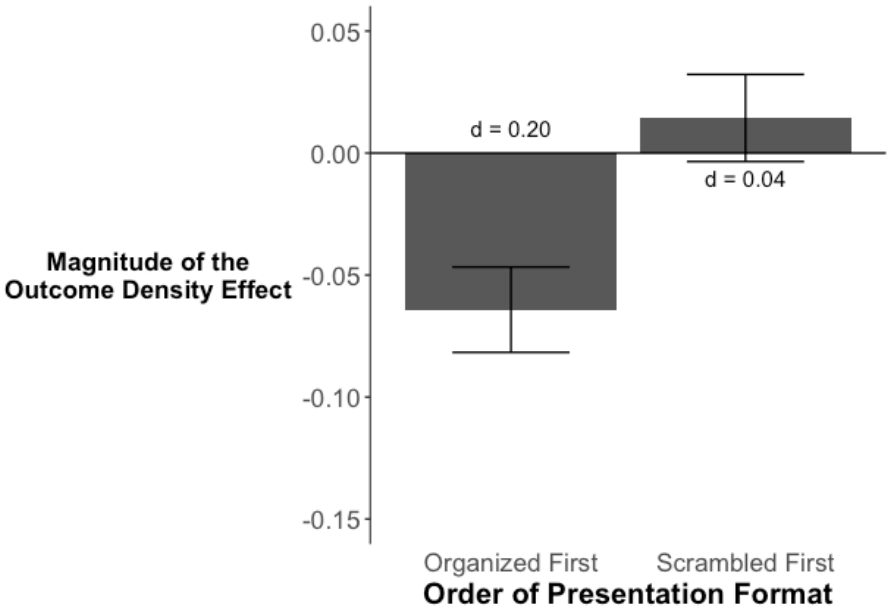


Figure 14. The magnitude of the outcome density effect represents the mean differences in causal power estimates between high and low outcome density minerals for each of the four conditions. Error bars indicate standard error and d indicates Cohen's dz (i.e., size of the outcome density effect).

[5] Parametric one-sample t-tests supported this interpretation, as mean differences in causal power estimates were not significantly different from zero in the scrambled-large sample, organized-small sample, and organized-large sample conditions.

Neither presentation format, $F(1, 467.02) = 0.72$, $p = .398$, $dz = .07$, nor sample size $F(1, 465.65) = 0.56$, $p = .454$, $dz = .06$, had a significant effect on the magnitude of the outcome density effect. Additionally, there were no significant interactions (all p's > 0.226). As with causal judgments, individuals do not use outcome density as a heuristic to inform estimates of causal power in cognitively demanding tasks.

Subset Analyses Based on Comprehension Check Responses

I conducted separate analyses based on participants' responses to the comprehension check questions. Of the 153 participants who responded to both comprehension check questions, 33 answered both questions incorrectly, 72 answered one question correctly, and 48 answered both questions correctly (see Table 8).

Table 8

Frequencies of Responses to Comprehension Check Questions

Category	Response	*N*
Independence of Alternative Causes	No (correct)	99
	Yes (incorrect)	60
Random Assignment	Same (correct)	73
	Greater Than (incorrect)	20
	Less Than (incorrect)	62
Overall Comprehension Check	None Correct	33
	One Correct	72
	Both Correct	48

Causal judgments. The effect of outcome density on causal judgments depended on how participants responded to the comprehension check questions (see Table 9). Within each group of participants, median causal judgments were equal to zero in every condition, as many participants gave causal judgments of zero for each mineral (both correct*:* $n = 20$ (41.67%), one correct: $n = 35$ (48.61%), none correct: $n = 12$ (36.36%).

Participants who answered one or both comprehension check questions correctly demonstrated an outcome density effect. Mean differences in causal judgments were positive, such that causal judgments were greater for the high than low outcome density conditions. However, participants who answered neither question correctly appeared insensitive to the outcome density manipulations. These results were supported by Wilcoxon-rank sum tests, which only showed significant differences in ranked causal judgments between the high and low outcome density conditions for the one correct and both correct groups[6] (see Table 10).

Table 9

Mean Causal Judgments Based on Comprehension Check Responses

			Both Correct ($N = 48$)		One Correct ($N = 72$)		None Correct ($N = 33$)	
Design	Sample Size	Outcome Density	*M*	*Mdiff*	*M*	*Mdiff*	*M*	*Mdiff*
Organized	24	Low	-6.67 *(37.92)*	18.13 *(57.79)* $d = 0.31$	-8.46 *(30.61)*	15.28 *(54.37)* $d = 0.28$	5.15 *(38.85)*	2.27 *(61.26)* $d = 0.04$
		High	11.46 *(35.70)*		6.82 *(34.56)*		7.42 *(41.43)*	
	72	Low	-5.00 *(38.25)*	16.88 *(52.66)* $d = 0.32$	-4.75 *(33.49)*	11.88 *(48.71)* $d = 0.24$	-1.97 *(35.84)*	12.12 *(59.62)* $d = 0.20$
		High	11.88 *(33.87)*		7.13 *(30.43)*		10.15 *(40.01)*	
Scrambled	24	Low	-10.23 *(35.13)*	29.06 *(51.57)* $d = 0.56$	-6.94 *(24.74)*	18.04 *(44.07)* $d = 0.41$	1.06 *(42.11)*	6.09 *(62.06)* $d = 0.10$
		High	18.83 *(39.13)*		11.10 *(32.63)*		7.15 *(42.98)*	
	72	Low	-3.77 *(40.53)*	21.73 *(61.67)* $d = 0.35$	-4.75 *(33.49)*	20.13 *(44.78)* $d = 0.45$	-11.55 *(44.54)*	16.79 *(71.87)* $d = 0.23$
		High	17.96 *(37.94)*		16.13 *(33.98)*		5.24 *(53.15)*	

Note. Causal judgments were made on a scale from -100 to +100. SDs in parentheses.

[6] These results were also supported by parametric t-tests. Mean differences in causal judgments were significantly different from zero in the one correct and both correct groups. This trend did not reach significance in the none-correct group.

Table 10

Causal Judgment Wilcoxon-Rank Sum Test Statistics by Comprehension Check Responses.

Design	Sample Size	Both Correct ($N = 48$)	One Correct ($N = 72$)	None Correct ($N = 33$)
Organized	24	$z = 2.03$	$z = 2.06$	$z = 1.15$
		$p = .021$	$p = .020$	$p = .875$
	72	$z = 2.69$	$z = 1.93$	$z = 0.81$
		$p = .004$	$p = .027$	$p = .208$
Scrambled	24	$z = 3.19$	$z = 3.13$	$z = 0.14$
		$p = .001$	$p = .001$	$p = .443$
	72	$z = 2.21$	$z = 3.54$	$z = 1.01$
		$p = .013$	$p < .001$	$p = .156$

Note. Causal judgments were made on a scale from -100 to +100. SDs in parentheses.

Contrary to my predictions, participants who demonstrated an understanding of experimental design by answering both questions correctly were still susceptible to the outcome density effect. This suggests that the outcome density effect is pervasive and independent of understanding random assignment or the independence of alternative causes. I hypothesized that if participants had an incomplete understanding of experimental design (i.e., answered one or neither question correct), they would be more sensitive to changes in outcome density. While there was an effect of outcome density in the one correct group, this was not replicated in the none-correct group.

Effect of cognitive demand. Cognitive demand manipulations only affected the influence of outcome density on causal judgments in the group of participants who correctly answered one of the comprehension check questions ($n = 72$). As can be seen in Table 11, there was a significant three-way interaction between sample size, presentation format, and the order of presentation format for the one-correct group, $F(1, 210) = 5.79$, $p = .017$.

Table 11

Magnitude of the Outcome Density Effect on Causal Judgments for the One-Correct Group

Order of Presentation Format	Sample Size	M_{diff}	SD_{diff}
Organized First			
Organized	24	0.67	47.78
	72	5.77	48.49
Scrambled	24_a	20.44	42.01
	72_a	13.74	40.14
Scrambled First			
Organized	24	32.55	57.28
	72	19.09	48.71
Scrambled	24	15.21	46.87
	72	27.67	49.28

Note. Mean differences in causal judgments between high and low outcome density conditions for the group of participants (n = 72) who correctly answered one comprehension check question. Subscript indicates a significant difference in the magnitude of the outcome density effect via Tukey's honestly significant difference test.

As depicted in Figure 15, this three-way interaction is driven by differences in causal judgments for the scrambled condition if participants first saw the organized block of minerals. In the organized-first group, the magnitude of the outcome density effect for the scrambled condition (i.e., the second block they reviewed) was greater for the small sample size than the large sample size. A Tukey's HSD (honest significant difference) test revealed this difference to be significant, $t(209.99) = 2.64$, $p = .044$, $d = 0.16$. There were no other significant pairwise comparisons in the organized-first group (all ps > .208) or the scrambled-first group (all ps > .148).

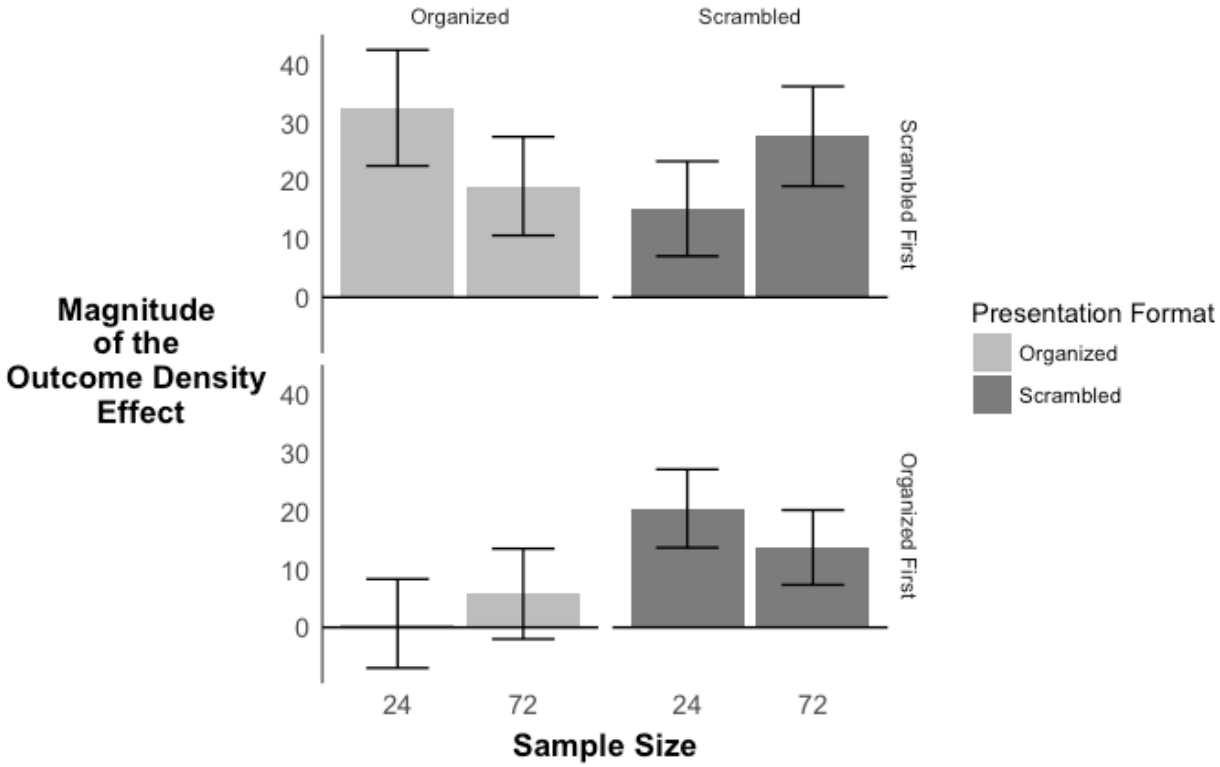


Figure 15. The magnitude of the outcome density effect for causal judgments (mean differences between high and low outcome density conditions) in the one-correct group. Error bars indicate standard error.

Because these findings do not provide a straightforward interpretation of how cognitive demand affects the magnitude of the outcome density effect, they do not support the use of outcome density as a heuristic when individuals have an incomplete understanding of experimental design (i.e., answered only one comprehension check question correctly).

For participants answering both questions correct (n = 48) and those answering neither question correct (n = 33), there were no effects of the cognitive demand manipulations (all ps > .118). Together, these findings suggest that the magnitude of the outcome density effect is independent of cognitive demand manipulations even if participants have an incomplete understanding of experimental design.

Causal power. For each group of participants, mean causal power estimates were close to zero in each condition (see Table 12). Of the 161 participants, many in each group had causal

power estimates equal to zero for each mineral [both correct: 21 (47.72%), one correct: 39 (60.00%), neither correct: 11 (36.67%)]. As such, median causal power estimates were equal to zero in each condition.

Table 12

Mean Causal Power Estimates Based on Comprehension Check Responses

			Both Correct ($N = 48$)		One Correct ($N = 72$)		None Correct ($N = 33$)	
Design	Sample Size	Outcome Density	*M*	*Mdiff*	*M*	*Mdiff*	*M*	*Mdiff*
Organized	24	Low	0.08 *(0.28)*	-0.06 *(0.31)*	0.06 *(0.27)*	-0.05 *(0.37)*	0.01 *(0.30)*	0.07 *(0.34)*
		High	0.02 *(0.18)*	$d = 0.20$	0.00 *(0.24)*	$d = 0.14$	0.08 *(0.26)*	$d = 0.20$
	72	Low	0.09 *(0.28)*	-0.07 *(0.41)*	0.03 *(0.27)*	-0.03 *(0.21)*	0.10 *(0.24)*	-0.03 *(0.39)*
		High	0.03 *(0.21)*	$d = 0.17$	-0.01 *(0.24)*	$d = 0.16$	0.07 *(0.28)*	$d = 0.07$
Scrambled	24	Low	0.09 *(0.27)*	-0.03 *(0.30)*	0.04 *(0.24)*	-0.06 *(0.24)*	0.04 *(0.31)*	-0.03 *(0.34)*
		High	0.06 *(0.25)*	$d = 0.10$	-0.01 *(0.23)*	$d = 0.23$	0.01 *(0.28)*	$d = 0.08$
	72	Low	0.09 *(0.31)*	-0.02 *(0.45)*	0.00 *(0.27)*	0.04 *(0.25)*	0.08 *(0.31)*	-0.02 *(0.30)*
		High	0.08 *(0.23)*	$d = 0.03$	0.04 *(0.20)*	$d = 0.16$	0.08 *(0.35)*	$d = 0.06$

Note. Causal power estimates were calculated using the generative (see Equation 5) and preventive (see Equation 6) formulas. SDs in parentheses.

Outcome density had no effect on estimates of causal power within the three groups of participants. This is consistent with the whole-group analyses, in which there was no significant effect of outcome density, and was confirmed by Wilcoxon-rank sum tests (see Table 13)[7].

[7] Because of tied rank and/or zero values, exact p-values could not be determined. These results were supported by parametric one-sample t-tests, which did not reveal mean differences between causal power estimates for high and low outcome density conditions to be significantly different from zero.

Table 13

Causal Power Estimate Wilcoxon-Rank Sum Test Statistics by Comprehension Check Responses

Design	Sample Size	Both Correct ($N = 48$)	One Correct ($N = 72$)	None Correct ($N = 33$)
Organized	24	$z = 1.12$	$z = 0.50$	$z = 0.39$
		$p = .132$	$p = .309$	$p = .350$
	72	$z = 0.75$	$z = 0.48$	$z = 0.99$
		$p = .227$	$p = .315$	$p = .839$
Scrambled	24	$z = 0.78$	$z = 1.54$	$z = 0.23$
		$p = .218$	$p = .061$	$p = .410$
	72	$z = 0.17$	$z = 1.54$	$z = 0.82$
		$p = .433$	$p = .224$	$p = .794$

Effect of cognitive demand. As I found for the entire sample, manipulations of sample size and presentation format had no effect on the magnitude of the outcome density effect in any of the three groups. For participants answering both questions correct ($n = 48$), participants answering one question correct ($n = 72$), and those answering neither question correct ($n = 33$), there were no effects of presentation format or sample size (all ps > .088). Thus, causal power estimates are not sensitive to manipulations of cognitive demand even when participants have an incomplete understanding of experimental design.

I did observe a non-significant, but marginal effect of counterbalancing in the one-correct group: outcome density had a greater effect on causal judgments if participants first saw the organized block ($M_{diff} = -.06$, $SD_{diff} = .27$) than for participants who first saw the scrambled block ($M_{diff} = .01, .27$), $F(1, 69.69) = 3.73$, $p = .058$, $dz = 0.31$, for the main effect of counterbalancing order. This pattern fits the findings for the entire sample, but was not replicated in the both-correct or none-correct groups (ps > .098).

Conditional Probability Questions

To determine if participants accurately tracked the frequencies of headaches, I analyzed participants' responses to conditional probability questions regarding the background rate (how

many headaches in a group of 100 people who did not receive the mineral), generative power (how many headaches in a group of 100 people who did receive the mineral), and preventive power (how many would no longer have headaches in a group of 100 people who received the mineral). Participants only responded to the generative or preventive power questions if they made a positive or negative causal judgment for a mineral, respectively.

For the generative power and background rate questions, normative frequency estimates would be 33 in the low outcome density condition and 67 in the high outcome density condition. For the preventive power question, normative frequency estimates would be 67 in the low outcome density condition and 33 in the high outcome density condition. Table 14 depicts the average frequency estimates across all conditions. To assess the effects of outcome density, sample size, presentation format, and the order of presentation format (i.e., counterbalancing order), I conducted separate 2x2x2x2 ANOVA's on raw responses to each of the questions.

Table 14

Mean Frequency Estimates for Conditional Probability Questions

			Background Rate		Generative		Preventive	
Design	Sample Size	Outcome Density	*N*	*M* *(SD)*	*N*	*M* *(SD)*	*N*	*M* *(SD)*
Organized	24	Low	159	36.07 *(16.11)*	18	36.44 *(23.00)*	35	58.51 *(22.85)*
	24	High	160	53.34 *(19.82)*	37	59.62 *(18.11)*	12	38.50 *(25.81)*
	72	Low	159	37.33 *(17.05)*	21	36.14 *(24.93)*	35	61.29 *(18.86)*
	72	High	161	55.85 *(19.37)*	43	53.72 *(21.48)*	8	46.88 *(23.01)*
Scrambled	24	Low	159	39.02 *(20.13)*	15	42.40 *(24.09)*	32	63.56 *(20.66)*
	24	High	160	53.84 *(22.32)*	41	60.00 *(19.07)*	11	37.36 *(25.52)*
	72	Low	160	38.41 *(20.68)*	23	47.17 *(29.50)*	39	55.03 *(19.59)*
	72	High	160	53.78 *(19.67)*	51	61.76 *(20.77)*	10	51.50 *(28.19)*

Background rate. As seen in Table 15, the effect of outcome density on frequency estimates about the background rate of headaches was modulated by an interaction with the order of presentation format, $F(1, 1104.19) = 22.07$, $p < .001$. Frequency estimates were significantly greater for the high than the low outcome density minerals in both the organized-first, $d = 1.06$, and scrambled-first, $d = 0.63$, conditions, supported by a Tukey's HSD test.

Table 15

Mean Background Rate Frequency Estimate by Outcome Density and Counterbalancing Order

Order of Presentation Format	Outcome Density	*M*	*SD*
Organized First	Low	33.63	17.23
	High$_a$	54.61	22.05
Scrambled First	Low	42.00	19.03
	High$_a$	53.78	18.30

Note. Subscript indicates non-significant differences via Tukey's HSD test.

Pairwise comparisons also revealed that frequency estimates for low outcome density minerals were significantly greater in the scrambled-first than the organized-first group, $d = 0.46$. The initially scrambled information may have given the appearance of more headaches in the low outcome density condition, which carried over to the organized block of minerals. There was no significant difference in responses to the background rate question for the high outcome density minerals between the organized-first and scrambled-first group, $d = 0.04$**.** The 2x2x2x2 ANOVA did not reveal other main effects or interactions to be significant (all *p*s > .419). Overall, these results suggest that participants processed the background rate of the minerals and made responses reflecting the difference between the low and high outcome density conditions.

Generative power. As expected, frequency estimates about generative power (see Table 14) were greater in the high ($M = 58.87$, $SD = 20.08$) than the low ($M = 40.83$, $SD = 25.77$) outcome density conditions, $F(1, 226.30) = 45.21$, $p < .001$, $d = 0.79$. There was also a significant main effect of presentation format, such that frequency estimates were greater in the

scrambled (M = 56.39, SD = 23.34) than the organized (M = 49.84, SD = 23.29) conditions, F(1, 195.19) = 4.99, p = .027, d = 0.28. This provides additional support for the idea that the scrambled design gave an impression of more headaches.

Additionally, there was a significant interaction between sample size and the order of presentation format, F(1, 197.47) = 4.38, p = .038. However, Tukey's HSD tests did not reveal comparisons between any of the conditions to be significantly different from each other (all ps > .097). There were no other significant main effects or interactions (all ps > .133).

Preventive power. Frequency estimates for the preventive power questions (see Table 14) were greater in the low (M = 59.38, SD = 20.55) than the high (M = 43.00, SD = 25.56) outcome density conditions. A 2x2x2x2 ANOVA revealed the main effect of outcome density to be significant, F(1, 160.69) = 11.33, p = .001, d = .706. There were no additional significant main effects or interactions (all ps > .072).

Reaction Time Data

To evaluate the manipulation of cognitive load, I assessed the amount of time participants spent reviewing stimuli prior to making a causal judgment and the amount of time participants spent making a causal judgment. I conducted all analyses using the natural log of the reaction time data. If participants spent more time spent reviewing data and making causal judgments in the scrambled and/or large sample minerals, these conditions may be more cognitively demanding than the organized and/or small sample conditions. Therefore, longer reaction times in the scrambled and large sample conditions would support the presentation format and sample size conditions as successful manipulations of cognitive load.

Time spent reviewing stimuli. The amount of time participants spent reviewing stimuli depended on sample size, presentation format, the order of presentation format, and outcome

density (see Table 16). The analysis of log reaction time data revealed two significant interactions: an interaction between presentation format and sample size, $F(1, 1113) = 7.93$, $p = .005$, and an interaction between presentation format and the order of presentation format, $F(1, 1113) = 164.53$, $p < .001$.

Table 16

Mean Time (milliseconds) Spent Reviewing Stimuli

Design	Sample Size	Outcome Density	*M* (ms)	*SD*
Organized	24	Low	9.35	0.80
		High	9.41	0.87
	72	Low	9.55	0.90
		High	9.67	0.85
Scrambled	24	Low	9.61	0.72
		High	9.75	0.83
	72	Low	10.07	0.75
		High	10.16	0.91

Note. Table includes the natural log of raw reaction time data.

As shown in Table 17, participants spent more time looking at the larger than the smaller sample sizes. Tukey's HSD tests revealed that the difference in reaction times between small and large sample sizes was greater in the scrambled, $d = 0.55$, than the organized, $d = 0.27$, conditions. All remaining pairwise comparisons were significant with the exception of the average time spent reviewing the organized-large sample and scrambled-small sample conditions. Overall, these results suggest that presentation format and sample size conditions successfully manipulated cognitive load.

Table 17

Mean Time Spent Reviewing Mineral Data by Presentation Format and Sample Size

Presentation Format	Sample Size	*M* (ms)	*SD* (ms)
Organized	24	9.38	0.83
	72_a	9.61	0.88
Scrambled	24_a	9.68	0.78
	72	10.12	0.83

Note. Analyses used the natural log of raw reaction time data. Subscript indicates non-significant differences in confidence judgments via Tukey's HSD test.

The effect of presentation format was also tempered by the order of presentation format (see Table 18). Tukey's HSD tests revealed that the organized-first participants spent longer looking at organized stimuli than the scrambled-first group, $d = 0.67$, and the scrambled-first participants spent longer looking at scrambled stimuli than the organized-first group, $d = 0.49$ (ps < .001).

Table 18

Mean Time Spent Reviewing Mineral Data by Presentation Format and Counterbalancing Order

Counterbalancing Order	Presentation Format	*M* (ms)	*SD* (ms)
Organized-First	Organized$_a$	9.76	0.75
	Scrambled$_a$	9.70	0.82
Scrambled-First	Organized	9.21	0.89
	Scrambled	10.10	0.80

Note. Analyses used the natural log of raw reaction time data. Subscript indicates non-significant differences in confidence judgments via Tukey's HSD test.

In the organized-first group, however, there was no difference between the amount of time looking at the scrambled and organized conditions, $d = 0.08$, $p = .669$. All other pairwise comparisons were significantly different ($p < .001$). It is possible that when participants first reviewed the organized (i.e., less demanding) stimuli, subsequent scrambled stimuli were easier to interpret.

Finally, participants spent more time reviewing stimuli in the high outcome density ($M = 9.75$, $SD = 0.91$) than the low outcome density ($M = 9.64$, $SD = 0.84$) conditions, although the size of this effect was small, $F(1, 1113) = 12.64$, $p < .001$, $d = 0.12$. In addition to presentation format and sample size, this suggests that outcome density may have also affected cognitive load.

Time spent making causal judgments. Cognitive demand manipulations also affected the amount of time participants spent making causal judgments (see Table 19).

Table 19

Mean Time Spent Making Causal Judgments (in milliseconds)

Design	Sample Size	Outcome Density	*M*	*SD*
Organized	24	Low	8.89	0.80
		High	8.86	0.83
	72	Low	8.92	0.84
		High	8.95	0.88
Scrambled	24	Low	8.84	0.84
		High	8.79	0.84
	72	Low	8.93	0.89
		High	8.86	0.85

Note. Table includes the natural log of raw reaction time data.

As can be seen in Table 20, there was a significant three-way interaction between sample size, presentation format, and the order of presentation format, $F(1, 1113) = 4.49$, $p = .034$. In both the organized-first and scrambled-first groups, participants spent more time making causal judgments in the first block of minerals. There was no effect of sample size on time spent making causal judgments in the organized-first group, as revealed by Tukey's HSD tests. In the scrambled-first group, sample size only affected reaction time in the scrambled condition, where participants spent longer reviewing the larger sample sizes, $p = .028$, $d = 0.24$.

Table 20

Mean Time Spent Making Causal Judgments for Scrambled-First Participants

Presentation Format	Sample Size	*M* (ms)	*SD* (ms)
Organized-First			
Organized	24_b	9.16	0.80
	72_b	9.24	0.82
Scrambled	24_c	8.55	0.80
	72_c	8.52	0.75
Scrambled-First			
Organized	24_a	8.58	0.72
	72_a	8.60	0.77
Scrambled	24	9.10	0.78
	72	9.29	0.81

Note. Analyses used the natural log of raw reaction time data. Subscript indicates non-significant differences in confidence judgments via Tukey's HSD test.

Confidence Judgments

As can be seen in Table 21, participants were slightly more confident in the results from the laboratory when outcome density was low (M = 5.45, SD = 2.44) than when outcome density was high (M = 5.28, SD = 2.51), $F(1, 1102.25) = 5.64$, $p = .018$, $d = .07$, for the main effect of outcome density.

Table 21

Mean Confidence Judgments

Design	Sample Size	Outcome Density	*M*	*SD*
Organized	24	Low	5.65	2.48
		High	5.30	2.55
	72	Low	5.56	2.39
		High	5.39	2.54
Scrambled	24	Low	5.51	2.47
		High	5.29	2.52
	72	Low	5.09	2.41
		High	5.15	2.43

The analysis of confidence judgments revealed two significant interactions: an interaction between presentation format and counterbalancing order, $F(1, 1102.25) = 5.04$, $p = .025$, and an interaction between presentation format and sample size, $F(1, 1102.25) = 3.59$, $p = .058$. As seen in Table 22, the effect of format order on confidence judgments was limited to the scrambled condition. Participants in the organized-first group were less confident about scrambled results from a laboratory than participants in the scrambled-first group. A post-hoc Tukey's HSD test revealed this difference to be significant, $p = .003$, $d = 0.22$. No other pairwise comparisons were significant (all ps > .342).

Table 22

Mean Confidence Judgments by Presentation Format and Counterbalancing Order

Order of Presentation	Presentation Format	*M*	*SD*
Organized-First	Organized	5.21	2.53
	Scrambled$_a$	5.17	2.67
Scrambled-First	Organized$_a$	5.74	2.42
	Scrambled	5.36	2.20

Note. Subscript indicates significant differences in confidence judgments via Tukey's HSD test.

The effect of presentation format was also modulated by a marginal interaction with sample size. As seen in Table 23, participants were slightly less confident in data from the larger than smaller sample size conditions. A Tukey's post-hoc test revealed this difference to be driven by reduced confidence in the scrambled-large sample condition than the organized-small sample, $d = 0.14$, organized-large sample, $d = 0.15$, and scrambled-small sample, $d = 0.11$, conditions.

Table 23

Mean Confidence Judgments by Presentation Format and Sample Size

Presentation Format	Sample Size	*M*	*SD*
Organized	24$_a$	5.47	2.52
	72$_b$	5.48	2.46
Scrambled	24$_c$	5.40	2.49
	72$_{abc}$	5.12	2.42

Note. Subscript indicates significant differences in confidence judgments via Tukey's HSD test.

Discussion

Summary

In this experiment, I investigated the possible role of outcome density as a heuristic that individuals use to reason about a cause-outcome relationship in a cognitively demanding causal learning task. There were pervasive outcome density effects for causal judgments across all conditions, adding to the prevalence of the outcome density effect for non-contingent causes in the literature (e.g., Buehner et al., 2003). Unexpectedly, this pattern was not replicated for estimates of causal power.[8] Overall, the magnitude of the outcome density effect was independent of cognitive load for both causal judgments and causal power estimates. Thus, use of outcome density as a heuristic cannot explain the pervasive outcome density effects found for causal judgments.

I also assessed the magnitude of the outcome density effect with regards to participants' understanding of experimental design. Specifically, I hypothesized that individuals who correctly answered questions regarding random assignment and/or the independence of alternative causes would be less susceptible to the outcome density effect. The results did not support this hypothesis, as there were pervasive outcome density effects for causal judgments by participants who answered only one question correctly ($N = 72$) *and* participants who answered both questions correctly ($N = 48$). There was no evidence of outcome density effects in the group of participants who answered neither question correctly ($N = 33$). In line with the analysis of the entire sample, outcome density had no effect on causal power estimates when analyzed by comprehension check responses.

[8] There was a reversal of the outcome density effect in the scrambled-small sample condition for causal power estimates (i.e., greater causal power estimates in the low outcome density condition). Because the effect was small ($dz = 0.15$) and only found in one condition, changes in $p(o)$ do not appear to affect causal power estimates overall.

As with the whole-group analyses, the results did not support differential reliance on outcome density as a heuristic between three groups. In the group that correctly answered one question, there seemed to be an effect of presentation format that was dependent on the order of presentation format and sample size. Because these results were not straightforward, we cannot assume that the one-correct group relied more on outcome density as a heuristic in the more cognitively demanding (i.e., scrambled) condition. In line with the whole-group analyses, there was no effect of cognitive demand manipulations on causal power estimates in any of the three groups.

Confidence and Sample Size

Although outcome density did not affect causal power estimates, there were pervasive outcome density effects for causal judgments. If the outcome density effect is not due to increased cognitive load, then why were causal judgments greater in the high outcome density conditions? Here, there was no effect of sample size on causal judgments. However, previous research suggests that causal judgments increase as the size of the sample increases (e.g., Liljeholm & Cheng, 2009). According to the statistical law of larger numbers, data becomes increasingly reliable as the number of observations increase (Van Overwalle & Van Rooy, 2001). As such, modulations in causal judgments for non-contingent relationships may be due to changes in the perceived reliability of a small or large sample (see Buehner & Cheng, 1997; Liljeholm & Cheng, 2009).

The influence of reliability on causal judgments may stem from how researchers pose causal questions (Liljeholm & Cheng, 2009). According to the conflation hypothesis (Buehner & Cheng, 1997), the wording of specific questions may lead individuals to consider both their assessment of causal strength and their belief in the reliability of the data when making a causal

judgment. In Buehner et al.'s (2003) first experiment, participants made a judgment about how strongly they thought a cause generated an outcome. The wording of this question is ambiguous, as it is unclear whether the participant is rating the strength of the cause to produce the outcome or the strength of their belief in the relationship.

In the latter interpretation, when a causal question is ambiguous, participants may conflate causal judgments with confidence in the information provided. Thus, if the overall size of the sample increases, causal judgments of a non-contingent relationship would be closer to zero. For generative and preventive causes, causal judgments would increase as sample size increases, thus increasing reliability in the strength of the cause. To test this, Buehner and Cheng (1997) asked participants to make causal strength judgments based on information from either 16 individual trials (Experiment 1) or a summary of 100 trials (Experiment 2). Because there were reduced outcome density effects in the larger sample condition, Buehner and Cheng (1997) suggested that participants were more confident that the relationship was non-contingent.

In the current experiment, the conflation hypothesis would predict reduced outcome density effects and increased confidence judgments for the larger sample conditions. However, there were no effects of sample size on the magnitude of the outcome density effect, as outcome density effects were pervasive across all conditions. Furthermore, participants were marginally less confident in the scrambled condition when the sample size was large. As there was no effect of sample size on confidence judgments in the organized condition, these findings do not support increased confidence in larger samples.

The current experiment manipulated objective sample size, but there is a possibility that it is not objective sample size that is most important, but rather virtual sample size. Liljeholm and Cheng (2009) introduce the notion of virtual sample size as the number of trials in which a cause

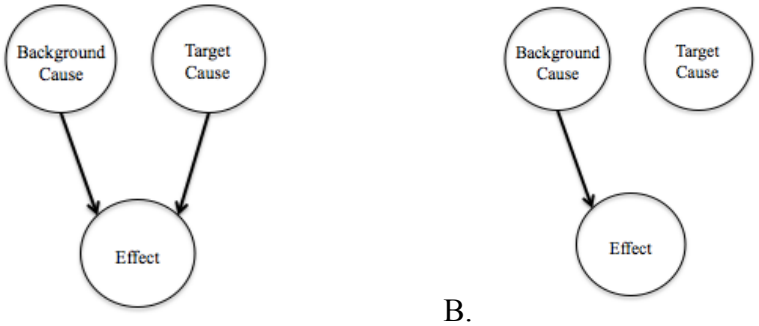
can prove its power. Recall the gardener example, in which a gardener applied Fertilizer A to 6 of 12 plants and Fertilizer B to 6 of 12 different plants. In plot A, 4 of 6 fertilized and 4 of 6 unfertilized plants grow, whereas in plot B, 2 of 6 fertilized and 2 of 6 unfertilized plants grow.

Using the information from the unfertilized plants, the gardener would expect 4 of 6 fertilized plants in plot A to grow because of alternative causes. Thus, Fertilizer A only had 2 cases (i.e., plants) in which it could prove its generative strength. In contrast, the gardener would only expect 2 of 6 unfertilized plants in plot B to grow due to alternative causes. Fertilizer B would therefore have 4 cases in which it could prove its generative strength. Both Fertilizer A and Fertilizer B are non-causal, but the gardener may be less confident in the Fertilizer A data because there were less cases in which it could prove its generative power. Thus, the gardener may be more likely to give a causal judgment of zero for Fertilizer B, because the data for Fertilizer B as a non-contingent cause is perceived as more reliable.

For contingent relationships, however, increases in virtual sample size should lead to increased generative or preventive causal judgments. For generative causes, virtual sample size is equal to the number of instances in which the outcome is absent prior to the cause. For preventive causes, virtual sample size is equal to the number of instances in which the outcome is present prior to the cause. If the cause proves its strength in more cases, this would increase reliability in its generative or preventive power, thereby increasing causal judgments. In the current experiment, however, participants simultaneously evaluated whether a given cause was preventive or generative and were not given information about the number of outcomes prior to the cause. Therefore, it is unclear what virtual sample size participants would have used on any give trial.

Ambiguity of Causal Judgment Question

Proposals of the conflation hypothesis assume that conflation of reliability and strength is due to the ambiguity of the causal judgment question (e.g., Liljeholm & Cheng, 2009). Griffiths and Tenenbaum (2005), however, dispute the influence of ambiguity. Instead, causal judgments should incorporate information about the strength of the cause *and* confidence based on the reliability of the sample. Thus, the ambiguous nature of the question should have no effect on causal judgments. Their causal support model posits that individuals base causal judgments on whether the observed data supports the target cause as present in the presence of the effect and alternative background causes (Graph A of Figure 16) or absent in the presence of the effect and alternative background causes (Graph B of Figure 16). The extent to which the data supports the cause as producing the outcome is equal to the log of the probability of the data given Graph A versus that given Graph B (see Equation 7).



A. B.

Figure 16. According to the causal support model (Griffiths & Tenenbaum, 2005), individuals make causal judgments using the probability that the evidence given Graph B outweighs the evidence of an effect given Graph A.

$$\text{Support} = \log\left(\frac{P(D|\text{Graph } 1)}{P(D|\text{Graph } 0)}\right) \qquad (7)$$

Neither the power PC model nor the ΔP rule accounts for effects of sample size, although the conflation hypothesis (e.g., Liljeholm & Cheng, 2009) offers an explanation for why sample size may affect causal power estimates. Griffiths and Tenenbaum's (2005) model of causal support considers that sample size may have multiple influences on causal judgments. When

exposed to a larger sample of data, individuals will adjust their causal judgments as they are introduced to new evidence and become more certain of these beliefs. As such, the causal support model suggests that the influence of sample size is inextricable from beliefs about a cause and an outcome.

As previously discussed, sample size did not affect causal judgments nor causal power estimates in the current experiment. Therefore, the conflation hypothesis is unable to explain possible reasons for these findings. It is possible, however, that outcome density effects for causal judgments were due to the ambiguous nature of the question. In Buehner et al.'s (2003) first experiment, participants made a causal judgment about how strongly they thought a cause prevented or caused (i.e., generated) an outcome. The authors note that the wording of this question made it unclear whether participants were making a judgment about the strength of the cause in the presence of alternative causes (thus making a contingency judgment) or making a judgment about the strength of the cause in the absence of alternative causes (thus making a causal power judgment). In a second experiment, Buehner et al. (2003) changed the wording of the question. Instead, participants were asked to estimate the number of outcomes if the putative cause was introduced to 100 cases. Overall, participants' causal power estimates were more normative in comparison to Experiment 1.

These findings are analogous to the results of the present study. There were pervasive outcome density effects when participants made a causal judgment about the extent to which a mineral influenced headaches on a scale from -100 (*the mineral has a strong influence on preventing headaches*) to +100 (*the mineral has a strong influence on producing headaches*). However, outcome density had little to no effect on participants' causal power estimates calculated from responses to frequency estimates about the generative strength, preventive

strength, or background rate of headaches for the mineral. Thus, outcome density effects were predominantly found when the causal judgment question was ambiguous.

However, the ambiguous nature of the question cannot fully explain outcome density effects for causal judgments. Participants demonstrated outcome density effects for causal judgments across all conditions. If outcome density is due to ambiguity, then it should not occur in each condition. For example, when the information is organized, the non-contingent nature of the relationship is clearer and ambiguity of the question should not have an effect.

Alternatively, individuals could be relying on a cell *A* strategy, a heuristic where causal judgments are based on the frequency of the joint presence of the cause and outcome (e.g., Schustack & Sternberg, 1981). Although it is not always used as the sole information to inform causal judgments, individuals tend to weigh cell *A* information more heavily than information from the other cells (see Figure 2) when making a causal judgment (Wasserman, Dorner, & Kao, 1990). Using the cell *A* strategy or outcome density to guide causal judgments are two distinctive strategies, as the outcome density effect assumes individuals are relying on both cell *A* and cell *C* (i.e., the probability of the outcome in the absence of the cause). The cell *A* strategy would correspond more with findings regarding the order in which individuals rely on cell information, in which cell use is ordered cell *A* > cell *B* > cell *C* > cell *D* (i.e., cell *B* is weighted more heavily than cell *C*, see Wasserman et al., 1990).

Implications and Future Directions

Our findings do not suggest the use of outcome density as a heuristic to make causal judgments about a non-contingent cause when cognitive demand is high. It is possible, however, that these findings are due to the nature of the stimuli. At the beginning of the study, several participants made unprompted comments about the nature of the stimuli. Therefore, I added a

question to the end of the study that explicitly asked participants if they noticed anything. Of the 89 participants who responded, 51 (57.30%) noted that there were the same number of headaches in the group that received the mineral and the group that did not receive the mineral. Additionally, 28 (31.46%) stated that because the results were the same for both groups, there was no relationship between the mineral and headaches.

Thus, the manipulations of cognitive demand may not have influenced the magnitude of the outcome density effect because the non-contingent relationship was clearly visible (i.e., same number of headaches in the presence and the absence of the cause). Future studies should evaluate this for non-contingent causes in which cell *A* and cell *C* are unequal, making the non-contingent relationship less discernible. Furthermore, this would allow researchers to evaluate whether these findings are due to reliance on a cell *A* strategy rather than outcome density as a whole.

Perhaps the most important finding in this experiment is that outcome density effects were pervasive for causal judgments but not causal power estimates. This may be due to the ambiguous nature of the question, although it is not certain why the outcome density effects would persist in all cognitive demand conditions.

Finally, the current experiment cannot entirely rule out outcome density as a heuristic because the current experiment solely investigated non-contingent relationships. Outcome density effects are also prevalent for generative and preventive relationships and thus, outcome density should also be investigated with regard to cognitive manipulations for causal relationships (e.g., Buehner et al., 2003).

References

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.

Allan, L. G., Siegel, S., & Tangen, J. M. (2005). A signal detection analysis of contingency data. *Learning & Behavior*, *33*(2), 250-263.

Allan, L. G., Hannah, S. D., Crump, M. J. C., & Siegel, S. (2008). The psychophysics of contingency assessment. *Journal of Experimental Psychology: General, 137*(2), 226-243.

Blanco, F., Matute, H., & Vadillo, M. A. (2013). Interactive effects of the probability of the cue and the probability of the outcome on the overestimation of null contingency. *Learning & Behavior*, *41*(4), 333-340.

Buehner, M. J., & Cheng, P. W. (1997). Causal induction: The power PC theory versus the Rescorla-Wagner model. In *Proceedings of the nineteenth annual conference of the Cognitive Science Society* (pp. 55-60). Hillsdale, NJ: Erlbaum.

Buehner, M. J., Cheng, P. W., & Clifford, D. (2003). From covariation to causation: a test of the assumption of causal power. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(6), 1119-1140.

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

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

Crump, M. C., Hannah, S. D., Allan, L. G., & Hord, L. K. (2007). Contingency judgments on the fly. *The Quarterly Journal Of Experimental Psychology*, *60*(6), 753-761. doi:10.1080/17470210701257685

Fielder, K. (2009). Pseudocontingencies: An integrative account of an intriguing cognitive illusion. *Psychological Review 116*(1), 187-206.

Finucane, M. L., Alhakami, A., Slovic, P., & Johnson, S. M. (2000). The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making*, *13*(1), 1-17.

Fleig, H., Meiser, T., Ettlin, F., & Rummel, J. (2017). Statistical numeracy as a moderator of (pseudo)contingency effects on decision behavior. *Acta Psychologica 174*, 68-79.

Göbel, S. M, Walsh, V., & Rushworth, M. F. (2001). The mental number line and the human angular gyrus. *Neuroimage, 14*(6), 1278-1289.

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology, 51*, 334-384.

Haahr, M. (1998). *List randomizer*. Retrieved from www.random.org.

Jenkins, H. M., & Ward, W. C. (1965). Judgment of contingency between responses and outcomes. *Psychological Monographs: General and Applied*, *79*(1), 1-17.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. *Heuristics and biases: The psychology of intuitive judgment*, 49-81.

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.

LeFevre, J. A., Sadesky, G. S., & Bisanz, J. (1996). Selection of procedures in mental addition: Reassessing the problem size effect in adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(1), 216-230.

Liljeholm, M., & Cheng, P. W. (2009). The influence of virtual sample size on confidence and causal-strength judgments. *Learning, Memory*, *35*(1), 157-172.

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

Matute, H., Steegen, S., & Vadillo, M. A. (2014). Outcome probability modulates anticipatory behavior to signals that are equally reliable. *Adaptive Behavior*, *22*(3), 207-216.

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgments of numerical inequality. *Nature 215*, 1519-1520.

Musca, S. C., Vadillo, M. A., Blanco, F., & Matute, H. (2010). The role of cue information in the outcome-density effect: Evidence from neural network simulations and a causal learning experiment. *Connection Science*, *22*(2), 177-192.

Nieder, A., & Merten, K. (2007). A labeled-line code for small and large numerosities in the monkey prefrontal cortex. *The Journal of Neuroscience, 27*(22), 5986-5993.

Perales, J. C., & Shanks, D. R. (2003). Normative and descriptive accounts of the influence of power and contingency on causal judgement. *The Quarterly Journal of Experimental Psychology*, *56*(6), 977-1007.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/.

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

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*(1), 3–22.

Torchiano, M. (2017). effsize: Efficient effect size computation. R package version 0.7.1. Retrieved from https://CRAN.R-project.org/package=effsize

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

Vallée-Tourangeau, F., Payton, T., & Murphy, R. A. (2008). The impact of presentation format on causal inferences. *European Journal of Cognitive Psychology*, *20*(1), 177-194.

Van Overwalle, F., & Van Rooy, D. (2001). When more observations are better than less: A connectionist account of the acquisition of causal strength. *European Journal of Social Psychology*, *31*(2), 155-175.

Wasserman, W., Dorner, 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.

Wasserman, E. A., Elek, S. M., Chatlosh, D. L., & Baker, A. G. (1993). Rating causal relations: Role of probability in judgments of response-outcome contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(1), 174-188.

Ward, W. C., & Jenkins, H. M. (1965). The display of information and the judgment of contingency. *Canadian Journal of Experimental Psychology*, *19*, 231-241.

White, P. A. (2003). Making causal judgments from the proportion of confirming instances: The pCI rule. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(4), 710-727.

Appendix A

Imagine the following: A pharmaceutical company is developing an allergy medicine comprised of several minerals. The company is working with 12 different laboratories to study the effect that each individual mineral has on headaches. Each laboratory is responsible for investigating the effects of one mineral.

Now imagine that you work for the pharmaceutical company. It is your job to evaluate the results of each study and determine what effect each mineral has on headaches.