The Influence of Reductionist Information on Perceptions of Scientific Validity

by

Rebecca Rhodes

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Psychology)
in the University of Michigan
2015

Doctoral Committee:

Professor Priti R. Shah, Chair
Professor Susan A. Gelman
Professor J. Frank Yates
Associate Professor Carolyn Yoon

ACKNOWLEDGMENTS

Firstly, to my primary advisor, Priti Shah, who has been an exceptional mentor to me over the past five years and without whose encouragement and guidance I would not have been able to complete this dissertation or this degree; to my committee members: Susan Gelman, Frank Yates, and Carolyn Yoon, whose sage remarks have enriched this work significantly; to Thad Polk, who kindly mentored me and taught me how to conduct rigorous functional neuroimaging research – an experience which, I now realize, helped inspire the questions underlying the topic of this dissertation; to Fernando Rodriguez, a collaborator who first alerted me to the incipient "seductive allure of neuroscience" studies, which I could not resist exploring further; to members of the Basic and Applied Cognition Lab and Computational and Cognitive Neuroscience Lab, who have listened to dozens of practice presentations over the years and whose insights have contributed greatly to the development of this project and many others; to those who mentored me at the Educational Testing Service and gave me incredible opportunities to develop professionally as a scientist, which have benefited this work; to my friends, who are my role models and who have provided invaluable emotional support in many ways, from faithfully serving as Netflix binge-watching partners to lending an ear without hesitation; and finally, to my family, who have always been my loudest cheerleaders, especially during this journey – a nontrivial feat considering the 1,448 miles between us.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The ability to reason scientifically about evidence is an important skill for many everyday decisions, ranging from whom to choose in the next presidential election to whether or not to immunize your child. Evidence for the importance of scientific reasoning can be found in continued attempts to improve the teaching of reasoning skills within the educational system. However, in order to develop effective strategies for reasoning about scientific evidence, it is important to understand lay conceptions about what it means for something to be "scientific". In the present work, I examine the influence that reductionist evidence – that is, evidence that comes from micro-level processes, such as biological or neurological processes – has on perceptions of scientific validity. Across eight experiments, I demonstrate that reductionist evidence tends to be viewed as more explanatory and more conclusive than comparable evidence from macro-level processes, such as psychological processes. Interestingly, the preference for reductionist information does not go away with education. In fact, people with greater scientific literacy are even more likely to assume that reductionist information is superior to macro-level information. I interpret this finding as evidence that the preference for reductionist information is not irrational, but instead an expected consequence of traditional science curricula. I demonstrate several important implications of reductionist preference. For example, this preference increases the likelihood of making causal inferences from the results of research studies that suggest micro-level – as opposed to

macro-level – mechanisms, and it can decrease the size of the sample one needs to feel confident in accepting conclusions from these studies. I relate these findings to current pervasive issues in the scientific community, such as publication biases and the prevalence of underpowered studies utilizing reductionist approaches. I also discuss educational strategies that could encourage holistic thinking about science – specifically, emphasizing science as a tool for thinking strategically about everyday phenomena, regardless of the level of analysis, rather than a collection of discrete facts obtained by the use of technology and equipment.

CHAPTER 1

Introduction

Much of the behavioral science that lay readers are exposed to aims to influence their behavior in some way (e.g., “Set Goals and You May Just Live Longer”, “The Eyes are the Window to Your Potential Soulmate”, “To Work Better, Work Less”), and how readers evaluate the science is likely to have a significant influence on the decision they make from it. Frequent claims are made without solid scientific justification, and careful evaluations of popular science help combat the influence of misleading evidence. Lay readers’ everyday scientific reasoning skills, however, are not what they should be – for example, a recent Gallup poll revealed that 75% of Americans hold at least one pseudoscientific belief (Moore, 2005). Moreover, even if one has the tools to reason scientifically, scientific reasoning is an effortful process (Evans, 2003; Schiffrin & Schneider, 1977; Sloman, 1996) and the degree to which one employs such skills can be influenced by a number of factors, both dispositional and contextual (Cacioppo & Petty, 1982; Frederick, 2005; Lord, Ross, Lepper, 1979; Stanovich & West, 1997). Given the large influx of research reports that grace news feeds on a daily basis, many making claims that aim to change people’s behavior or beliefs in some way, it is important to understand the assumptions people hold about the kinds of evidence that confer scientific legitimacy. In the context of research evaluation, the present work examines how contextual factors – specifically, the level of analysis at which the evidence or explanation is reported – and individual differences affect judgments of perceived scientific validity and explanation quality.

Everyday Scientific Reasoning

Research has examined the influence of several dispositional factors on everyday scientific reasoning, such as the ability to coordinate theory and evidence (Kuhn, 2001), the ability to reason independently of prior beliefs, (Stanovich & West, 1997), and the ability to apply statistical reasoning to everyday situations (Nisbett, Fong, Lehman, & Cheng, 1987). Given that most people encounter scientific information online in the context of media reports, recent work has also examined how various features of scientific reports influence readers' evaluations of scientific quality. For example, research has found that the inclusion of scholarly references, use of the passive voice, and the presence of a methods section makes a report seem more scientific (Thomm & Bromme, 2012). More specifically, some work has found that the *type* of methods reported can influence evaluations of the research. For example, research has suggested that the presence of neuroscience can make a scientific explanation seem better and the presence of a mathematical equation can make a research study seem higher in quality (Eriksson, 2012; Weisberg, Keil, Goodstein, Rawson, and Gray, 2008). These findings raise the possibility that the evidential level of analysis is an important factor to which people attend. It is possible that people may inherently value reductionist evidence more so than evidence from a higher level of analysis. Although some research suggests this may be the case, a more thorough examination is needed. Furthermore, if it is the case that people value micro-level information more than macro-level information, it is important to know whether this preference is driven by superficial cues and cognitive laziness or a deliberate conclusion that micro-level information is superior. Distinguishing between these possibilities could inform mitigation strategies. In this dissertation, I aimed to gain insight into people's assumptions about micro- and macro-level information and examine how individual differences interact with evidential level of analysis to influence scientific reasoning.

Levels of Analysis

One assumption people may have is that evidence that results from a micro-level process, such as a biological or neurological process, as opposed to a macro-level process, such as a psychological process, confers more legitimacy to a phenomenon. People typically have little direct experience with micro-level processes, and processes at this level often require formal training to fully understand. In contrast, people have extensive experience with macro-level processes such as physical and mental states, so much so that they often develop their own theories about how these processes work (e.g., Olson & Bruner, 1996). The validity of an explanation or piece of evidence does not necessarily depend on its level of analysis. However, the public fascination with neuroscience and genetics, and the tendency to use evidence from these fields to legitimize certain phenomena (O'Connor, Rees, & Joffe, 2012; Racine, Bar-Ilan, & Illes, 2005), suggests that micro-level evidence may be perceived as fundamentally superior to macro-level evidence. Such an assumption could pose challenges to everyday scientific reasoning. For example, consumers could be duped by pseudoscientific claims that appeal to micro-level processes to boost their perceived legitimacy. Currently there are a number of "brain-training" programs aimed at improving cognitive function. In their advertising, some of these programs claim to be based on neuroscience – an example of utilizing micro-level processes to legitimize the program. However, evidence that such programs accomplish what they claim to is mixed, so the use of neuroscience as a marketing tool in this instance is potentially misleading (Chancellor & Chatterjee, 2011). Whether the suggestion of a micro-level basis improves the perceived validity of these claims is an important research question to address.

Causal Reasoning and Explanation

Judgments of scientific validity involve making causal attributions for a set of factors and a target effect; therefore, before one can address the effect of reductionist information on perceptions of scientific validity and explanation quality, it is important to understand why reductionist information could affect causal reasoning in the first place. One reason is that moving the level of analysis from the macro- to the micro-level could appear to uncover the causal pathways that give rise to a phenomenon. In other words, people may believe that micro-level evidence or explanations provide more information about mechanism than do macro-level evidence/explanations. People pay great attention to mechanism when making causal attributions (Newsome, 2003). For example, when people must identify which of several factors caused an effect, researchers have found that people's decisions rely more on the mechanistic relationship of each factor to the target effect than the degree of covariation between each factor and the target effect (Ahn, Kalish, Medin, & Gelman, 1995). Further evidence for the importance of mechanism can be seen in a series of juror decision making studies in which verdicts (a causal judgment) depend on how well the evidence fits into a coherent story (Pennington & Hastie, 1986, 1992). When given evidence about a hypothesized factor and a target effect, people have a natural tendency to construct a coherent story about how that factor could have elicited the effect (Ahn & Bailenson, 1996). Although the importance of mechanism for causal inference is well established, it is unclear whether the level of analysis at which the mechanism takes place influences the perceived validity of the phenomenon at hand. Thus, an important question addressed in this dissertation is whether people are more likely to believe in a phenomenon when they are given a micro-level, as opposed to a macro-level, mechanism.

Studies examining genetic attributions and stereotype endorsement provide some evidence that micro-level information may be more likely to elicit causal judgments than macro-level information. Genetic evidence is often used as an explanation for group differences – such as gender and racial differences – on a number of human attributes. The causes of these group differences likely involve a number of interacting factors, both sociocultural (macro-level) and genetic (micro-level). Several research studies have demonstrated that people are more likely to endorse group stereotypes if they read a genetic explanation for group differences (Bastian & Haslam, 2006; Brescoll & LaFrance, 2004; Keller, 2005). In other words, people are more likely to make the inference that belonging to a certain group causes a particular outcome if the suggested mechanistic pathway is genetic.

If micro-level mechanisms are perceived as being of higher quality or more valid than macro-level mechanisms, it is reasonable to assume that micro-level information could improve the quality of a scientific explanation. Indeed, Weisberg et al. (2008) found that the presence of micro-level information – in this case, neuroscience details – significantly improved the perceived quality of explanations for psychological phenomena. In their study, Weisberg and colleagues presented participants with descriptions of 18 psychological phenomena and an accompanying explanation. They manipulated whether the quality of the explanation was bad (circular) or good and also whether the explanation involved irrelevant neuroscience information or not. Interestingly, irrelevant neuroscience information had the largest effect on bad explanations. Given the importance of mechanism in understanding causation, one reason for this dissociation may lie in the relative contribution the neuroscience information made to the causal structure of the explanation. More specifically, the good explanations already identified a causal – albeit behavioral – mechanism, so the inclusion of neuroscience information made less of a contribution for these explanations than it did for the bad explanations, which did not identify a causal mechanism at all.

Whether neuroscience or other micro-level information also improves the perceived validity of a phenomenon has yet to be addressed in the literature. There are important differences between judging explanation quality and scientific validity. Explanation quality is a retrospective judgment that could be primarily driven by whether a mechanism was identified; it is possible that the level of analysis of the mechanism (micro vs. macro) may not matter as much as whether a mechanism was identified at all. The findings in Weisberg et al. (2008) could be explained by this hypothesis, and the present work tests this hypothesis directly. Perceived validity, however, is a prospective judgment similar to causal attribution but in a slightly stronger form. Validity implies that causation is not due to error but is a true reflection of the real world. The degree to which neuroscience information, or other micro-level information, affects perceptions of scientific validity is important because these perceptions have significant implications for everyday decision-making.

The Gap Between Perceived and Real Understanding

Although people might believe that micro-level mechanisms result in better explanation quality, the reality is that most people do not understand micro-level pathways; therefore, it is unlikely that the perception of explanatory depth translates to real understanding of causation for the lay consumer. For example, most people acquire their knowledge about genetics from the popular media (Dar-Nimrod & Heine, 2011), so they are unlikely to have a real understanding of the suggested genetic mechanisms underlying group differences. Thus, the assumption that micro-level information improves explanation quality creates fertile ground for being influenced by pseudoscientific jargon.

Although people are unlikely to fully understand micro-level mechanisms, their perceived understanding could still be high. Support for this assertion comes from Rozenblit and Keil (2002), who demonstrated an illusion of explanatory depth. Rozenblit and Keil (2002) found that people initially overestimate their knowledge of how devices work, and only after being forced to delineate the step-by-step mechanism do their estimates decrease and become more accurate indicators of their level of knowledge. The literatures on overconfidence (Dunning, Griffin, Milojkovic, & Ross, 1990; Fischhoff, Slovic, & Lichtenstein, 1977) and the feeling of knowing (Koriat, 1993) provide additional examples of this dissociation between real and perceived performance. As suggested by the findings from Weisberg et al. (2008), it is possible that the level of analysis of an explanation could influence *perceived* understanding of a phenomenon, even if it does not increase real understanding. Furthermore, if readers think they have a mechanistic understanding of a phenomenon, they might be likely to believe the phenomenon is valid, as well.

Neuroscience Information

One type of reductionist information the current work focuses on is neuroscience information. A few studies have looked at the effect of neuroscience information on scientific reasoning, but there are a number of unanswered questions. As previously discussed, Weisberg and colleagues demonstrated that irrelevant neuroscience information can make bad explanations of psychological phenomena seem better. Michael, Newman, Vuorre, Cumming, & Garry (2013) recently ran a series of five replications of this study and estimated the effect size of neuroscience information to be 0.40, 95% CI [0.23, 0.57]. Another type of study has focused on the effect of brain images, rather than neuroscience text (McCabe & Castel, 2008). McCabe and Castel had their participants read three fictional articles about brain imaging studies (accompanied by a brain image, a bar graph, or nothing else) and asked them to rate the writing quality, the aptness of the article's title, and the scientific reasoning in the article. The presence of a brain image improved ratings of writing quality and scientific reasoning. More recent studies, however, have been unable to replicate these findings (Gruber & Dickerson, 2012; Hook & Farah, 2013; Michael et al., 2013). Hook and Farah (2013) found no main effect of brain images on research evaluations but predicted that the influence of brain images could be moderated by dualistic beliefs – specifically, people who think of the mind and brain as separate entities may be the ones who are fascinated to see evidence that the brain is involved in complex cognitive and emotional phenomena. However, they found that dualistic beliefs did not predict the effect of brain images on research evaluations. Furthermore, a recent meta-analysis comprising the original data from McCabe and Castel (2008), as well as data from 10 replications which used a mixture of media (online vs. paper), subject pools (Mechanical Turk, undergraduates, high school students, general public) and compensation strategies ($0.30, $0.50, course credit, movie voucher, no compensation), found that brain images exerted little to no influence on credibility (Michael et al., 2013).

While the previous set of findings may seem contradictory, an important difference between the Weisberg et al. (2008) studies and the brain image studies is that the brain image studies were primarily testing the effect of adding a brain image to an article that already contained neuroscience language, whereas the Weisberg et al. (2008) studies were testing the effect of adding neuroscience language to an explanation of a behavioral phenomenon. The null effects of brain images may simply indicate that they do not provide additional value beyond that provided by neuroscience language. Importantly, the explanations that were most affected by neuroscience in the Weisberg et al. (2008) studies were those which did not identify a causal mechanism but simply explained the phenomenon by restating the results. Thus, it is unclear whether macro-level information would elicit similar effects if it could be construed as a causal mechanism. The relative influence of neuroscience or other micro-level evidence compared to evidence from a macro-level process has yet to be investigated.

Another shortcoming in previous studies examining the influence of neuroscience information is that they did not control for several important factors that are likely to affect research evaluations. As Table 1.1 illustrates, the outcome variables of the fictional research studies and the ways in which neuroscience information was integrated into the fictional research studies have been inconsistent across studies, which could contribute to the mixed findings. One goal of this dissertation was to systematically manipulate these features of the research study and examine whether they moderate the influence of neuroscience and other micro-level information. Finally, another important factor that has been neglected in several previous research studies is the role of individual differences. Although some studies have examined the roles of dualistic beliefs (Hook & Farah, 2013) and analytical thinking (Fernandez-Duque, Evans, Colton, & Hodges, in press) in predicting susceptibility to neuroscience information – and failed to find significant relationships – a broader examination is needed.

Table 1.1
Stimulus Features in Previous Neuroscience Studies

	Features of Fictional Research study				
Paper	Type of Integration	Outcome Variable	Subjects	Comparison of Interest	Outcome Variable(s) of Interest
Weisberg et al. (2008)	Explanation	16 cognitive, 2 non-cognitive	Human	Neuroscience text vs. no additional text	Quality of explanations
McCabe & Castel (2008)	Evidence & Explanation	Cognitive	Human	Brain image vs. bar graph vs. text; brain image vs. topographical maps; Brain image vs. text	Credibility & reasoning
Gruber & Dickerson (2008)	Evidence	Pseudo-cognitive	Human	No image vs. several different types of images, including brain	Credibility & reasoning
Hook & Farah (2013)	Evidence	2 cognitive, 4 non-cognitive	Human	Brain image vs. bar graph vs. control photo	Credibility & reasoning
Diekmann et al. (in press)	Evidence	Non-cognitive	Human	Neuroscience text vs. no additional text	Interestingness
Fernandez-Duque et al. (2014)	Explanation	16 cognitive, 2 non-cognitive (used stimuli from Weisberg, et al. (2008)	Human	Brain image + neuroscience text vs. neuroscience text only vs. no additional text ; neuroscience text vs. social science text vs. no additional text; neuroscience text vs. hard science text vs. social science text	Quality of explanations

Individual Differences in Reasoning

Although previous studies suggest that micro-level information might be perceived more favorably than macro-level information, the influence of micro-level information on any one individual is likely moderated by a number of factors. This section reviews four important factors that contribute to individual differences in reasoning: prior beliefs, flexibility in thinking, Type 2 processing, knowledge, and essentialist beliefs. The experiments that follow investigate the influence of each of these factors on preferences for micro-level information.

Prior Beliefs

Much research has shown that personal theories or prior beliefs can influence one's depth of reasoning. Lord, Ross, and Lepper (1979) recruited participants who were either proponents or opponents of capital punishment, a polarizing issue. Participants read a description of a research study that was either supportive (found a decrease in murder rates) or not supportive (found an increase in murder rates) of capital punishment. One of the key findings Lord, Ross, and Lepper found was that participants were more convinced by the research study that was congruent with their prior attitudes. Similarly, Klaczynski and Narasimham (1998) found that people used sophisticated thinking strategies to disconfirm evidence that contradicted their religious beliefs. Findings such as these have been interpreted as theory-motivated reasoning, which means that people evaluate evidence in a way that upholds their prior theories of the world (Klaczynski, 2000). Specifically, people who disagree with a claim tend to be more critical of it and engage analytical processes, and people who agree with a claim are less likely to be critical of it and more likely to process it heuristically. Motivated reasoning is common in everyday reasoning situations, so it is conceivable that prior beliefs and/or experiences relating to a research article's claim could affect whether people evaluate the information analytically or heuristically. As a result, the type of reasoning an individual uses could affect his/her susceptibility to reductionist information. In the experiments that follow, participants were typically asked to indicate their beliefs about the claim(s) being made in the research article prior to reading it.

Cognitive Flexibility

Although prior beliefs tend to affect the depth of reasoning one will use to evaluate evidence, they do not affect everyone. Some individuals are skilled at reasoning in a more objective way, independent of their own personal beliefs. The tendency to do this can be measured by the Actively Open-Minded Thinking Scale (AOT; Stanovich & West, 1997). This 41-item scale asks about people's ability to think flexibly and be open to new information, regardless of what they personally believe. A high score on the AOT scale reflects more sophisticated thinking dispositions; specifically, it indicates a motivation to have accurate beliefs, even if that means changing one's current beliefs. Although AOT performance is correlated with cognitive ability, the two constructs are separable. Performance on the AOT predicts data-driven thinking during argument evaluation tasks, even after partialling out the variance associated with cognitive ability (Stanovich & West, 1997). Given that people who score highly on the AOT scale are more likely to reason in a data-driven, as opposed to a belief-driven way (Stanovich & West, 1997), it is possible that these people might be less influenced by reductionist information.

Type 2 Processing

The Cognitive Reflection Test (CRT; Frederick, 2005) is a widely used measure of one's ability to suppress an intuitive response, resulting from heuristic processing, in favor of a more deliberate response. This test consists of three items that tend to elicit automatic, but incorrect, answers. The correct answer requires more thinking than it initially seems. The CRT is correlated both with cognitive ability (Frederick, 2005; Toplak, West, & Stanovich, 2011) and rational thinking measures such as syllogism problems with belief bias (Toplak, West, & Stanovich, 2011). CRT performance also predicts performance on many heuristics and biases tasks (Cokely & Kelley, 2009; Frederick, 2005; Toplak, West, & Stanovich, 2011). Importantly, Toplak and colleagues found that the CRT predicts rational thinking and performance on heuristics and biases tasks after partialling out the variance associated with assessments of intelligence, thinking dispositions, executive functions, and cognitive skills. Thus, people who score highly on the CRT can be categorized as people who are more likely to engage in rational, analytic thinking. If reductionist information is processed heuristically, it is possible that people who score highly on the CRT would be less influenced by reductionist information.

Knowledge

Education and knowledge are important indicators of reasoning ability. Previous research suggests that domain-relevant knowledge might inoculate one against the influence of neuroscience. Weisberg et al. (2008) administered their stimuli to an expert population, defined as individuals who had completed an advanced degree in cognitive psychology or a related field, and found that they were not influenced by irrelevant neuroscience information. It is possible that people who have higher methodological and/or scientific knowledge may base their evaluations more on the research methodology and less on the extraneous information suggesting micro- or macro-level processes. If so, these individuals may be less influenced by micro-level information. The stimuli used in the current work mostly involved research findings, so I expected that individuals' knowledge of science and methodological principles would be an important predictor of their ability to reason about the evidence. The present work examined the influence of several different types of knowledge: methodological knowledge, as indicated by familiarity with principles such as random assignment, selection bias, and sample size; brain knowledge, as indicated by basic neuroanatomy questions from introductory psychology textbooks; and scientific knowledge, as indicated by knowledge about basic biology and physics.

Essentialist Beliefs

Research suggests that some people believe human attributes result from inalterable underlying essences (Gelman, 2003). For example, someone with strong essentialist beliefs might be more likely to believe that gender differences on a variety of domains are due to biological, rather than social, causes. Essentialist beliefs have been found with regard to a variety of human attributes, including social categories (Dar-Nimrod & Heine, 2011; Mahalingam, 2003) and personality (Halsam, Bastian, & Bissett, 2004). Bastian and Haslam (2006) developed an 18-item Biological Essentialism Scale which assesses the extent to which participants hold essentialist beliefs. More specifically, the scale measures the extent to which a person believes that human traits are determined biologically, that traits are discrete, and that traits can be determined quickly. People who have strong essentialist beliefs may also be more likely to reduce complex human attributes to a reductionist cause. If so, it is conceivable that people with strong essentialist beliefs would show stronger preferences for micro-level information.

Processes Underlying the Neuroscience Effect

Some research studies have tried to examine the processes underlying the neuroscience effect. Weisberg et al. (2008) suggested that neuroscience could act as a seductive detail, distracting people from paying attention to other methodological details present in the study, but this possibility has yet to be tested empirically. Another possibility could be that people simply prefer language that sounds technical, but extant research suggests that this is not the case. Fernandez-Duque et al. (in press) gave people the descriptions of the same phenomena used in Weisberg et al. (2008) and had them read explanations based on social science, neuroscience, or another hard science. The hard science information was essentially technical jargon and did not clearly relate to the psychological phenomenon being studied, whereas the neuroscience information was also technical but was arguably related to the psychological phenomenon in a more straightforward way. Fernandez-Duque and colleagues found that neuroscience explanations were still more satisfying than hard science explanations, suggesting that technical language is not enough to elicit the effect. Instead, another factor may be at play – people may believe that the brain is the best explanation for mental phenomena. Hook and Farah (2008) ruled out the possibility that dualistic beliefs moderate the neuroscience effect, and Fernandez-Duque et al. (in press) found that analytical thinking did not protect against the neuroscience effect.

Through an in-depth look at individual differences and systematic manipulations of the features of the fictional research studies, this work sheds light on the processes underlying the preference for micro-level information. Specifically, I provide insight into whether the influence of micro-level information is due to heuristic thinking or deliberative strategies. For example, Experiments 2.1 through 3.2 (Chapters 2 and 3) directly assessed whether neuroscience information acts as a seductive detail. If so, this might suggest that the influence of neuroscience is due to an attentional bias. Experiments 2.1 through 3.2 (Chapters 2 and 3) also investigated the relationship between micro-level preferences and sophisticated thinking dispositions, as measured by the CRT and AOT scales. If people who have less sophisticated thinking dispositions are the ones who are more influenced by micro-level information, this might suggest that it is a result of a heuristic reasoning process; in contrast, if people who have more sophisticated thinking dispositions are more influenced by micro-level information, this might suggest that the valuation of micro-level information is the result of a deliberative strategy. Similarly, Experiments 2.1 through 4.2 (Chapters 2, 3, and 4) examined the influence of various types of knowledge: methodological knowledge, brain knowledge, and scientific literacy. If knowledge acts as an inoculation, this might again suggest that the preference is due to a heuristic reasoning process, whereas if knowledge is associated with greater micro-level preference, this suggests a more deliberative process. Finally, Experiments 2.1 through 3.4 (Chapters 2 and 3) examined the role of prior beliefs and prior behavior. Given that people are more likely to be critical of information that they do not already believe in, and less critical of information that they do believe in, assessing the influence of prior beliefs on micro-level preference will provide further insight into how micro-level information affects scientific reasoning.

Understanding whether the appeal of micro-level information is due to heuristic or deliberative processes is important for several reasons. If the preference is due to a heuristic process, this might suggest that the preference depends heavily on the context and, in situations where one is at risk for committing a bias, one can take efforts to adopt a more deliberative mindset and evaluate the evidence more objectively. Indeed, many strategies exist for mitigating other cognitive biases, and strategies could likely be developed for overcoming the influence of alluring micro-level information. It might also be the case that individual differences in rational thinking abilities, such as AOT and CRT performance, can mitigate the influence of micro-level information. If, however, the preference for micro-level information is due to a deliberative process, it suggests that this effect is due not to cognitive laziness but beliefs about the validity of evidence that comes from different scientific fields. If this is the case, it could be challenging to mitigate the influence of micro-level information when it is not an appropriate criterion to use for judging scientific validity.

Research Questions

In this work, I addressed several research questions. Experiments 2.1 through 3.2 (Chapters 2 and 3) examined the influence of neuroscience information compared to irrelevant information or psychology information on perceptions of mechanistic understanding and scientific validity. These studies also investigated the interplay of prior beliefs, AOT, CRT, and methodological knowledge. Experiment 3.3 (Chapter 3) examined the influence of neuroscience information on mechanistic understanding and scientific validity compared to other technical, but not brain-based information, and also examined the role of scientific literacy. Experiment 3.4 (Chapter 3) examined whether the influence of neuroscience information on perceived mechanistic understanding and scientific validity is limited to phenomena that are explicitly cognitive in nature. Experiment 4.1 (Chapter 4) examined the extent to which individuals prefer micro-level information across a variety of research scenarios, and whether the way in which the micro-level information is integrated into the research study influences ratings of perceived quality and understanding. Finally, Experiment 4.2 (Chapter 4) investigated the implications of micro-level preferences by examining whether they decrease the number of subjects one needs in a research study to be sufficiently confident in its conclusion.

References

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

Ahn. W., & Bailenson, J. (1996). Causal attribution as a search for underlying mechanisms: An explanation of the conjunction fallacy and the discounting principle. *Cognitive Psychology, 31,* 82-123.

Bastian, B., & Haslam, N. (2006). Psychological essentialism and stereotype endorsement. *Journal of Experimental Social Psychology*, *42*, 228-235.

Brescoll, V., & LaFrance, M. (2004). The correlates and consequences of newspaper reports of research on sex differences. *Psychological Science*, *15*, 515-520.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*, 116.

Chancellor, B., & Chatterjee, A. (2011). Brain branding: When neuroscience and commerce collide. *AJOB Neuroscience*, *2*, 18-27.

Cokely, E. T., & Kelley, C. M. (2009). Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation. *Judgment and Decision Making*, *4*, 20–33.

Dar-Nimrod, I., & Heine, S. J. (2011). Genetic essentialism: On the deceptive determinism of DNA. *Psychological Bulletin, 137,* 800-818.

Dunning, D., Griffin, D. W., Milojkovic, J. D., & Ross, L. (1990). The overconfidence effect in social prediction. *Journal of Personality and Social Psychology*, *58*(4), 568.

Eriksson, K. (2012). The nonsense math effect. *Judgment and Decision Making, 7*(6), 746-749.

Evans, J. St. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences*, *7*, 454-459.

Fernandez-Duque, D., Evans, J., Colton, C., & Hodges, S. D. (in press). Superfluous neuroscience information makes explanations of psychological phenomena more appealing. *Journal of Cognitive Neuroscience*.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 552-564.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*, 25-42.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought.* Oxford: Oxford University Press.

Gruber, D., & Dickerson, J. A. (2012). Persuasive images in popular science: Testing judgments of scientific reasoning and credibility. *Public Understanding of Science, 21,* 938-948.

Haslam, N., Bastian, B., & Bissett, M. (2004). Essentialist beliefs about personality and their implications. *Personality and Social Psychology Bulletin, 30*, 1661–1673.

Hook, C. J., & Farah, M. J. (2013). Look again: Effects of brain images and mind-brain dualism on lay evaluations of research. *Journal of Cognitive Neuroscience, 25,* 1397-1405.

Keller, J. (2005). In genes we trust: The biological component of psychological essentialism and its relationship to mechanisms of motivated social cognition. *Journal of Personality and Social Psychology, 88,* 686-702.

Klaczynski, P. A. (2000). Motivated scientific reasoning biases, epistemological beliefs, and theory polarization: A two-process approach to adolescent cognition. *Child Development, 71,* 1347-1366.

Klaczynski, P. A., & Narasimham, G. (1998). The development of self-serving reasoning biases: Ego-protective versus cognitive explanations. *Developmental Psychology, 34*, 175-187.

Koriat, A. (1993). How do we know what we know? The accessibility model of the feeling of knowing. *Psychological Review, 100,* 609-639.

Kuhn, D. (2001). How do people know? *Psychological Science, 12*(1), 1-8.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, *37*, 2098-2109.

Mahalingam, R. (2003). Essentialism, culture, and beliefs about gender among the Aravanis of Tamil Nadu, India. *Sex Roles, 49*, 489–496.

McCabe, D. P., & Castel, A. D. (2008). Seeing is believing: The effect of brain images on judgments of scientific reasoning. *Cognition, 107,* 343-352.

Michael, R. B., Newman, E. J., Vuorre, M., Cumming, G., Garry, M. (2013). On the (non)persuasive power of a brain image. *Psychonomic Bulletin & Review, 20,* 720-725.

Moore, D. W. (2005, June 16). Three in four Americans believe in paranormal. *Gallup Poll News Service*. Retrieved from http://www.gallup.com/poll/16915/three-four-americans-believe-paranormal.aspx.

Newsome, G. L. (2003). The debate between current versions of covariation and mechanism approaches to causal inference. *Philosophical Psychology, 16*, 87-107.

Nisbett, R. E., Fong, G. T., Lehman, D. R., & Cheng, P. W. (1987). Teaching reasoning. *Science, 238*, 625-631.

O'Connor, C., Rees, G., & Joffe, H. (2012). Neuroscience in the public sphere. *Neuron, 74*, 220-226.

Olson, D. R., & Bruner, J. S. (1996). Folk psychology and folk pedagogy. *The Handbook of Education and Human Development*, 9-27.

Pennington, N., & Hastie, R. (1986). Evidence evaluation in complex decision making. *Journal of Personality and Social Psychology, 51*, 242-258.

Pennington, N., & Hastie, R. (1992). Explaining the evidence: Tests of the story model for juror decision making. *Journal of Personality and Social Psychology, 62,* 189-206.

Racine, E., Bar-Ilan, O., & Illes, J. (2005). fMRI in the public eye. *Nature Reviews Neuroscience, 6*, 159-164.

Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science, 26,* 521-562.

Shiffrin, R. M. & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review, 84*, 127-190.

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

Stanovich, K. E., & West, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology, 89,* 342-357.

Thomm, E., & Bromme, R. (2011). "It should at least seem scientific!" Textual features of "scientificness" and their impact on lay assessments of online information. *Science Education, 96*(2), 187-211.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition, 39,* 1275-1289.

Weisberg, D. S., Keil, F. C., Goodstein, J., Rawson, E., & Gray, J. R. (2008). The seductive allure of neuroscience explanations. *Journal of Cognitive Neuroscience, 20*, 470-477.

CHAPTER 2

Explaining the Alluring Influence of Neuroscience Information on Scientific Reasoning

Introduction

The *New York Times* recently featured an article arguing that people are not merely addicted to their iPhones, but actually feel love for them in the same way they feel love for a significant other (Lindstrom, 2011). This conclusion was based on an fMRI experiment that found similar levels of insular cortex activity when individuals thought about their significant others and when they thought about their iPhones. Explaining behavior by citing data from neuroscience has become a common trend in the media. The degree to which neuroscience evidence is particularly alluring has been a hot topic in research on scientific reasoning. Specifically, some work has suggested that providing neuroimaging data or irrelevant neuroscience explanations makes readers more likely to believe, and be less critical of, scientific information (McCabe & Castel, 2008; Weisberg, Keil, Goodstein, Rawson, & Gray, 2008). Thus, one of the potential consequences of highlighting neuroscience information is that it may persuade individuals to believe a claim to be more valid than what the evidence actually implies.

Although some recent studies have found that neuroscience information or images affect readers' evaluations of scientific studies, others have not found neuroscience to have a significant impact (Farah & Hook, 2013; Green & Cahill, 2012; McCabe &

Castel, 2008; Weisberg et al., 2008). For example, although one experiment found significant effects of brain images on judgments of reasoning (McCabe & Castel, 2008), another found brain images to be no more influential than other types of images (Gruber & Dickerson, 2012). One possible reason for this discrepancy is that individual factors may influence how, or even whether, neuroscience information affects judgments. For instance, previous research on reasoning finds that individuals' prior beliefs predict the tendency to engage in belief-biased reasoning. In situations where the presented evidence is incongruent with one's prior beliefs, individuals may be more motivated to be critical and, therefore, less susceptible to the influence of neuroscience. Additional factors like scientific knowledge and thinking dispositions are also related to how individuals reason about claims, and those with more knowledge and sophisticated thinking styles may be less influenced by the presence of neuroscience information. In this experiment, we shed light on inconsistencies in the literature by exploring how the influence of neuroscience information may be moderated by individuals' prior beliefs and their dispositions towards critical thinking.

The Influence of Neuroscience Information on Reasoning

Several recent studies have suggested that neuroscience information can influence the way people evaluate scientific evidence. Weisberg et al. (2008) found that including an irrelevant sentence that contained neuroscience information made individuals evaluate poor explanations more favorably. Similarly, McCabe & Castel (2008) found that pictures of brain activations, compared to bar graphs or topographical brain maps, presented with a news article made individuals give higher ratings of the article's scientific reasoning quality. Similar findings have been found in practical contexts, such as jury decision-making (Green & Cahill, 2012). However, more recent studies have only found trivial effects of neuroscience (Hook & Farah, 2013; Michael, Newman, Vuorre, Cumming, & Garry, 2013).

Several explanations have been offered for the potential effect of neuroscience. One possibility offered by Weisberg et al. (2008) is that neuroscience information acts as a seductive detail. Literature on seductive details reveals that text that is considered fascinating but irrelevant can actually impair one's ability to encode the important details of instructional material (Harp & Mayer, 1998; Rey, 2012). People tend to view neuroscience with fascination (Tallis, 2012) and, if it acts as a seductive detail, it is possible that neuroscience may distract people from paying due attention to other important details of a research study. Alternatively, another explanation may be that people prefer a reductionist explanation of complex phenomena (Keil, 2006; McCabe & Castel, 2008; Weisberg et al., 2008). For example, Keil (2006, p. 242) discusses an "illusion of explanatory depth" in which people can be misled into thinking they understand how complex systems work when they can visualize parts of that system. In other words, the concrete nature of neuroscience explanations may lead to a false sense of clarity and understanding, heightening its perceived value. To clarify the mechanism by which neuroscience information influences the evaluation of scientific evidence, the present studies test the possibility that irrelevant neuroscience information impairs the recognition of flawed evidence and/or increases perceived understanding of a phenomenon.

Prior Beliefs and Reasoning

Prior beliefs have a robust influence on our tendency to be critical of new information. Lord, Ross, & Lepper (1979) showed that, when giving participants the same data summaries from empirical studies, those who were initially in favor of the claim rated the study as more convincing evidence than those who disagreed with the claim. Other research has found similar effects; for example, people who receive preference-inconsistent information are more likely to give sophisticated and skeptical responses than those who receive preference-consistent information (Ditto & Lopez, 1992), and people are more likely to use statistical principles such as base rates and the law of large numbers when the attainment of a desired solution requires that they do so (Ginossar & Trope, 1987; Sanitioso & Kunda, 1991). Dual process explanations suggest that belief-biased reasoning results from activation of the heuristic system when evidence is congruent with personal beliefs and the analytic system when evidence is incongruent (Evans, 2003; Klaczysnki, 2000; Kunda, 1990;). As such, people with congruent beliefs might be especially likely to use neuroscience information (even if irrelevant), as additional evidence to support their point of view, without considering whether that evidence actually improves the claim. Additionally, given that individuals are more likely to be critical of claims they do not agree with, it is possible that holding incongruent prior beliefs inoculates one against any potential influence of neuroscience. Although prior beliefs share an important relationship with reasoning, they have been largely ignored in studies looking at the influence of neuroscience information.

Thinking Dispositions and Reasoning

Another factor relevant to understanding how neuroscience information influences reasoning is individual thinking dispositions. Individuals differ in their ability to be flexible in their thinking and consider evidence that contradicts their beliefs. For example, research suggests that individuals high in actively open-minded thinking are less susceptible to being influenced by their prior beliefs (Stanovich & West, 1997). Additionally, individuals differ in their ability to override automatic responses and engage in deliberative processing and this ability predicts performance on numerous classic heuristics and biases tasks (Toplak, West, & Stanovich, 2011). Both of these thinking dispositions are related to reasoning abilities and may influence the extent to which individuals are affected by neuroscience information and prior beliefs.

Overview

We conducted two experiments to examine the influence of irrelevant neuroscience information on lay reasoning in the context of public news reports about a scientific finding. Experiment 2.1 assessed the influence of neuroscience information on news article evaluations among participants who already had prior beliefs (either congruent or incongruent) about the claim, controlling for individual differences in thinking dispositions and knowledge. Experiment 2.2 controlled for the number of words in the news article and measured evaluations for individuals with congruent, incongruent, and neutral prior beliefs. To address potential processes by which irrelevant neuroscience information affects evaluations, Experiment 2.2 tested whether neuroscience distracts people from attending to important details, increases one's feeling of understanding, or both.

Experiment 2.1

The goal of Experiment 2.1 was to examine the influence of neuroscience information on scientific reasoning after controlling for other variables that affect everyday reasoning. We predicted that prior beliefs (congruent/incongruent with the claim), knowledge of methodological principles, and thinking dispositions (e.g., the ability to think flexibly and the ability to override prepotent responses) would predict how favorably participants evaluate the evidence (Klaczynski, 2000; Stanovich & West, 1997; Toplak, West, & Stanovich, 2011), and that more consistent effects of neuroscience information would emerge after accounting for the unique variance associated with these factors.

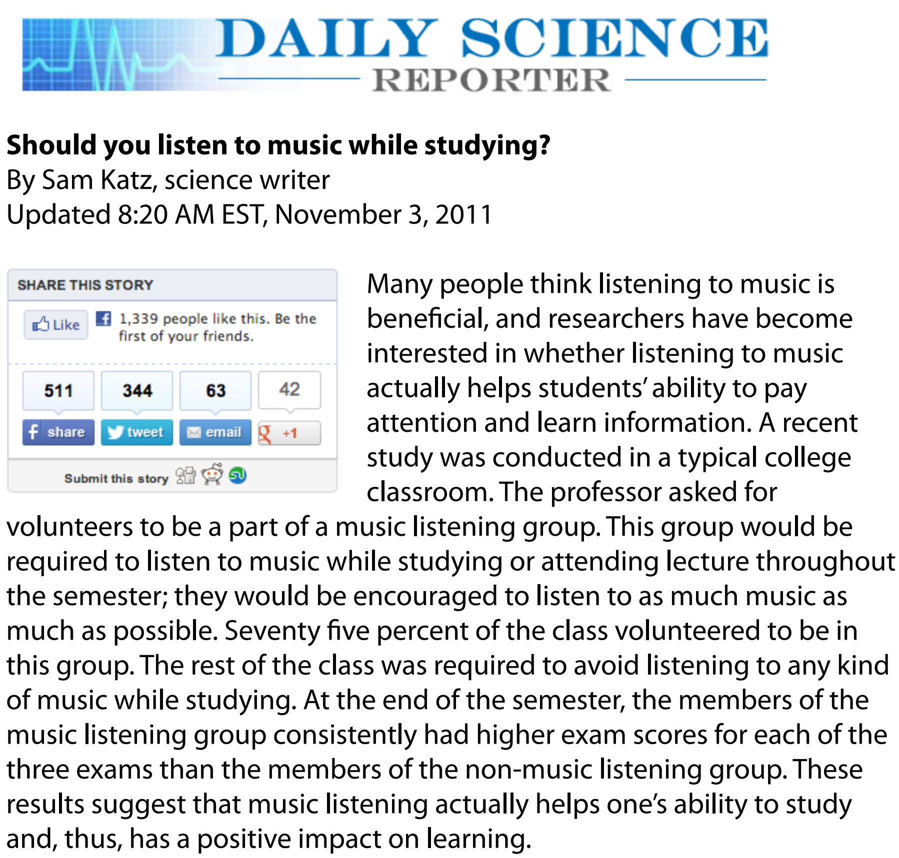
Method

Participants

The experiment was conducted online through Amazon's Mechanical Turk, a crowdsourcing system in which thousands of users can complete tasks for monetary compensation and that has been shown to yield high-quality data (Buhrmester, Kwang, & Gosling, 2011). Participants were 201 adults (110 females; median age = 30; range = 18-72) from the U.S. Participants were told that they would spend 10-15 minutes reading a brief news clipping and answering some questions. Participants were then given a URL that randomly assigned them to the neuroscience (n = 103) or control (n = 98) condition. Five participants were removed for indicating that they spent little effort on the tasks in the experiment. Approximately half of the participants (n = 100 and 110, respectively) had earned a four-year college degree and reported having taken a statistics class at some point in their education. Additionally, the majority of participants (n = 158, 176, and 183, respectively) reported being at least somewhat familiar with the principles of the scientific method and the importance of random samples and the law of large numbers. Participants were compensated $0.80.

Materials and Procedure

We constructed a news-like article that introduced a type of claim often encountered in media reports of scientific studies. The article claimed that listening to music while studying was beneficial for learning and provided evidence to support that claim. Similar to Klaczysnki (2000), the evidence consisted of a study with a significant sampling error: the participants in the fictitious study self-selected themselves into the two conditions, creating a clear selection bias (see Figure 2.1.1).


DAILY SCIENCE
REPORTER

Should you listen to music while studying?
By Sam Katz, science writer
Updated 8:20 AM EST, November 3, 2011

Many people think listening to music is beneficial, and researchers have become interested in whether listening to music actually helps students' ability to pay attention and learn information. A recent study was conducted in a typical college classroom. The professor asked for volunteers to be a part of a music listening group. This group would be required to listen to music while studying or attending lecture throughout the semester; they would be encouraged to listen to as much music as much as possible. Seventy five percent of the class volunteered to be in this group. The rest of the class was required to avoid listening to any kind of music while studying. At the end of the semester, the members of the music listening group consistently had higher exam scores for each of the three exams than the members of the non-music listening group. These results suggest that music listening actually helps one's ability to study and, thus, has a positive impact on learning.

Figure 2.1.1. Research study description seen by all participants.

Participants in both conditions read the research study description. Similar to the approach used by Weisberg et al. (2008), the neuroscience condition saw the research study description preceded by the following two sentences that contained neuroscience jargon but did not provide any clear explanation for the effect described in the research study:

> Years of neuroscience research have made it clear that listening to music is associated with distinct neural processes. Functional MRI scans reveal that listening to music engages cortical areas involved in music and sound perception, and this activation is thought to be present even while doing other tasks, such as studying or learning new information.

Prior beliefs measure. Before reading the article, participants were first screened about their prior beliefs about the claim and were asked to choose one from the following list: a) listening to music has a negative impact on studying/learning; b) listening to music has no impact on studying/learning; c) listening to music has a positive impact on studying/learning; d) I have no expectation about the relationship between listening to music and studying/learning. Participants who chose option (a) or (b) were classified as having incongruent prior beliefs (n = 98) and those who chose (c) were classified as having congruent prior beliefs (n = 98). Participants who chose option (d) were not eligible for the study (n = 85). Prior beliefs were coded such that a higher score indicated congruent prior beliefs.

Evaluation measures. After reading the article, participants rated the quality of the article (1 = Very poor, 5 = Very good), the quality of the research study (1 = Very poor, 5 = Very good), and how convincing the article was as evidence of the claim (1 = Completely unconvincing, 7 = Completely convincing). Participants were also asked to justify their convincingness ratings in their own words. The first author (blind to condition) and an independent rater (blind to both condition and hypotheses) coded whether these open-ended justifications mentioned the methodological flaw. Inter-rater agreement was good (Kappa = .73, $p < .001$).

Thinking disposition measures. After reading the article, participants completed the 41-item Actively Open-Minded Thinking (AOT) scale (Stanovich & West, 1997) and the three-item Cognitive Reflection Test (CRT; Frederick, 2005). High scores on the AOT scale indicate more flexible and open-minded thinking dispositions, and high scores on the CRT reflect an ability to engage in deliberative over automatic processing.

Knowledge measure. Finally, participants completed the following series of questions measuring overall knowledge and familiarity with scientific reasoning principles: 1) Are you familiar with the general principles of the scientific method? (1 = Not at all familiar, 5 = Very familiar); 2) What is the highest grade or year of school you completed? (1 = Elementary school only, 9 = Advanced graduate work or Ph.D); 3) Are you familiar with the idea that, for the purpose of research, one must have a large enough sample size to draw generalizations about the results? (1 = Not at all familiar, 5 = Very familiar); 4) Are you familiar with the idea that, for the purpose of research, one must select a random sample of participants from the population of interest? (1 = Not at all familiar, 5 = Very familiar). To create an overall knowledge score, scientific method, sample size, random sample, and education variables were all converted to z-scores and the average was computed.

Results

Overall, participants were poor at identifying the methodological flaw, regardless of whether neuroscience information was present. Forty participants (20.4%) mentioned the methodological flaw when explaining their ratings of convincingness, and this percentage did not differ significantly between the neuroscience and control conditions (17.8% vs. 23.1%, respectively, $X^2(1, N = 196) = 0.56$, $p = .4$). Group means for the three evaluation measures can be seen in Table 2.1.1.

Table 2.1.1
Group Means for Evaluation Measures

Condition	*n*	Quality of Study	Quality of Article	Convincingness
Control	95	2.81 (1.07)	3.30 (0.90)	3.94 (1.76)
Neuroscience	101	3.13 (1.04)	3.63 (0.84)	4.24 (1.75)

Note. Standard deviations in parentheses.

In order to examine the influence of neuroscience information on reasoning, we conducted a multivariate analysis of variance (MANOVA) on the three evaluation measures. This revealed an overall effect of condition, $F(3,192) = 2.73$, $p < .05$; Hotelling's $T^2 = .04$, partial $\eta^2 = .04$. Neuroscience resulted in higher ratings of article quality, $t(194) = -2.63$, $p < .01$, $d = .37$, and ratings of study quality, $t(194) = -2.16$, $p < .05$, $d = .30$. Neuroscience did not have a significant effect on convincingness, $t(194) = -1.19$, $p = .23$.

To control for individual differences, we constructed three regression models predicting ratings of convincingness, quality of the article, and quality of the study, with knowledge, prior beliefs, AOT, and CRT entered as covariates (Table 2.1.2). Condition became a significant predictor of convincingness, study quality ratings, and article quality ratings. In all cases, the presence of neuroscience resulted in more favorable evaluations. As expected, prior beliefs significantly predicted all three types of evaluations.

Incongruent beliefs about the claim resulted in less favorable evaluations and congruent beliefs resulted in more favorable evaluations. In addition, participants with more methodological knowledge and higher scores on the AOT and CRT scales consistently gave less favorable evaluations.

Table 2.1.2
Regression Models Predicting Evaluation Measures After Controlling for Individual Differences

	Model 1			Model 2			Model 3		
	DV: Convincing			Quality of the Article			Quality of the Study		
Predictor	*B* (*se*)	*p*	f^2	*B* (*se*)	*p*	f^2	*B* (*se*)	*p*	f^2
Condition: Neuroscience	0.44 (0.22)	<.05	.02	0.36 (0.11)	<.01	.04	0.40 (0.13)	<.01	.04
Prior beliefs	1.26 (0.21)	<.001	.17	0.33 (0.11)	<.01	.04	0.39 (0.13)	<.01	.04
Methodological knowledge	-0.41 (1.16)	<.01	.03	-0.12 (0.08)	.16	.01	-0.27 (0.10)	<.01	.03
AOT	-0.46 (0.18)	<.05	.04	-0.20 (0.09)	<.05	.02	-0.40 (0.11)	<.001	.03
CRT	-0.27 (0.09)	<.01	.04	-0.11 (0.05)	<.05	.03	-0.14 (0.05)	<.05	.06

Note. Model 1 Fit: $F(5, 188) = 14.42$, $p < .001$, $R^2 = .27$. Model 2 Fit: $F(5, 188) = 6.86$, $p < .001$, $R^2 = .15$. Model 3 Fit: $F(5, 188) = 11.46$ $p < .001$, $R^2 = .23$. Cohen's f^2 is a measure of local effect size and represents the proportion of variance uniquely accounted for by one predictor, above and beyond all other predictors (Cohen, 1988). Cutoff values for small, medium, and large effects are .02, .15, and .35, respectively.

There was a significant interaction between CRT and condition for predicting convincingness ratings ($B = 0.36$, $se(B) = 0.17$, $p < .05$) and a marginal interaction between CRT and condition for predicting study quality ($B = 0.19$, $se(B) = 0.11$, $p = .07$). To investigate the nature of the interaction for convincingness ratings, we compared the influence of condition on convincingness ratings for individuals with the lowest CRT score (score of 0) and the highest CRT score (score of 3). Among individuals with the lowest CRT score, condition had no effect on ratings ($M = 4.78$, $SD = 1.41$ for control condition, $M = 4.61$, $SD = 1.72$ for neuroscience condition, $t(75) = 0.48$, $p = .63$).

However, among individuals with the highest CRT score, convincingness ratings were significantly higher for the neuroscience condition ($M = 4.0$, $SD = 1.67$) than for the control condition ($M = 3.1$, $SD = 1.33$), $t(49) = -2.02$, $p < .05$). There was no evidence of interactions between other individual difference measures (prior beliefs, knowledge, and AOT) and the presence of neuroscience (*p*'s ranged from .23 to .79).

Discussion

Experiment 2.1 showed that neuroscience information had some effect on all three of the evaluation measures after individual differences were taken into account. Individual differences in prior beliefs, methodological knowledge, and thinking dispositions were consistently significant predictors of article evaluations, and the effect sizes were small-to-moderate. Interestingly, this experiment suggested that the influence of neuroscience is relatively independent of individual differences. This was surprising, since individual differences were expected to play a significant role in the way people responded to irrelevant neuroscience information. The only case in which this was true was for convincingness ratings, where there was a significant interaction between condition and CRT score. However, our prediction was that individuals with more sophisticated thinking dispositions would be less influenced by neuroscience information and, if anything, we found some evidence of the opposite: although higher CRT scores led to lower convincingness ratings for the control condition, convincingness ratings remained relatively high in the neuroscience condition, regardless of CRT score.

Experiment 2.1 also showed that participants were relatively poor at identifying the methodological flaw, as less than a quarter cited the selection bias in their justification of their convincingness rating. Performance was equally poor for both conditions, which provides some preliminary evidence against the possibility that neuroscience simply distracts people from attending to the methodological details of a research study. We test this distraction hypothesis more directly in Experiment 2.2.

It should be noted that, consistent with a recent meta-analysis (Michael et al., 2013), the effect sizes of neuroscience were small. However, it is possible that these effect sizes were a result of our sample selection. We limited our sample to people who were already convinced one way or the other about the effect of music on studying, expecting to find that the influence of neuroscience would be exaggerated in these groups. We found no evidence that prior beliefs exacerbate the influence of neuroscience; instead, it is possible that people with strong prior beliefs are simply less influenced by additional information. As such, the effects of neuroscience may be more pronounced among participants who have no prior expectation of how listening to music should affect studying/learning. Experiment 2.2 addresses this issue by including participants with congruent, incongruent, or neutral prior beliefs about the claim.

Finally, a limitation of Experiment 2.1 was that the article in the neuroscience condition contained more words, so it is unclear whether an increase in ratings was due to the neuroscience jargon or simply because more information was present. A proper control for word count is needed to rule out this possibility. Additionally, regarding the lack of interactions between condition and individual differences, it is possible that we neglected to include other relevant individual difference measures in the model. For example, it is possible that the influence of neuroscience may be present only for people do not have adequate knowledge about the brain. We address this possibility in Experiment 2.2.

Experiment 2.2

Experiment 2.2 had four objectives: 1) eliminate the potential confound of article word count, 2) examine whether the effect of irrelevant neuroscience is greater among participants with neutral prior beliefs compared to participants with congruent and incongruent prior beliefs, 3) examine whether irrelevant neuroscience distracts people from attending to the details of the research methodology, and 4) test the possibility that irrelevant neuroscience inflates one's feeling of understanding a behavioral phenomenon.

Methods

Participants

Four hundred U.S. participants were recruited from Amazon's Mechanical Turk (188 females; median age = 31; range = 18-72). Six participants were excluded for indicating that they spent little effort on the experiment and five participants were excluded for having completed a previous version of the experiment; thus, all participants were seeing the stimuli for the first time. Approximately half of the participants (48.5%) reported that they had attained a four-year college degree. Additionally, fifty four percent of participants reported having taken at least one basic statistics or research methodology course, and most people reported being familiar with the principles of the scientific method, random sampling, and the importance of sample size (79.6%, 91.2%, and 94.0%, respectively).

Materials and Procedure

The protocol for Experiment 2.2 was similar to Experiment 2.1. Participants were randomly assigned to the neuroscience (n = 204) or control (n = 185) condition and read a research study summary. Preceding the research summary was a paragraph that either contained neuroscience jargon (neuroscience condition) or described the popularity of listening to music while studying (control condition). The additional paragraph in the control condition is presented below:

> Although some people prefer to work in silence, many people opt to listen to music while working or studying. In fact, due to the increased mobile access to music, a brief glimpse into a library or coffee shop will reveal dozens of individuals poring over their laptops and books with earphones wedged into their ears.

In contrast to the articles used in Experiment 2.1, both articles contained the same number of words (220). Participants were compensated $1.10.

Prior beliefs measure. We assessed prior beliefs at the beginning of the experiment using the same questions from Experiment 2.1. Participants were categorized as having congruent (*n* = 99), incongruent (*n* = 98), or neutral (*n* = 192) beliefs about the claim that listening to music improves studying/learning. We predicted that neuroscience would be most influential for participants with neutral prior beliefs.

Evaluation measures. In addition to rating study quality, article quality, and convincingness of the article, participants also rated the quality of the scientist ("Please rate the quality of the scientist who conducted the study described in the article") and how well they understood why music might have an influence on learning ("On a scale of 0 to 100, how well did this article help you understand WHY music may have an impact on learning/studying?"). If participants have a preference for reductionist explanations, we predicted that neuroscience language, even if irrelevant, would make them think they have a better understanding of the phenomenon. The ratings for all five dependent variables were done on a sliding scale that ranged from 0 to 100% to allow more flexibility in responses.

Participants also justified their convincingness ratings in their own words and responses were coded according to whether the methodological flaw was mentioned, in the same manner as in Experiment 2.1. Inter-rater agreement was excellent (Kappa = .94, $p < .001$).

Individual differences. We again collected measures of AOT, CRT, and methodological knowledge. Additionally, because the influence of neuroscience may depend on how much knowledge of the brain one has, we included five questions about neuroanatomy, collected from introductory psychology textbook companion websites (Morris & Maisto, 2002; Schacter, Gilbert, & Wegner, 2009). The questions used are listed in the Appendix. The brain knowledge score was computed by summing the number of questions they answered correctly ($M = 2.37$, $SD = 1.26$).

Recall measure. To test the possibility that neuroscience information distracts people from recalling the details of the study, we measured participants' free recall of the study they read about at the beginning of the experiment, following a typical protocol used in seductive details literature (see Harp & Mayer, 1997). We identified four main idea units in the study description: 1) Participants in the study volunteered/self-selected into the conditions, 2) 75% of the class was in the music listening condition, 3) The students who listened to music received higher grades, and 4) The conclusion that listening to music improves studying. Participants received a '1' for each main idea unit they were able to recall, and points were summed for a total possible score of 0-4. A primary rater (blind to the condition) coded all responses, and reliability was measured by having a second rater (blind to the hypotheses and condition), code 75% of the responses (Kappa = 0.72, $p < .001$).

Results

Overall, 34% of participants identified the methodological flaw in the research study, and this percentage did not differ between the neuroscience and control conditions (M = 32.8% and M = 37.5%, respectively, $X^2(1, N = 389) = .52, p > .4$). The mean evaluation scores for each condition can be seen in Table 2.2.1. A MANOVA on the five evaluation measures revealed a significant overall effect of condition, $F(5, 382) = 20.49$, $p < .001$; Hotelling's $T^2 = .26$, partial $\eta^2 = .21$. There were significant effects of neuroscience on ratings of scientist quality, $t(386) = -4.59$, $p < .001$, $d = .46$ and self-assessed understanding of the mechanism by which music might impact learning, $t(386) = -9.27$, $p < .001$, $d = .94$. In other words, the presence of neuroscience information increased perceived understanding of the mechanism underlying the effect. Neuroscience also had a marginal effect on ratings of article quality, $t(386) = -1.82$, $p = .06$, $d = .18$, and no effect on ratings of study quality or convincingness ($t(386) = -1.01$, $p = .31$ and $t(386) = -1.23$, $p = .21$, respectively).

Table 2.2.1
Group Means for Evaluation Measures

Condition	*n*	Quality of Study	Quality of Article	Convincing	Quality of Scientist	Understanding of Mechanism
Control	185	54.11	67.74	53.42	44.04	20.25
		(23.17)	(17.53)	(22.30)	(23.78)	(25.82)
Neuroscience	203	56.44	70.90	56.17	54.41	46.48
		(21.95)	(16.61)	(21.37)	(20.59)	(29.51)

Note. Standard deviations in parentheses.

As in Experiment 2.1, we constructed separate linear regression models predicting each of the evaluation measures and controlling for differences in AOT, CRT, prior beliefs, methodological knowledge, and brain knowledge. The pattern remained the same; the presence of neuroscience predicted ratings of article quality, scientist quality, and mechanistic understanding. In all cases, the addition of neuroscience information increased ratings, and the effect size was medium-to-large for the mechanistic understanding variable. Condition was not a significant predictor for ratings of study quality or convincingness. The full models for these measures are presented in Table 2.2.2.

Table 2.2.2
Regression Models Predicting Evaluation Measures After Controlling for Individual Differences

	Model 1			Model 2			Model 3			Model 4			Model 5		
	DV: Convincing			Quality of the Article			Quality of the Study			Quality of the Scientist			Mechanistic Understanding		
Predictor	*B* (*se*)	*p*	f^2	*B* (*se*)	*p*	f^2	*B* (*se*)	*p*	f^2	*B* (*se*)	*p*	f^2	*B* (*se*)	*p*	f^2
Condition: Neuroscience	2.66 (2.10)	.20	.00	3.08 (1.65)	.06	.01	2.20 (2.15)	.30	.00	10.29 (2.14)	<.001	.05	26.12 (2.60)	<.001	.26
Prior beliefs	4.54 (1.48)	<.01	.02	0.67 (1.17)	.56	.00	2.37 (1.52)	.12	.00	2.39 (1.51)	.11	.01	1.44 (1.85)	.43	.00
Methodological knowledge	-6.55 (1.47)	< .001	.04	-3.94 (1.16)	<.001	.02	-7.05 (1.51)	<.001	.06	-7.66 (1.50)	<.001	.07	-5.68 (1.84)	<.01	.01
AOT	-3.32 (1.76)	.06	.01	-4.55 (1.39)	<.01	.02	-5.54 (1.81)	<.01	.03	-4.69 (1.80)	.01	.02	-12.44 (2.20)	<.001	.07
CRT	-2.32 (0.88)	<.01	.01	-1.92 (0.70)	<.01	.01	-2.27 (0.91)	<.05	.02	-0.89 (0.90)	.32	.00	-2.19 (1.10)	< .05	.00
Brain knowledge	0.29 (0.83)	.72	.00	0.65 (0.66)	.32	.00	0.08 (0.86)	.91	.00	0.01 (0.85)	.98	.00	-1.01 (1.04)	.29	.00

Note. Model 1 Fit: $F(6, 378) = 9.18$, $p < .001$, $R^2 = .12$. Model 2 Fit: $F(6, 378) = 8.40$, $p < .001$, $R^2 = .11$. Model 3 Fit: $F(6, 378) = 10.12$, $p < .001$, $R^2 = .13$. Model 4 Fit: $F(6, 378) = 12.18$, $p < .001$, $R^2 = .16$. Model 5 Fit: $F(6, 378) = 29.49$, $p < .001$, $R^2 = .31$.

As expected, being more knowledgeable about methodological principles and having more sophisticated thinking dispositions (measured by AOT and CRT) resulted in less favorable evaluations for all measures. Prior beliefs were positively associated with ratings of convincingness, such that having congruent beliefs led to higher ratings. Surprisingly, knowledge of the brain did not predict evaluations. Additionally, there were no significant interactions between condition and brain knowledge, methodological knowledge, AOT, or CRT (*p*'s ranged from .17 to .88).

To provide an overall test of our hypothesis that the influence of neuroscience is larger for those with neutral prior beliefs, we first re-coded the prior beliefs measure assigning a ‘0’ to participants with neutral prior beliefs and a ‘1’ for participants with either congruent or incongruent beliefs. We ran a MANOVA on all five evaluation measures with condition, beliefs, and the beliefs x condition interaction entered as predictors and found no significant interaction between beliefs and condition, $F(5, 380) = 0.65$, $p = .66$; Hotelling’s $T^2 = .01$, partial $\eta^2 = .01$. Looking more specifically at all three prior beliefs groups, Figure 2.2.1 shows the ratings of the five evaluation measures broken down by each beliefs group – congruent, incongruent, and neutral. As Figure 2.2.1 illustrates, the effect of neuroscience is substantial for ratings of scientist quality and mechanistic understanding, regardless of prior beliefs.

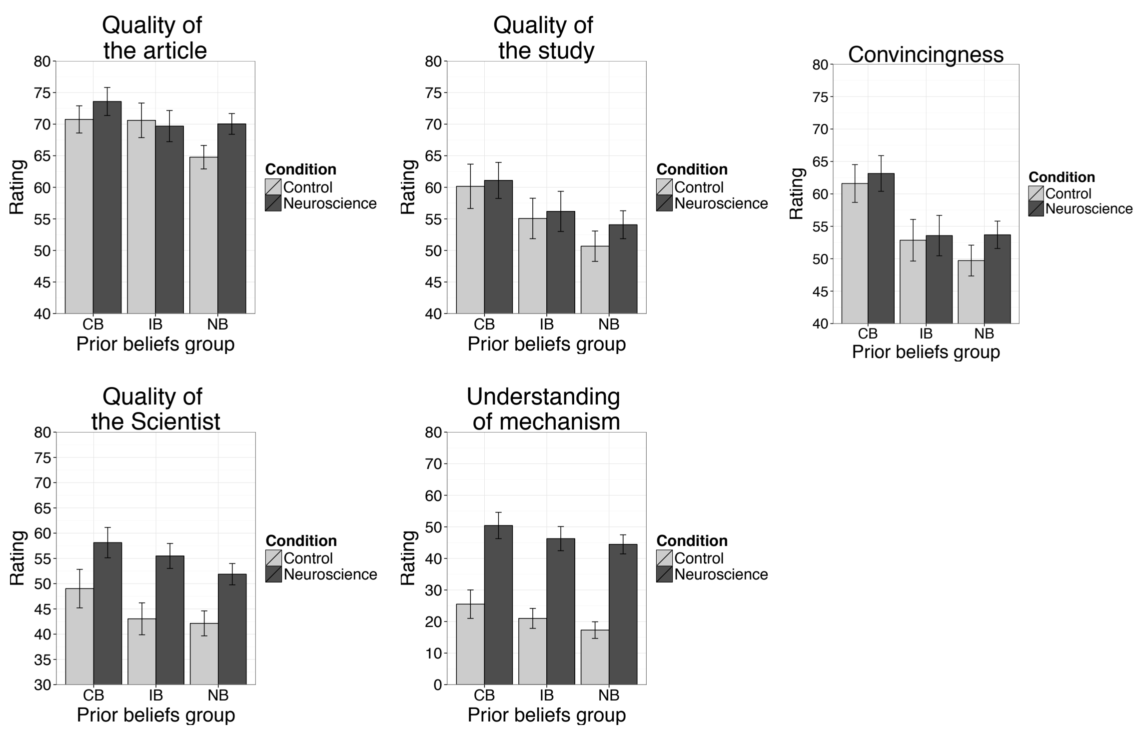


Figure 2.2.1. Means of Each Evaluation Measure, by Condition and Prior Beliefs Subgroup. CB = Congruent prior beliefs. IB = Incongruent prior beliefs. NB = Neutral prior beliefs. Bars represent standard errors of the mean.

To assess whether participants in the neuroscience condition were distracted from the methodological details of the study, we compared the recall of the methodological flaw (that participants self-selected into the study conditions) for both conditions. There was not a significant difference in the recall of the flaw between the control (M = .42, SD = .49) and neuroscience (M = .37, SD = .48) conditions, $t(386) = 1.03$, $p = .29$. We also looked more generally at the total number of main idea units recalled in both conditions, and again found no difference between the control (M = 2.04, SD = 1.16) and neuroscience (M = 1.90, SD = 1.06) conditions, $t(386) = 1.11$, $p = .26$. These results suggest that neuroscience did not interfere with participants' ability to recall the key points of the research study.

Discussion

Experiment 2.2 showed a general effect of neuroscience information, even after controlling for the number of words in the articles. In particular, neuroscience increased ratings of scientist quality by 10% and improved mechanistic understanding by 26%. Similar to Experiment 2.1, we found that individual differences such as prior beliefs, knowledge, and thinking dispositions predicted evaluations. However, we again found no evidence to suggest that one's prior beliefs about a claim (whether congruent, incongruent, or neutral) moderate the effect of neuroscience.

Somewhat surprisingly, brain knowledge did not predict evaluations. It is possible that the questions we selected to measure brain knowledge were too difficult for participants, although we reasoned that many people would have had exposure to these brain knowledge questions in any introductory biology or psychology class. Nevertheless, future studies could explore different measures of brain knowledge. Experiment 2.2 also replicated the finding from Experiment 2.1 that participants in both the neuroscience and control condition perform equally poorly at identifying the methodological flaw in the study, suggesting that neuroscience is not particularly distracting. Experiment 2.2 provided further evidence against the distraction hypothesis by showing that neuroscience did not affect the ability to recall the details of the study.

General Discussion

The present experiment investigated the influence of neuroscience information on evaluations of scientific evidence after controlling for individual differences. Overall, we found an effect of irrelevant neuroscience on evaluations that is small for subjective ratings such as article quality, study quality, and convincingness of the evidence. However, we did find moderate-to-large effects for ratings of scientist quality and mechanistic understanding.

The small effects of neuroscience information on subjective ratings are consistent with recent literature (Farah & Hook, 2013; Hook & Farah, 2013; Michael et al., 2013), but the large effect on understanding of the mechanism represents a novel finding. This self-assessed understanding of the mechanism provides insight into how irrelevant neuroscience information may influence the way people are thinking about the relevant variables in the study. Specifically, the indication that participants' understanding improved suggests that they are making causal associations between listening to music and quality of studying; as such, the fact that neuroscience information increased these ratings of understanding suggests that it may be influencing participants' understanding about causation in the experiment. Future studies should further explore this effect on causal understanding.

Interestingly, we found no evidence to suggest that the influence of neuroscience interacts significantly with individual difference measures. Consistent with research on belief-biased reasoning (Klaczynski, 2000; Lord, Ross, & Lepper, 1979), participants with congruent prior beliefs were more convinced by the evidence than participants with incongruent prior beliefs, regardless of whether neuroscience was present. We reasoned that having neutral prior beliefs would allow for more movement in the evaluations and a potentially bigger effect size of neuroscience; however, Experiment 2.2 showed that this was not the case. Additionally, although methodological knowledge and thinking dispositions often predicted evaluations, we found no evidence to suggest that having more knowledge or more sophisticated thinking styles mitigated the influence of irrelevant neuroscience information.

Another goal of the present research was to investigate potential mechanisms for the effect of neuroscience information on reasoning and critical thinking. We tested the possibility that neuroscience behaves as a seductive detail (cf. Rey, 2012), but found little evidence to suggest that neuroscience itself distracts people from paying attention to the methodological aspects of the article. A minority of participants were able to identify the methodological flaw in the article, regardless of condition; additionally, performance on the recall measure was comparable for both conditions. Instead, we found evidence to suggest that the reductionist nature of neuroscience information may mislead people into thinking they are getting more information than they actually are. Although the neuroscience information in the present experiment was not particularly relevant or informative, participants indicated that it increased their understanding of the relationship between listening to music and studying, suggesting that they do not understand the limitations of inferences made from neuroimaging evidence. It is also worth noting that this effect may not be specific to neuroscience. For example, Eriksson (2012) demonstrated that meaningless mathematical equations made journal abstracts more likely to be accepted. Taken together, these results support the notion that reductionist explanations, even if not fully understood, appear to be providing valuable information and may even convince people that they understand a phenomenon better.

Given the popularity of neuroimaging and the attention it receives in the press, it is important to understand how people are weighting this evidence and how it may or may not affect people's decisions. While the effect of neuroscience is small in cases of subjective evaluations, its effect on the mechanistic understanding of a phenomenon is compelling. Future studies should continue to examine the extent to which this is a neuroscience-specific effect or, more generally, an effect of any kind of concrete or reductionist explanation.

References

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality data. *Perspectives on Psychological Science*, *6*, 3-5.

Cohen, J. E. (1988). *Statistical power analyses for the behavioral sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Ditto, P. H., & Lopez, D. F. (1992). Motivated skepticism: Use of differential decision criteria for preferred and non preferred conclusions. *Journal of Personality and Social Psychology*, *63*(4), 568-584.

Eriksson, K. (2012). The nonsense math effect. *Judgment and Decision Making*, *7*(6), 746-749.

Evans, J. St. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences*, *7*, 454-459.

Farah, M. J., & Hook, C. J. (2013). The seductive allure of "seductive allure". *Perspectives on Psychological Science*, *8*(88), 88-90.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25-42.

Ginossar, Z., & Trope, Y. (1987). Problem solving in judgment under uncertainty. *Journal of Personality and Social Psychology*, *52*(3), 464-474.

Green, E., & Cahill, B. S. (2012). Effects of neuroimaging evidence on mock juror decision making. *Behavioral Sciences and the Law*, *30*, 280-296.

Gruber, D., & Dickerson, J. A. (2012). Persuasive images in popular science: Testing judgments of scientific reasoning and credibility. *Public Understanding of*

Science, *21*(8), 938-948.

Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology*, *90*(3), 414-434.

Hook, C. J., & Farah, M. J. (2013). Look again: Effects of brain images and mind–brain dualism on lay evaluations of research. *Journal of Cognitive Neuroscience, 25***,** 1397–1405.

Keil, F. C. (2006). Explanation and understanding. *Annual Review of Psychology*, *57*, 227-254.

Klaczynski, P. A. (2000). Motivated scientific reasoning biases, epistemological beliefs, and theory polarization: A two-process approach to adolescent cognition. *Child Development*, *71*(5), 1347-1366.

Klaczynski, P. A., & Robinson, B. (2000). Personal theories, intellectual ability, and epistemological beliefs: Adult age differences in everyday reasoning biases. *Psychology and Aging*, *15*(3), 400-416.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480-498.

Lindstrom, M. (2011, September 30). You love your iPhone. Literally. The New York Times. Retrieved from http://www.nytimes.com/pages/opinion/index.html.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, *37*(11), 2098-2109.

McCabe, D. P., & Castel, A. D. (2008). Seeing is believing: The effect of brain images on judgments of scientific reasoning. *Cognition*, *107*. 343-352.

Morris, C. G., & Maisto, A. A. (2002). *Psychology: An Introduction*. Retrieved from http://cwx.prenhall.com/bookbind/pubbooks/morris5/.

Michael, R. B., Newman, E. J., Vuorre, M., Cumming, G. & Garry, M. (2013). On the (non)persuasive power of a brain image. *Psychonomic Bulletin and Review, 20*, 720–725.

Rey, G. D. (2012). A review of research and a meta-analysis of the seductive detail effect. *Educational Research Review*, *7*, 216-237.

Sanitioso, R., & Kunda, Z. (1991). Ducking the collection of costly evidence: Motivated use of statistical heuristics. *Journal of Behavioral Decision Making*, *4*, 161-178.

Schacter, D. L., Gilbert, D. T. & Wegner, D. M. (2009). *Introducing Psychology*. Retrieved from http://bcs.worthpublishers.com/schacterbrief1e/default.asp#t_579071.

Stanovich, K. E., & West, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology*, *89*(2), 342-357.

Tallis, R. (2012). *Aping mankind: Neuromania, Darwinitis and the misrepresentation of humanity*. London: Acumen Publishing.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, *39*, 1275-1289.

Weisberg, D. S., Keil, F. C., Goodstein, J., Rawson, E., & Gray, J. R. (2008). The seductive allure of neuroscience explanations. *Journal of Cognitive Neuroscience*, *20*(3), 470-477.

Appendix

Brain Knowledge Questionnaire

Question	Source
1. Functional neuroimaging (fMRI) directly measures the: a) Neural activity of the brain during a specific task **b) Activity of oxygenated hemoglobin in the blood in the brain and body** c) Electrical activity in the brain d) Different function of the brain's two hemispheres. *Note. Both (a) and (b) were accepted as correct answers. 2. The hippocampus, amygdala, and hypothalamus are all part of the: **a) limbic system** b) brainstem c) cerebral cortex d) association areas e) somatosensory cortex	Schacter, Gilbert, & Wegner (2011)
3. The occipital lobe receives and interprets __________________ information. a) pain b) auditory **c) visual** d) bodily position 4. What structure connects the two hemispheres of the brain and coordinates their activities? a) amygdala b) reticular formation **c) corpus callosum** d) hippocampus 5. A single long fiber extending from the cell body that carries outgoing messages is called a/an: **a) axon** b) nerve c) terminal d) dendrite	Morris & Maisto (2011)

CHAPTER 3

Seeing Behavior Through the Brain: Evidence of Neurorealism

Introduction

In the media, fMRI data are commonly referenced as meaningful neural correlates of complex human behaviors (e.g., "Political Views Tied to Brain Structure", "Do Sad Movies Make You Cry? Blame Your Brain"). In many cases, these neural correlates are technically not explanatory or causal in nature. Nonetheless, neural data are frequently treated as direct proof that behavioral findings are real – a phenomenon that has been named 'neurorealism' (Racine, Bar-Ilan, & Illes, 2005). What implications does neuroscience have for lay understanding of human behavior? Do such low-level descriptions of behavioral phenomena actually have an impact on our conceptions of causality? Here we provide evidence that the presence of a neural correlate of a behavioral phenomenon increases the likelihood that people will infer a causal relationship between the behavioral variables and believe in the phenomenon.

There are a number of reasons why the public might find neuroscience data fascinating: brain scans may be interpreted as a clear window into the mind, people may prefer biological accounts of complex behavioral phenomena, and the technical vocabulary in neuroscience information may promote a feeling of intellectual fluency (Beck, 2010; Roskies, 2007; Trout, 2008). Empirical research has provided support for the idea that neuroscience information contains some value. For example, irrelevant neuroscience information can improve the quality of bad explanations of psychological phenomena (Weisberg, Keil, Goodstein, Rawson, & Gray, 2008), interacting with a brain scanner can make people believe in mind reading (Ali, Lifshitz, & Raz, 2014), and images of brains have been shown to lend more credibility to research (Keehner, Mayberry, & Fischer, 2011; McCabe & Castel, 2008; but see Farah & Hook, 2013; Gruber & Dickerson, 2012; Hook & Farah, 2013; Michael, Newman, Vuorre, Cumming, & Garry, 2013). Laypeople typically encounter neural data as simplistic explanations of seemingly complex, mystifying behavioral phenomena such as love, sex, and political orientation. To what extent does this reductionist framing have implications for lay understanding of human behavior and causal inference? Previous research has yet to examine whether the presence of neural correlates can influence causal reasoning, and the present research seeks to address this question.

Why might neural correlates be erroneously interpreted as causal pathways for behavior? Correlational reasoning can be difficult due to the fact that people have a natural, and often unconscious, tendency to seek causes for events, and inferences are often based on limited information (Kuhn, Phelps, & Walters, 1985; White, 1989). Correlational reasoning may be especially difficult when the data involves a behavior and its neural correlates. People tend to seek out mechanistic information when determining whether two events or variables are causally related (Ahn, Kalish, Medin, & Gelman, 1995; Koslowski & Masnick, 2010), and the reductionist nature of a neural correlate may give it the appearance of providing mechanistic information for the behavior in question (Keil, 2006). Consider, for example, the finding that stereotypical beliefs tend to become more pronounced when genetic explanations of group differences are provided (Bastian

& Haslam, 2006; Dar-Nimrod & Heine, 2011). One explanation for this finding could be that reductionist information, such as the suggestion of a genetic pathway, appears to provide more direct evidence of causation; as such, people may be more likely to assume that one behavioral variable (e.g., gender) determines another behavioral variable (e.g., math ability), when evidence is provided at a genetic, as opposed to a psychological, level. The present work tested whether a neural correlate is perceived as evidence for a causal relationship between two behavioral variables.

Overview

In four experiments, we assessed the influence of a neural correlate on perceived understanding of a phenomenon and the likelihood of making a causal inference. We compared the influence of a neural correlate to several control conditions: the absence of a correlate (Experiment 3.1), the presence of a psychological correlate (Experiments 3.2, 3.3, and 3.4), and the presence of an eye tracking correlate (Experiment 3.3). Experiment 3.4 also tested the possibility that the presence of a neural correlate is only influential for understanding phenomena that are easily related to the mind. In Experiments 3.1 through 3.3, we predicted that participants who received a neural correlate would indicate greater perceived understanding of the behavioral phenomenon and would be more likely to expect replication of that phenomenon in a future research study. In Experiment 3.4, we predicted that neural information would no longer be privileged when the study's dependent variable is non-cognitive.

Experiment 3.1

Methods

Participants

We used Amazon's Mechanical Turk to collect data from 472 participants. Participants were paid $1.10. The average time to read the news article was 88.19 seconds; 20 participants were excluded for taking less than 10 seconds or longer than 5 minutes to read the article. The average time to complete the survey was 20.72 (SD = 8.27) minutes, and we excluded participants whose time to complete the survey exceeded 2.5 standard deviations from the mean (n = 12). Finally, to ensure that participants in our sample paid attention to the task, we excluded 91 participants who answered an attention check item incorrectly, based on recommendations of Oppenheimer, Meyvis, & Davidenko (2009). Our final data set consisted of 349 participants (118 participants in control condition, 116 in the neuroscience text condition, and 115 in fMRI condition).

Procedure and Materials

Participants read a fictional news article that described the results of a behavioral research study, and we manipulated whether a neural correlate or no correlate was present (Figure 3.1.1). The fictional research study examined the effect of listening to classical music while studying on academic performance and concluded that listening to classical music had a positive effect on learning. The study contained an obvious methodological flaw in that the participants in the study self-selected themselves into the conditions; as a result, the internal validity of the study was threatened by the fact that the two groups being compared were likely unequal on other variables. The methodological flaw was included to avoid ceiling effects in the study ratings.

DAILY SCIENCE REPORTER

Should you listen to classical music while studying?
By Sam Katz, science writer
Updated 8:20 AM EST, November 3, 2011

[manipulated text]

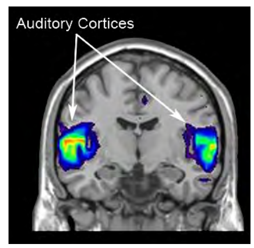
[manipulated photo]

Many people think listening to music is beneficial, and researchers have become interested in whether listening to classical music actually helps students' ability to pay attention and learn information. A recent study was conducted in a typical college classroom. The professor asked for volunteers to be a part of a music listening group. This group would be required to listen to classical music while studying or attending lecture throughout the semester; they would be encouraged to listen to as much classical music as much as possible. Seventy five percent of the class volunteered to be in this group. The rest of the class was required to avoid listening to any kind of music while studying. At the end of the semester, the members of the music listening group consistently had higher exam scores for each of the three exams than the members of the non-music listening group. These results suggest that classical music listening actually helps one's ability to study and, thus, has a positive impact on learning.

Neural Text: Years of neuroscience research have made it clear that listening to music is associated with distinct neural processes. Functional MRI scans reveal that listening to music engages cortical areas involved in music and sound perception, and this activation is thought to be present even while doing other tasks, such as studying or learning new information.

Control Text: Although some people prefer to work in silence, many people opt to listen to music while working or studying. In fact, due to the increased mobile access to music, a brief glimpse into a library or coffee shop will reveal dozens of individuals poring over their laptops and books with earphones wedged into their ears.

fMRI photo:



Control photo:

Figure 3.1.1 Research description and images used in Experiment 3.1.

The presence of a neural correlate was manipulated in a paragraph preceding the research study description. In the control condition, participants read a short description of how common it is to see people listening to music while studying in libraries and coffee shops and saw a picture of a young male student studying with headphones on. In the two neural conditions, participants read a short paragraph explaining how the auditory cortex is activated during sound perception and saw one of two photos: the picture of the young male student studying (neural text condition) or a coronal fMRI slice depicting activation in the auditory cortices (fMRI condition). We manipulated these photos in the neural conditions to test whether visual depiction of the neural correlate has an additive effect on causal inference.

Quality ratings. To assess whether neural information makes the quality of the research seem better, we asked participants to rate the quality of the study, article, and scientist. In addition, participants rated the convincingness of the research study as evidence of the positive impact of classical music on studying/learning. All ratings were on a sliding scale of 0-100%.

Mechanistic understanding. Participants were asked to rate their mechanistic understanding of the research findings on a 0-100% sliding scale ("On a scale of 0 to 100, how much did this article help you understand why classical music may have an impact on learning/studying?"). High ratings were interpreted as identification of a causal mechanism.

Expectation of replication. Participants were also asked to predict the results of a new research study testing the same claim that listening to music improves learning. Participants were told that the new study used an improved methodology – the students would be randomly assigned to the conditions. Participants could endorse one of three predictions for the new study: 1) the group who listened to classical music would receive higher exam scores; 2) the group who listened to classical music would receive lower exam scores; 3) there would be no difference in exam scores between those who did and did not listen to classical music. Responses were dichotomized; endorsement of the first option was coded as "1" and endorsement of the second or third option was coded as "0".

Individual differences. We measured individual differences in methodological knowledge, brain knowledge, prior beliefs, and thinking dispositions to examine whether these factors moderate the effect of neuroscience. Participants answered questions about methodological/statistical and brain knowledge, indicated their beliefs about the effect of listening to music on studying/learning, and completed the Actively Open Minded Thinking (AOT; Stanovich & West, 1997) scale and the Cognitive Reflection Test (CRT; Frederick, 2005).

Methodological/statistical knowledge. Knowledge of methodology and statistics was assessed with questions about familiarity with the scientific method ("Are you familiar with the general principles of the scientific method?"), issues of sample size ("Are you familiar with the idea that, for the purpose of research, one must have a large enough sample size to draw generalizations about the results?"), and random sampling ("Are you familiar with the idea that, for the purpose of research, one must select a random sample of participants from the population of interest?"). All responses were on a 5-point Likert scale (1 = Not at all familiar, 5 = Very familiar). An average, standardized knowledge score was computed.

Brain knowledge. Knowledge of the brain was assessed with five questions about basic neuroanatomy taken from introductory psychology textbooks (see Rhodes et al., 2014 for a list of questions). The brain knowledge score reflected the number of correctly answered questions.

Actively Open-Minded Thinking. The AOT scale consisted of 41 items measuring people's tendency to be flexible in their thinking. Participants indicated their agreement (1 = Disagree strongly, 6 = Agree strongly) with statements such as "A group which tolerates too much difference of opinion among its members cannot exist of long" and "Abandoning a previous belief is a sign of strong character". Higher scores on the AOT scale represent an ability to evaluate evidence objectively, independent of prior beliefs.

Cognitive Reflection Test. The Cognitive Reflection Test (CRT) measures the ability to inhibit an automatic, prepotent response in favor of a more deliberate answer. The CRT score reflected how many of the following three questions they answered correctly: "A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?", "If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?", "In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?"

Prior beliefs. Given that people are more likely to use higher order scientific reasoning with evidence that contradicts their personal theories, the effect size of neuroscience may depend on whether participants already agree with the article's claim – that listening to music while studying improves performance (Klaczynski, 2000). Before participants read the article, we asked about their prior beliefs regarding music and quality of studying ("What do you think is the relationship between listening to classical music and studying/learning?"). Participants who thought music had a positive impact on studying were categorized as having congruent prior beliefs (n = 118). Participants who thought that music had a negative impact or no impact on studying were categorized as having incongruent prior beliefs (n = 114). Participants who had no expectation of how music affects students were categorized as having neutral prior beliefs (n = 117).

Results

We found no differences between the two neural conditions on any of our dependent measures – convincingness of the study, quality of study, quality of article, quality of scientist, mechanistic understanding, or expectations of replication (ps > .50) –

so these conditions were grouped into an aggregate neural condition for all further analyses. Relative to the control condition, the presence of a neural correlate significantly increased the likelihood of expecting that the new study would replicate previous results ($B = 0.57$, $p < .05$, Table 3.1.1). 53.07% of participants who saw the neural correlate expected replication, compared to 42.24% in the control condition. The neural correlate also increased participants' understanding of the reason why listening to music improved studying, $t(347) = -7.00$, $p < .001$, $d = 0.79$. On a scale of 0-100%, participants in the control condition rated their understanding of the reason for the effect as 17.08% (*SD* = 24.76), while those in the neuroscience condition rated their understanding as 39.29% (*SD* = 29.52).

Table 3.1.1
Logistic Regression Predicting Expectation of Replication, Controlling for Individual Differences

	B	SE	p	OR (95% CI)
Condition: Neuroscience	0.57	0.26	.02	1.78 (1.06, 3.00)
Prior beliefs: Incongruent	-1.86	0.31	<.001	0.15 (0.08, 0.28)
Prior beliefs: Neutral	-1.62	0.31	<.001	0.19 (0.10, 0.35)
AOT	-0.54	0.23	.02	0.58 (0.36, 0.92)
CRT	-0.38	0.11	<.001	0.67 (0.54, 0.84)
Methodological Knowledge	-0.29	0.15	.06	0.74 (0.54, 1.01)

Note. Brain knowledge was not a significant predictor ($p > .50$) and was removed from the model. No interactions with condition were significant (*p*s > .60). $R^2 = .17$ (Hosmer-Lemeshow), .21 (Cox-Snell), .28 (Nagelkerke). Model $\chi^2(6) = 83.76$, $p < .001$.

We found that the presence of neural correlates had no effect on study quality, article quality, or convincingness of the study (*p*s > .40). There was, however, a significant effect on scientist quality. Participants in the neural condition rated the scientist significantly higher than did participants in the control condition, $M = 47.93$ (*SD* = 25.36) vs. $M = 41.42$ (*SD* = 24.09), $t(347) = -2.30$, $p < .05$. Additionally, a logistic regression revealed that perceived understanding and perceived scientist quality independently predicted expectations of replication (*B* = 0.02, *SE(B)* = 0.01, $p < .01$ and *B* = 0.04, *SE(B)* = 0.01, $p < .001$, respectively).

Given that the neural condition was associated with significantly higher ratings of perceived understanding and scientist quality than the control condition, and that perceived understanding and scientist quality predicted expectations of replication, it is possible that one or both of these dimensions mediated the effect of condition on expectations of replication. We conducted mediation analyses testing the significance of perceived understanding and scientist quality as mediators. Unstandardized indirect effects were computed for each of 10,000 bootstrapped samples, and the 95% confidence interval was computed by determining the indirect effects at the 2.5th and 97.5th percentiles. The bootstrapped unstandardized indirect effect of perceived understanding was 0.05 (CI: 0.01 – 0.11), and the bootstrapped unstandardized indirect effect of scientist quality was 0.05 (CI: 0.01 – 0.09). The total indirect effect was calculated by adding the bootstrapped indirect effects of perceived understanding and scientist quality and was statistically significant (*B* = 0.11, *SE* = 0.33, $p < .01$, *CI*: [0.04, 0.17]). The direct effect of condition was no longer significant (*B* = 0.002, *SE* = 0.05, $p = .97$, *CI*: [-0.10,0.11], indicating that perceived understanding and perceptions of scientist quality fully mediated the effect of condition.

Discussion

Consistent with previous research (Rhodes, Rodriguez, & Shah, 2014), the presence of a neural correlate increased perceived understanding and perceived scientist quality. The fact that the presence of a neural correlate also increased expectations of replication suggests that participants who saw this information were more likely to attribute a causal relationship between the two variables in the behavioral research study: music-listening behavior and academic performance. Given that perceived understanding and perceived scientist quality were both mediators of this effect, the higher likelihood of causal inference in the neuroscience condition appears to be explained by greater perceived understanding and greater trust in the scientist in this condition than in the control condition. Indeed, the only reason why the mechanistic understanding of a phenomenon should be improved is if a possible causal relationship has been identified; therefore, the fact that understanding was higher in the neural condition implies that participants were attributing a causal, as opposed to a spurious or correlational, relationship between classical music and quality of studying.

One could argue that the presence of the neural correlate in the first experiment *did* provide a plausible causal explanation for the effect of listening to music on quality of studying. For example, perhaps participants believed that the effect of brain activity is not local, and that activation anywhere in the brain could increase efficiency in other areas. If so, it would then be plausible to assume that increased brain activity in the auditory cortex could have a positive effect on areas related to attention and memory, which could improve academic performance. If this is the case, it could explain why participants in the neural condition reported having a greater feeling of understanding of the phenomenon and why they were more likely to expect the study to replicate in the future. Moreover, this effect may be elicited not just by neural information but, instead, by any type of information that could be construed as providing a causal mechanism. In our second experiment, we compared the likelihood of making a causal inference when a neural correlate was present to when a behavioral correlate was present. Both types of correlates could be interpreted as causal mechanisms for the effect of listening to music on learning. If the type of mechanism suggested (behavioral vs. neural) has no effect on causal inference generation, there should be no differences in understanding and expectations of replication between the two conditions. However, if neural mechanisms are more likely to be treated as evidence of causation, causal inference should be more likely when the neural correlate is present.

Experiment 3.2

Methods

Participants

We recruited 484 participants from Amazon's Mechanical Turk. Participants were paid $1.50. The average time to read the news article was 76.27 seconds; we excluded 58 participants who spent less than 10 seconds or greater than 5 minutes reading the article. The average time it took to complete the entire survey was 12.82 (*SD* = 6.24) minutes, and we excluded 11 participants whose time to complete the survey was greater than 2.5 standard deviations from the mean. Finally, as in Experiment 3.1, we excluded 23 participants who answered an attention check item incorrectly. After these exclusions, our data set consisted of 392 participants (193 in behavioral condition, 199 in neural condition).

Materials and Procedure

Participants read the same fictional research study as in Experiment 3.1, except the dependent variable in the study was changed to "wordless music." The presence of neural and behavioral information was manipulated between-subjects (Figure 3.2.1). The neural condition contained the same neural correlate as in Experiment 3.1, and the behavioral condition contained a paragraph explaining how listening to music can increase feelings of happiness and life satisfaction. Both conditions contained pictures of data: the neural condition contained the fMRI slice from Experiment 3.1 and the behavioral condition contained bar graphs showing that people who listen to more music score higher on a subjective happiness measure.

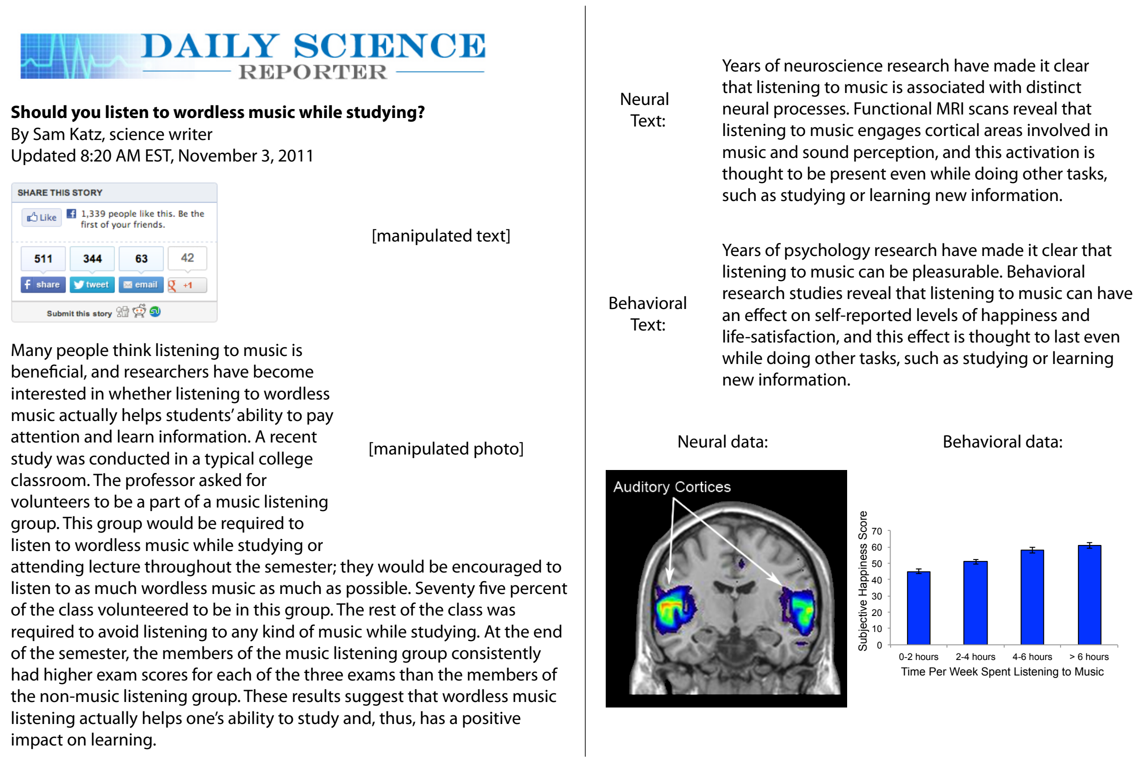


Figure 3.2.1. Research description and images used in Experiment 3.2.

Quality ratings and mechanistic understanding. After reading the article, participants rated the quality of the research and the scientist who conducted the study. Participants also rated their understanding of the reason why wordless music might have a positive impact on learning/studying. All ratings were on a 0-100% sliding scale.

Expectation of replication. As in Experiment 3.1, participants were asked about their expectation for the results of an improved study that randomly assigned students to the wordless music and no music conditions. Participants could endorse one of three options: 1) The wordless music group will score higher than the no-music group; 2) The wordless music group will score lower than the no-music group; 3) There will not be any difference in exam scores between the wordless music group and no-music group. Responses that endorsed the first option – that students in the wordless music condition would have higher exam scores – were coded as '1'; all other responses were coded as '0'.

Plausibility of causal pathways. Although the neural and behavioral correlates were not necessarily intended to be causal explanations of the music effect, they could be interpreted this way if participants believed in certain assumptions. These assumptions and their resulting causal pathways are illustrated in Figure 3.2.2. The first causal pathway is implied if participants believed that the effect of neural activation is not localized; if this is the case, then the fact that the auditory cortex is activated when people listen to music suggests that there could be increased stimulation in other areas of the brain, perhaps improving academic performance. Similarly, if participants believed that positive moods improve the quality of studying, then the fact that music is associated with more positive moods suggests that improved mood leads to better study quality, which could conceivably improve performance. We assessed the plausibility of these two causal pathways by asking participants to respond "True" or "False" to the following questions: 1) "If one specific region of the brain is activated (such as a region associated with sound perception), brain activity in other, unrelated regions (such as a region associated with memory or attention) will also increase as a result", and 2) "People tend to learn information better when they are in a happy mood."

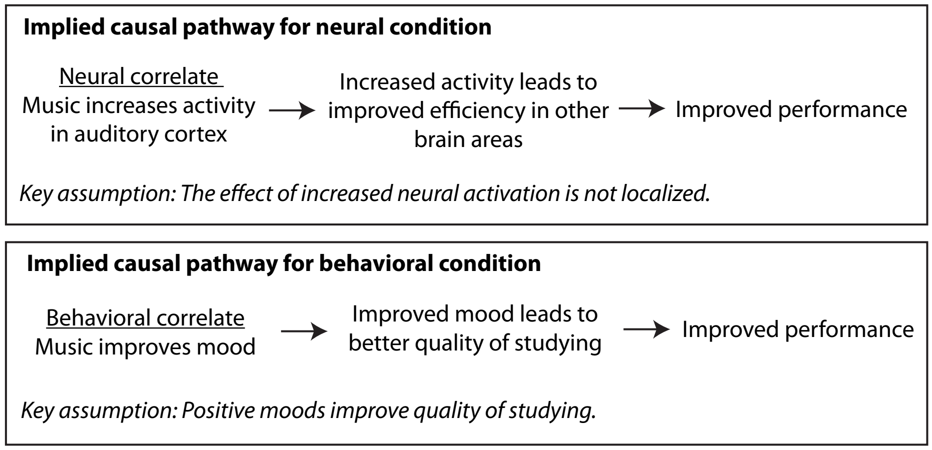


Figure 3.2.2. Causal pathways implied by neural and behavioral correlates. Although other potential pathways may exist as well, these pathways were selected as the most logical progressions based on the correlates provided.

Individual differences. We measured aspects of participants' methodological knowledge (specifically, familiarity with sample size and selection bias), as well as participants' habitual music-listening behavior to see whether these factors moderated the effect of neural information.

Music-listening behavior. Although prior beliefs did not moderate the effect of neuroscience in Experiment 3.1, we still anticipated that personal experience might be an important factor in determining how people reason about causation in the research study. As such, prior to reading the article, participants were asked how often they chose to listen to wordless music when they were working, studying, or concentrating on something specific over the last month (never, 20% of the time, 40% of the time, 60% of the time, 80% of the time, always). One hundred and sixty seven participants (42.60%) indicated that they never listened to wordless music while concentrating over the past month. These participants were categorized as “Non-listeners” and the remaining 225 participants were categorized as “Listeners”.

Understanding of sample size. Familiarity with the importance of sample size was measured after participants read the article. Participants completed an adapted, frequentist version of Kahneman and Tverksy’s hospital problem (Kahneman & Tversky, 1972; Sedlmeier & Gigerenzer, 1997). Participants who answered the item correctly received a score of 1; all other responses received a score of 0.

Understanding of selection bias. We measured understanding of selection bias by asking participants to read the following research scenario and identify the most serious threat to the validity of the radio station’s conclusions.

> Suppose a local radio station is constructing a poll about how people in a particular city currently feel about controversial political issues, such as gun control, affirmative action, and marriage equality. The station surveys the audience by asking listeners to call in and explain their position on each issue. The radio station then compiles the responses and posts the results on their website.

Participants could choose from four options: a) The study was not done in person, b) The study overrepresented individuals with strong opinions, c) The study was done only once, not multiple times, and d) The study does not take into account how listeners have previously voted. The correct answer is (b). Participants who selected this answer received a score of 1; all other responses received a score of 0.

Results

The presence of a neural correlate increased expectations of replication (B = 0.70, p < .05), but this was qualified by an interaction with habitual music listening behavior (B = -0.83, p < .05, Table 3.2.1). Among people who never listen to music while trying to concentrate (n = 167), people who read the neural version were more likely to expect replication (65.11%) than people who read the behavioral version (46.91%). Among people who habitually listen to music while working (n = 225), the neural and behavioral versions of the article resulted in similar expectations of replication (61.06% and 64.28%, respectively). Interestingly, these results suggest that people who do not already practice the behavior are most likely to interpret the neural correlate as evidence of causation; people who do practice the behavior are likely to believe in the effect regardless of whether neural information is present.

Table 3.2.1
Logistic Regression Predicting Expectation of Replication

	B	SE	p	OR (95% CI)
Condition: Neuroscience	0.70	0.32	.02	2.02 (1.08, 3.84)
Behavior: Listeners	0.71	0.30	.01	2.04 (1.13, 3.71)
Selection bias: Correct	-0.83	0.25	<.001	0.43 (0.25, 0.70)
Condition: Neuroscience X Behavior: Listeners	-0.83	0.42	.04	0.43 (0.18, 0.99)

Note. Sample size familiarity was not a significant predictor (B = -0.32, p = .12) and was removed from the model. Additionally, there were no interactions between condition and sample size or selection bias (ps > .62). R^2 = .04 (Hosmer-Lemeshow), .05 (Cox-Snell), .06 (Nagelkerke). Model $\chi^2(4) = 19.30$, $p < .001$.

We also found, as in Experiment 3.1, that participants who read the neural version rated their understanding of the phenomenon significantly higher than those who read the behavioral version (45.75% (SD = 29.75) vs. 32.32% (SD = 30.24), respectively, $t(390)$ = -4.43, $p < .001$, d = 0.44). The presence of a neural correlate had no effect on evaluations of research or scientist quality, ts < 0.50, ps > .70. Taken together with results from Experiment 3.1, we concluded that neural correlates do not have an effect on perceptions of research quality; instead, the primary influence is on causal understanding. As in Experiment 3.1, a logistic regression showed that ratings of causal understanding and scientist quality were both significant predictors of expectations of replication (B = 0.01, SE = 0.00, p < .01, and B = 0.02, SE = 0.00, p < .001, respectively).

Given that condition was a significant predictor of perceived understanding and that perceived understanding was a predictor of expectations of replication, we predicted that perceived understanding would be a significant mediator of the effect of condition on expectations of replication. However, because the effect of condition on expectations of replication was only significant for non-listeners, we performed the mediation analysis on a dataset consisting of only non-listeners. We tested the significance of the indirect effect using bootstrapping procedures. Unstandardized indirect effects were computed for each of 10,000 bootstrapped samples, and the 95% confidence interval was computed by determining the indirect effects at the 2.5th and 97.5th percentiles. The bootstrapped unstandardized indirect effect was 0.05, and the 95% confidence interval ranged from 0.01 to 0.11. Thus, the indirect effect was statistically significant. The direct effect of condition was no longer significant (B = 0.11, *CI*: [-0.04, 0.25]) indicating that perceived understanding fully mediated the effect of condition on expectations of replication for non-listeners.

An alternative explanation for the increased understanding and expectations of replication in the neural condition could be that the suggested behavioral mechanism was perceived as less plausible than the suggested neural mechanism. However, 95.66% percent of participants thought that positive moods could improve study quality and 73.72% of participants endorsed the idea that neural activation is not localized, suggesting that both of the suggested mechanisms were considered plausible. The fact that the presence of the neural correlate increased mechanistic understanding more than the presence of the behavioral correlate, despite the fact that both mechanisms were considered plausible, provides additional evidence that neural information is perceived as inherently causal.

Discussion

Experiment 3.2 replicated the results of Experiment 3.1 and provided additional evidence that the information neuroscience provides is unique and more influential for perceived understanding than is psychology information. Plausible mechanisms were provided in both conditions in the present experiment, but perceived understanding and expectations of replication were significantly greater in the neural condition than the psychology condition, particularly among participants who did not already engage in the behavior being promoted.

One explanation for the effects of condition on perceived understanding and expectations of replication could be that the neuroscience information sounded more technical and scientific than the psychology information. Most people are probably aware that neuroimaging uses sophisticated technology, and people may assume that information that is gained through more sophisticated technology has more validity and explanatory power. Experiment 3.3 tests this hypothesis by adding a third condition, which replaced the neuroscience information with information that is gathered through another sophisticated, but not brain-based, technology: eye tracking. If the effect of neuroscience is due simply to the fact that it sounds more technical, we should see analogous effects on understanding and expectations of replication for the eye tracking condition. If however, there is something unique about neuroscience information as an explanation, we should expect the neuroscience condition to report greater understanding and expectations of replication than the psychology and eye tracking conditions.

It is important to note that in Experiment 3.2, neural information was primarily influential for participants who did not already engage in the behavior that the news article promoted. One explanation for this finding may be that people who choose to listen to music while they work or study already believe in some causal effect that music has on their productivity; as such, the type of causal mechanism (behavioral vs. neural) does not make a difference on their perception of causation. It is not immediately clear why people who avoid listening to music were more influenced by the neural than behavioral information. Research on motivated reasoning finds that people who are confronted with evidence that contradicts their personal theories provide more analytic (as opposed to heuristic) evaluations of that evidence (Evans, 2003; Klaczynski, 2000). It is possible that this is also true of people who do not currently engage in a behavior; if so, the more critical evaluations of the behavioral version of the article, and the valuation of the neural information, may have been the result of analytic processes. Alternatively, it is also possible that, because they did not habitually engage in the behavior, these participants were less engaged by the article and processed the neural information heuristically. Experiments 3 and 4 seek to replicate these prior behavior findings with new scenarios and further understand the extent to which people engage heuristic or analytic processes when evaluating neuroscience information.

Experiment 3.3

The objective of Experiment 3.3 was to replicate Experiment 3.2 using a novel fictional research scenario about the benefits of meditation for academic performance and also compare the influence of neuroscience information to another technical, but not brain-based methodology: eye tracking. We also included a scientific literacy scale as an additional individual difference measure. We were again interested in the influence of neuroscience information on perceived understanding and expectations of replication. In addition, we added a question that asked participants to directly compare neuroscience, eye tracking, and psychology evidence in terms of how well each piece of evidence supported the researchers' conclusion. By including this within-subjects measure of neuroscience preference, we hoped to examine the relationship between individual differences and neuroscience preference.

Method

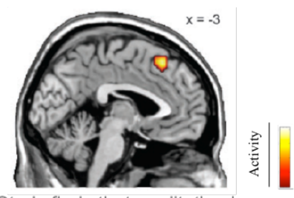
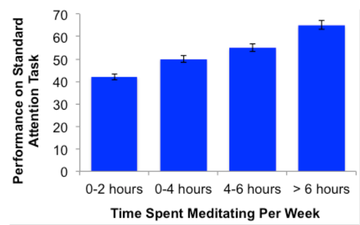
Participants

We recruited 673 participants (41.91% female, mean age = 33.99, range = 19-71) from Amazon's Mechanical Turk. To be eligible for the experiment, participants must have completed at least 100 tasks on Mechanical Turk and received an approval rating of at least 95%. More than half of the participants (57.33%) had obtained a college degree. We excluded 25 people whose total time to complete the survey exceeded the mean completion time (M = 14.97 minutes, SD = 9.74) by more than 2.5 standard deviations. To ensure the quality of our responses, we asked participants at the end of the survey how much effort they put into the experiment (1 = Almost no effort, 2 = Very little effort, 3 = Some effort, 4 = Quite a bit of effort, 5 = A lot of effort). We trimmed the data set to only include participants who indicated that they spent "Quite a bit of effort" or "A lot of effort" when completing the experiment; based on this criterion, 44 people were excluded. Research suggests that screening participants based on response time and selfreported attention is effective for removing careless responses (Meade & Craig, 2012). The final data set consisted of 604 participants. Participants were paid $1.10 for their participation.

Procedure

All participants read a brief description of a fictional research study investigating whether meditation improves academic performance. The set up of the fictional research study was similar to the music listening study in that it was conducted in a large classroom, participants volunteered to be part of either the experimental (meditation) group or control group, and participants in the meditation group showed better academic performance than did the control group at the end of the semester. Similar to the previous studies, we manipulated the paragraph preceding the research study description. This paragraph described a correlate of meditating that was derived from behavioral, neuroscience, or eye tracking methodologies. For example, the behavioral correlate was performance on standard attention tasks, the neuroscience correlate was neural activity in brain areas involved in regulating attention, and the eye tracking correlate was the ability to maintain one's gaze. Additionally, the paragraph was paired with a picture of behavioral, neuroscience, or eye tracking data, respectively. The text and images used in each condition can be seen in Table 3.3.1. Participants were randomly assigned to the psychology, neuroscience or eye tracking condition. Importantly, the same mechanism was suggested in each of the three conditions – that meditating improves academic performance because it improves attention. Thus, this design allowed us to more purely investigate the impact that the level of analysis and technical language of the suggested mechanism had on evaluation measures. The mean time spent reading the article was 1.17 minutes.

Table 3.3.1
Stimuli Used in Experiment 3.3

Text Manipulations		
Neuroscience paragraph	Years of neuroscience research have made it clear that meditating is associated with distinct neural processes. Functional MRI scans reveal that meditating regularly engages brain areas involved in processes like attention and awareness, and these effects are believed to be long lasting.	 fMRI Study finds that meditating is associated with greater neural activity in attention-related brain regions.
Psychology paragraph	Years of psychology research have made it clear that meditating is associated with distinct cognitive processes. Behavioral research studies reveal that meditating regularly can have positive effects on performance on standard attention tasks, and these effects are believed to be long lasting.	 Psychology study finds that meditating is associated with better performance on attention tasks.
Eye tracking paragraph	Years of psychology research have made it clear that meditating is associated with distinct cognitive processes. Eye tracking studies reveal that meditating regularly can improve the ability to maintain one's gaze during a task, and these effects are believed to be long lasting.	 Eye tracking study finds that meditating is associated with longer gaze maintenance, an index of visual attention.
Research Description Seen by All Participants		

New research is shedding light on whether meditating is effective for improving academic performance. In a recent study conducted in a large classroom, students volunteered to be part of a meditation group or a control group. Participants in the meditation group were instructed to meditate for 30 minutes every day for three months. Forty percent of the class volunteered to be in this group. The control group was required to avoid meditating during this time period. Researchers found that students in the meditation group showed a greater increase in their academic performance at the end of the semester than the control group. Researchers concluded that meditation can improve academic performance.

Evaluation Measures

Quality ratings. After reading the article, participants rated the quality of the research and the quality of the scientist, both on a sliding scale of 0% to 100%.

Perceived understanding. Participants rated the extent to which the article helped them understand the reason why meditation might improve academic performance, on a sliding scale of 0% to 100%.

Expectation of replication. To measure participants' causal inference, we asked participants to predict the most likely outcome of a new study that randomly assigned class members to the experimental and control groups. Participants selected among the following options: 1) The meditation group will have better academic performance than the control group, 2) The meditation group will have worse academic performance than the control group, or 3) There will not be any difference in academic performance between the meditation and control groups. Responses were dichotomized such that selecting the first option was coded as '1' and all other responses were coded as '0'. Participants also rated their confidence in their choice on a scale of 0 to 100%.

Preference for neuroscience evidence. Because participants in each condition were only exposed to one type of correlational evidence (psychology vs. neuroscience vs.

eye tracking), the design of the experiment did not allow comparisons of quality, perceived understanding, and expectations of replication within a given subject. To obtain a within-subjects measurement of the effect of the type of evidence, we presented participants with the question below. Participants selected the piece of evidence that most supported the researchers' conclusion.

> One reason why researchers think that meditation may improve academic performance is because meditation has been shown to improve attention/concentration. Suppose that different research studies have been conducted to examine how meditation influences attention/concentration. *Which piece of evidence provides the most support for the conclusion that meditation improves attention/concentration?*
>
> - Behavioral research study found that meditation improves performance on standard attention tasks used in cognitive psychology
> - Neuroimaging study found that meditation increases activity in brain areas associated with attention
> - Eye tracking study found that meditation improves one's ability to maintain his/her gaze during a task

Individual Differences

Prior behavior and beliefs. Before reading the article, participants were asked to indicate how often they have participated in some form of meditation over the past year (1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Often, 5 = All of the Time). For simplicity, selection of the first option was coded as 'Never meditate', and selection of any other option was coded as 'Sometimes meditate'. Additionally, participants were asked to indicate which of the following best described their belief about the effect of meditation on academic performance: 1) Meditating regularly can improve academic performance, 2) Meditating regularly can impair academic performance, 3) Meditating regularly has no effect on academic performance, 4) I have no opinion about the effect of meditation on academic performance. Participants who selected option 1 were categorized as having congruent beliefs, participants who selected option 2 or option 3 were categorized as having incongruent beliefs, and participants who selected option 4 were categorized as having neutral beliefs.

Methodological knowledge. Methodological knowledge was measured by familiarity with selection bias and the importance of sample size. Familiarity with these two concepts was measured in the same way as in Experiment 3.2.

Scientific literacy. Participants completed the Civic Scientific Literacy scale (Miller, 1998; see Appendix), which consisted of 11 questions measuring familiarity with basic physical and biological science. Seven items were simple true/false questions (e.g., "Lasers work by focusing sound waves"). The remaining four questions assessed understanding of basic probability and the scientific process. There was a ceiling effect on scientific literacy ($M = 0.87$, $SD = 0.13$ out of a possible 1.0, Cronbach's alpha = .47), so we combined the literacy scale with the two methodological questions to create a 12-item scale ($M = 0.83$, $SD = 0.14$, Cronbach's alpha = 0.54). As expected, scientific literacy was significantly and positively correlated with education level ($r(600)= .28$, $p < .001$).

Results

We ran a MANOVA with the perceived quality and understanding ratings as the dependent variables and condition, prior beliefs, prior behavior, and scientific literacy as predictors. Results of the MANOVA revealed significant overall effects of condition ($F(6, 1150) = 8.83$, $p < .001$, Hotelling-Lawley = 0.09) on perceived understanding but not on ratings of scientist and research quality (ps > .50, Figure 3.3.1). Post-hoc analyses using Tukey's HSD on the perceived understanding variable revealed significant effects of both the neuroscience and eye tracking conditions. Interestingly, there was greater perceived understanding in the neuroscience (M = 48.78, SD = 29.69) and eye tracking (M = 45.39, SD = 30.83) conditions than in the behavioral condition (M = 31.98, SD = 29.96, ps < .001)). The effect size for the neuroscience condition (d = .56) was greater than the effect size for the eye tracking condition (d = .44).

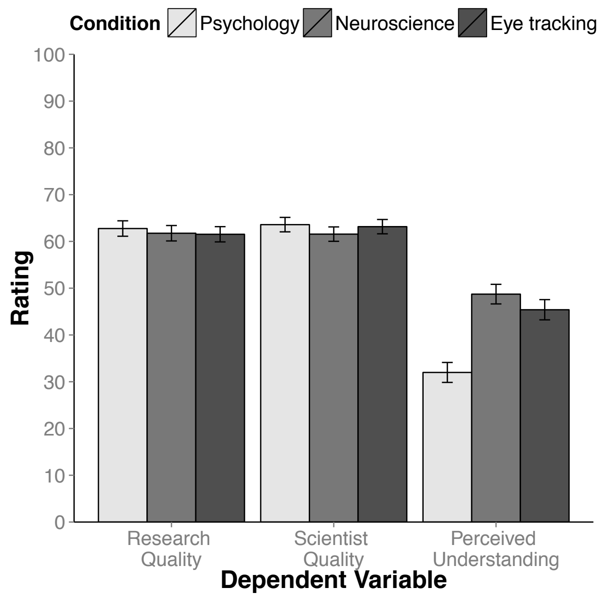


Figure 3.3.1. Effect of condition on reasoning measures.

Results from the MANOVA also revealed significant effects of prior behavior, prior beliefs, and scientific literacy on perceived understanding (Fs > 3.17, ps < .05) as well as ratings of research and scientist quality (Fs > 3.20, ps < .05). Specifically, participants who sometimes meditated reported having slightly greater perceived understanding of the phenomenon than participants who never meditated ($M_{\text{Sometimes}}$ = 55.53, $SD_{\text{Sometimes}}$ = 30.80, M_{Never} = 38.23, SD_{Never} = 30.91, p = .05), participants who had incongruent prior beliefs about meditation reported having a lower perceived understanding than participants with congruent prior beliefs ($M_{\text{Incongruent}} = 31.55$, $SD_{\text{Incongruent}} = 28.53$, $M_{\text{Congruent}} = 45.33$, $SD_{\text{Congruent}} = 31.25$, $p = .01$), and participants with lower than median scientific literacy reported having greater perceived understanding of the phenomenon than did participants with higher than median scientific literacy ($\text{M}_{\text{Low}} = 53.96$, $\text{SD}_{\text{Low}} = 32.16$, $\text{M}_{\text{High}} = 36.68$, $\text{SD}_{\text{High}} = 28.89$, $p < .001$). Prior behavior, prior beliefs, and scientific literacy had similar effects on the scientist and research quality ratings: participants who sometimes meditated reported higher scientist and research quality than did participants who never meditated (ps $< .01$), participants with incongruent beliefs rated the scientist and research quality lower than did participants with congruent beliefs (ps $< .05$), participants with neutral beliefs rated the research quality lower than participants with congruent beliefs ($p < .01$), and participants with lower than median scientific literacy rated the scientist and research more highly than did participants with higher than median scientific literacy (ps $< .001$).

To examine expectations of replication, we first excluded two participants who said they had zero confidence in their prediction. Aside from these two participants, overall people were fairly confident in their prediction; the mean confidence for those who expected replication was 69% ($SD = 22.24$) and the mean confidence for those who did not expect replication was 59% ($SD = 20.50$). We conducted a logistic regression with condition (neuroscience vs. eye tracking vs. psychology), prior beliefs (incongruent vs. congruent vs. neutral), prior behavior (never vs. sometimes), perceived understanding, and scientific literacy as predictors and found that the only significant effects were a main effect of prior beliefs (participants with neutral or incongruent prior beliefs were less likely to expect replication than participants with congruent beliefs, *p*s < .001) and a main effect of perceived understanding ($p < .001$).

Although there were not significant interactions between condition and prior beliefs or prior behavior, there appeared to be a trend in the predicted direction for prior behavior – participants who never meditated were slightly more likely to expect replication when neuroscience information was present rather than psychology or eye tracking information (*p*s = .17 and .18, respectively). We repeated the logistic regression and combined the psychology and eye tracking conditions into a single non-neuroscience condition to increase power (Table 3.3.2). This analysis revealed a significant main effect of condition ($p = .02$) and a marginally significant interaction between condition and prior behavior ($p = .09$), in addition to the main effects of prior beliefs and perceived understanding.

Table 3.3.2
Logistic Regression Predicting Expectations of Replication

	B (SE)	p	OR (CI)
Condition: Non-neuroscience	0.56 (0.24)	.02	1.75 (1.08, 2.86)
Prior Behavior: Never	0.46 (0.32)	.15	1.59 (0.83, 3.05)
Prior Beliefs: Incongruent	-1.48 (0.33)	.00	0.22 (0.11, 0.43)
Prior Beliefs: Neutral	-1.11 (0.20)	.00	0.32 (0.21, 0.48)
Perceived Understanding	0.01 (0.00)	.00	1.01 (1.00, 1.01)
Condition: Non-neuroscience X Prior Behavior: Never	-0.65 (0.39)	.09	0.51 (0.23, 1.11)

Note. $R^2 = .08$ (Hosmer-Lemeshow), .10 (Cox-Snell), .14 (Nagelkerke). Model $\chi^2(6) = 68.32$, $p < .001$. The effect of scientific literacy was not significant ($p = .55$) and was removed from the model.

Figure 3.3.2 illustrates that, contrary to Experiment 3.2, the interaction seems to be primarily driven by people who sometimes meditate – these individuals are actually more likely to expect replication when there is *no* neuroscience information present. We examined this interaction further and found that people who indicated that they sometimes meditate had a significantly higher education level than participants who said they never meditate ($p < .05$). Although the interaction between education and condition did not reach significance ($p = .12$), the fact that people who sometimes meditate were significantly different from people who never meditate in terms of education level suggests that the influence of prior behavior may be confounded by other individual differences related to education level that were not assessed.

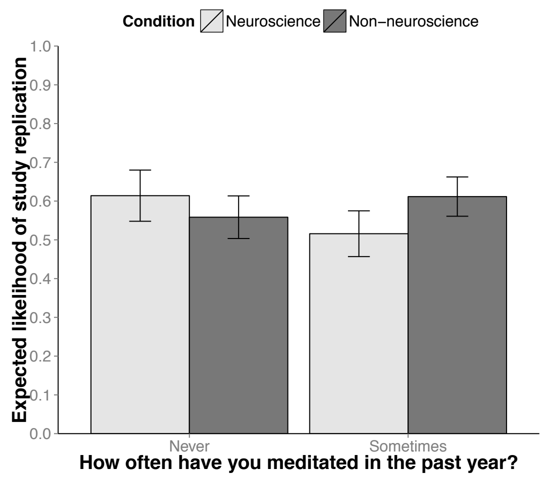


Figure 3.3.2. Predicted probability of expecting replication.

As Figure 3.3.3 illustrates, when asked which piece of evidence most supports the researchers' conclusion, participants were significantly influenced by the condition they were in. Individuals in the eye tracking condition were the most likely condition to select eye tracking evidence, individuals in the psychology condition were the most likely condition to select behavioral evidence, and individuals in the neuroscience condition were the most likely condition to select neuroscience evidence. These results suggest that the type of mechanism participants read in the article primed them to select the piece of evidence consistent with that mechanism. To examine whether there was a preference for neuroscience information while controlling for the possible influence of priming, we performed exact binomial tests on the eye tracking and psychology conditions, in each case excluding participants who chose the eye tracking and psychology evidence, respectively. For the eye tracking condition, participants were significantly more likely to choose neuroscience than psychology evidence, after those who chose eye-tracking evidence were removed (41.17% vs. 58.82%, $p < .001$). Similarly, for the psychology condition, participants were significantly more likely to choose neuroscience than eye tracking evidence, after those who chose psychology evidence were removed (90.91% vs. 9.09%, $p < .001$). Interestingly, in the neuroscience condition, participants were significantly more likely to choose the psychology evidence than the eye tracking evidence, after those who chose neuroscience evidence were removed (97.33% vs. 2.67%, $p < .001$). This latter finding suggests that the technical language present in the eye tracking evidence is not sufficient to make this evidence preferred over psychology evidence, and the finding from the psychology condition similarly shows that neuroscience is selected far more frequently than eye tracking evidence.

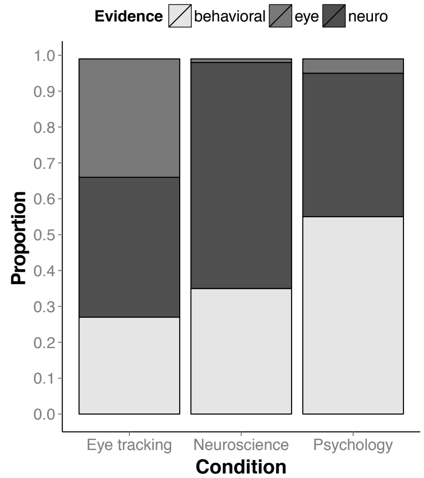


Figure 3.3.3. Within-subjects preference for neuroscience.

We also examined correlations between the choice of neuroscience evidence in the *non-neuroscience* conditions (to control for the influence of priming) and scientific literacy. People who were more scientifically literate were slightly more likely to select the neuroscience evidence over the eye tracking or psychology evidence, but this relationship did not reach significance, $r(398) = 0.08$, $p = .08$.

Discussion

We replicated our previous finding that neuroscience information, compared to psychology information, increased perceived understanding. Interestingly, we found a similar effect when comparing eye tracking information to psychology information. These results suggest that technical-sounding language may sound more mechanistic, which could increase perceived understanding. It is worth noting, however, that the effect size was larger for neuroscience information than eye tracking information, suggesting that brain-based explanations may still be perceived as providing more explanatory power.

Our finding that neuroscience evidence was perceived to provide greater support for the conclusion further suggests that there is something special about neuroscience information. The evidential support that neuroscience appears to provide cannot be purely accounted for by the technical language, since the eye tracking evidence was not perceived to be as supportive of the researcher's conclusion. Possibly, people are more familiar with the topic of neuroscience than eye tracking since it receives more attention in popular media. Alternatively, considering that the outcome variable in the research study was academic performance, which is likely perceived as a cognitive construct, people may believe that neuroimaging evidence is the best possible evidence for understanding what is going on in the mind. Fernandez-Duque, Evans, Christian, and Hodges (in press) described this as the "brain-as-engine-of-mind" hypothesis as opposed to the "prestige of science" hypothesis; that is, our findings may best be accounted for by the belief that the brain is the best explanation for mental phenomena, rather than the belief that brain imaging indicates better science. If this is the case, the marginally significant correlation between scientific literacy and frequency of selecting neuroscience evidence as most supportive suggests that people who are more scientifically literate – who presumably may have had more exposure to neuroscience evidence in the media – are slightly more likely to hold this "brain-as-engine-of-mind" belief.

An interesting question that follows is whether the brain would still be considered the best piece of evidence if the fictional study's dependent variable is no longer cognitive. Plenty of neuroimaging studies use brain evidence to help explain phenomena that are not explicitly cognitive in nature, such as those relating to self-regulation, emotion, and personality, to name just a few. It is possible that brain evidence may not appear as relevant for phenomena that are not explicitly cognitive in nature and, if this is the case, neuroscience information may no longer increase perceived understanding or appear to be superior evidence compared to behavioral information. Experiment 3.4 tests this possibility.

Finally, our failure to find a significant interaction between condition and prior behavior is interesting and could be due to the fact that meditation is not a common activity, and people who meditate may be different from people who do not meditate in ways that we did not measure. In a recent representative survey, only about 10 percent of people used meditation in the past year (National Health Interview Survey, 2008). In our sample, about 60 percent of people said they had meditated during the past 12 months (30.2% rarely, 21.35% sometimes, 7.61% often, 2.81% all of the time), which suggests that our sample is not representative, and our predicted influence of prior behavior may be more likely among a representative sample. Additionally, as mentioned previously, participants in our sample who meditated were significantly higher in education level than people who did not meditate. This difference in education level could have influenced the extent to which (and the direction in which) prior behavior interacts with condition. Experiment 3.4 controls for this possibility by using a different research context that we expected to be less dependent on education level and more applicable to a variety of people.

Experiment 3.4

In this experiment, we manipulated the presence of neuroscience and psychology information in the same way as in Experiments 3.1 and 3.2. However, in order to test whether the influence of neuroscience is limited to cognitive phenomena, the main dependent variable in the fictional research study was loneliness, a non-cognitive construct. Additionally, the fictional research study described a popular activity – eating comfort foods – that should be less associated with moderating factors such as education.

Method

Participants

We recruited 310 participants (46.13% female; mean age = 33.86) from Amazon's Mechanical Turk. To be eligible to complete the HIT, participants must have completed at least 100 tasks on Amazon Mechanical Turk and received at least a 95% lifetime approval rating. More than half of the participants (59.66%) had obtained a bachelor's degree. Similar to Experiment 3.3, we excluded eight participants whose time to complete the survey exceeded 2.5 standard deviations of the mean completion time (15.25 minutes, *SD* = 11.06). We also excluded five people who said they spent less than "quite a bit of effort" on the experiment and 14 people who incorrectly answered a comprehension question after reading the article. The final data set consisted of 273 participants. Participants were compensated $1.10 for their participation.

Procedure

Participants read a fictional research study that found that eating comfort foods alleviated feelings of loneliness. Researchers in the fictional study induced loneliness in the participants and asked them to volunteer to be in either a comfort food or regular food condition (selection bias). Those who volunteered to be in the comfort food condition could choose to eat pizza, ice cream, or cookies. The fictional study found that participants in the comfort food condition had a greater reduction in loneliness; therefore, the researchers concluded that eating comfort foods reduces loneliness.

Prior to reading this fictional research study, participants were randomly assigned to read a paragraph that contained either psychology or neuroscience details related to comfort food. The psychology version of the paragraph stated that eating comfort foods tends to bring up fond memories of one's childhood, such as family traditions or other special social events. The neuroscience version of the paragraph stated that eating comfort foods activates brain areas associated with remembering positive socioemotional events. Importantly, the mechanism suggested in both conditions was that comfort food makes one remember fond memories and/or events; the two mechanisms differed only in their level of analysis. The mean time spent reading the article was 1.42 minutes.

Table 3.4.1
Stimuli Used in Experiment 3.4

Neuroscience Paragraph	**Psychology Paragraph**
Years of neuroscience research suggest that certain foods can affect the kinds of things people think about. Specifically, brain scans show that when people view images of their favorite comfort foods, they show increased activation in brain areas (such as the medial temporal lobe) associated with remembering positive socioemotional memories. (50 words)	Years of psychology research suggest that certain foods can affect the kinds of things people think about. Specifically, psychology studies show that when people view images of their favorite foods, they tend to think about memories from their childhood such as family traditions and other significant social events. (48 words)
Research Description Seen by All Participants	

New research is shedding light on how comfort foods might cure feelings of loneliness. In a recent study, participants volunteered to be part of a comfort food group or a control group. At the beginning of the study, researchers made participants feel lonely by asking them to write about the last fight they had with a close friend or relative. Then, participants in the comfort food group were instructed to eat their favorite comfort food, choosing from pizza, ice cream, or cookies. Participants in the control group were instructed to eat a granola bar. Loneliness was measured before and after participants ate their respective foods. Researchers found that participants in the comfort food group showed a greater decrease in loneliness than participants in the control group. Researchers concluded that comfort food can reduce loneliness.

Evaluation Measures

Quality ratings and perceived understanding. After reading the article, participants rated the overall quality of the research and the quality of the scientist on scales of 0% to 100%. Participants also rated the extent to which the article helped them understand why comfort foods might reduce loneliness on a scale of 0% to 100%.

Expectation of replication. Participants were asked to indicate what would happen if researchers repeated the study using an improved methodology. With the improved methodology, participants in the fictional study would be randomly assigned to the comfort food group or control group, and participants' could choose to eat any comfort food they desired, rather than having to choose from just pizza, ice cream, or cookies. Participants could respond in the following ways: 1) The comfort food condition would show a greater reduction in loneliness than would the control condition, 2) The control condition would show a greater reduction in loneliness than would the comfort food condition, or 3) The comfort food condition and control condition would show equal reductions in loneliness. Participants who selected the first option were categorized as having expectations of replication (coded as '1'), while participants who selected the second or third option were categorized as not having expectations of replication (coded as ‘0’).

Preference for neuroscience evidence. We followed a similar approach as in Experiment 3.3. Participants were told that one reason researchers think comfort foods reduce loneliness is because comfort foods are associated with fond, sentimental memories. We then asked which piece of evidence most supports the conclusion that comfort foods increase the recall of fond, sentimental memories: 1) Neuroimaging study found that comfort foods were associated with increased activity in brain areas (such as the medial temporal lobe) associated with storing sentimental memories, or 2) Behavioral research study found that people who ate comfort foods were more likely to report thinking about sentimental memories than people who ate foods they enjoyed but did not associate with comfort.

Individual Differences

Prior behavior and beliefs. Prior to reading the fictional study, participants were asked to think about all the times they have felt lonely or sad over the past two months and indicate how often they ate comfort food(s) to help themselves feel better (never, 20% of the time, 40% of the time, 60% of the time, 80% of the time, or always). Responses were dichotomized such that ‘0’ indicated that participants never ate comfort food and ‘1’ indicated that participants did turn to comfort foods to help themselves feel better. To measure prior beliefs, participants were also asked to indicate which of the following best explains the relationship between eating comfort foods and loneliness: 1) Eating comfort foods tends to reduce loneliness, 2) Eating comfort foods tends to increase loneliness, 3) Eating comfort foods does not affect loneliness, and 4) I have no expectation about the relationship between comfort foods and loneliness. Participants who chose the first option were categorized as having congruent beliefs, participants who chose the second or third options were categorized as having incongruent beliefs, and participants who chose the fourth option were categorized as having neutral beliefs.

Methodological knowledge. Methodological knowledge was measured by familiarity with selection bias and the importance of sample size. Familiarity with these two concepts was measured in the same way as in Experiments 3.2 and 3.3.

Scientific literacy. Participants completed the same Civic Scientific Literacy scale (Miller, 1998) used in Experiment 3.3. Similar to Experiment 3.3, we combined the questions from the Civic Scientific Literacy Scale with two methodological questions to form a 12-item science literacy measure (M = 0.83, SD = 0.13, Cronbach's alpha = 0.49). As expected, scientific literacy was significantly and positively correlated with education level ($r(271) = 0.31$, $p < .001$).

Results

We conducted a MANOVA on the three rating scale variables with condition, prior beliefs, prior behavior, and scientific literacy as predictors. Interestingly, condition was not a significant predictor of any rating scale variable (Fs < 1.84, ps > .17). The lack of an effect on perceived understanding suggests that the contribution of neuroimaging evidence to perceived understanding might be limited to outcomes that are cognitive or explicitly related to the mind. There were, however, significant main effects of prior beliefs and scientific literacy (Fs > 4.91, ps < .01) as well as a marginally significant effect of prior behavior ($F(3, 253) = 2.16$, $p = .09$), and no significant interactions (ps > .40). One-way ANOVAs revealed significant effects of prior beliefs and scientific literacy on all three ratings: scientist quality, research quality, and perceived understanding (*F*s > 5.7, *p*s < .05). Post-hoc tests showed that participants who had lower than median scientific literacy rated the research quality, scientist quality, and perceived understanding higher than participants who had higher than median scientific literacy (*p*s < .05). Additionally, participants who had congruent prior or neutral prior beliefs rated the scientist and research quality lower than participants with congruent prior beliefs (*p*s < .01), and participants with neutral prior beliefs rated their perceived understanding lower than participants with congruent prior beliefs (*p* < .05).

As expected, and contrary to Experiment 3.3, there was no difference in education level between people who ate comfort food and people who didn't eat comfort food (*p* > .60). We examined participants' expectations of replication in a logistic regression (Table 3.4.2) and found, similar to Experiments 3.1 and 3.2, a significant main effect of prior beliefs (*p* < .001), marginally significant main effects of prior behavior (*p* = .08) and scientific literacy (*p* = .09), and a significant interaction between condition and prior behavior (*p* < .05).

Table 3.4.2
Logistic Regression Predicting Expectations of Replication

	B (SE)	*p*	*OR (CI)*
Condition: Neuroscience	-0.05 (0.32)	.85	0.94 (0.49, 1.79)
Prior Behavior: Never	-0.78 (0.42)	.06	0.45 (0.19, 1.06)
Prior Beliefs: Incongruent	-1.60 (0.34)	.00	0.20 (0.09, 0.39)
Prior Beliefs: Neutral	-1.61 (0.38)	.00	0.19 (0.09, 0.41)
Scientific Literacy	2.24 (1.13)	.04	9.43 (1.01, 90.16)
Perceived Understanding	0.01 (0.00)	.04	1.01 (1.00, 1.02)
Condition: Neuroscience X Prior Behavior: Never	1.32 (0.66)	.04	3.74 (1.04, 14.16)

Note. R^2 = .13 (Hosmer-Lemeshow), .15 (Cox-Snell), .21 (Nagelkerke). Model $\chi^2(7)$ = 44.98, *p* < .001.

Specifically, the presence of neuroscience information increased expectations of replication among participants who did not already engage in the behavior being promoted, but not among participants who did already engage in the behavior (Figure 3.4.1).

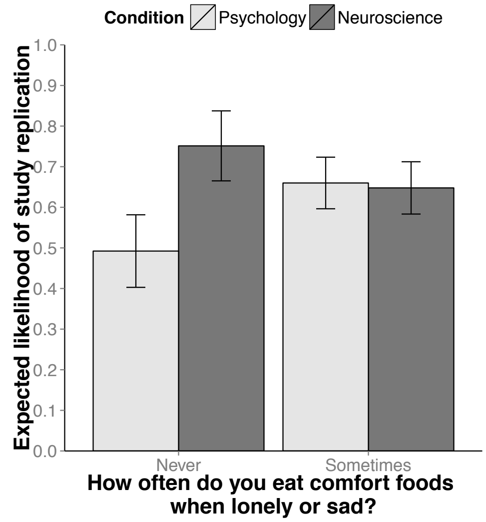


Figure 3.4.1. Predicted probability of expecting replication, controlling for scientific literacy and perceived understanding.

Similar to Experiment 3.3, participants indicated that neuroscience evidence provided better support for the researchers' conclusion than psychology evidence (71.78% vs. 28.21%). Interestingly, participants heavily favored neuroscience over psychology information regardless of the condition they were in (χ^2 (1, N = 273)= 0.0, p = 1). Additionally, we examined the correlation between scientific literacy and the selection of neuroscience in the psychology condition (to control for the influence of priming) and found that participants who were more scientifically literate were more likely to think that neuroscience evidence provided greater support for the conclusion, r(141) = 0.17, p = .03.

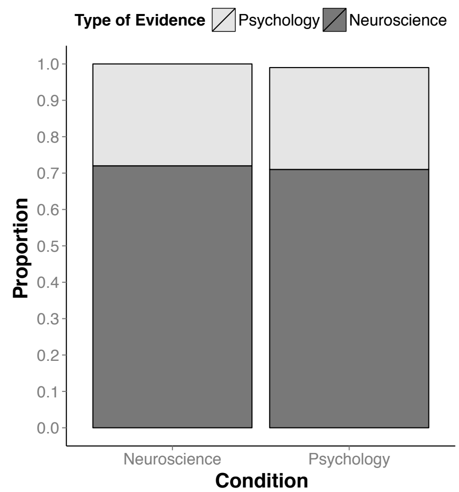


Figure 3.4.2. Within-subjects measure of neuroscience preference.

Discussion

The results from Experiment 3.4 suggest that the dependent variable being assessed in a neuroimaging study may affect the influence of that information; specifically, neuroimaging evidence may be most likely to increase perceived understanding of a phenomenon when that phenomenon is explicitly cognitive. These results provide additional support to the "brain-as-engine-of-the-mind" hypothesis (Fernandez-Duque et al., in press) and may shed light on why a recent experiment investigating the influence of neuroscience information found that it did not influence assessments of personality, a non-cognitive construct (Diekmann, Konig, & Alles, in press).

As expected, Experiment 3.4 also replicated the findings from Experiments 3.1 and 3.2 related to prior behavior – the influence of neuroscience information seems to be primarily influential for participants who do not yet engage in the behavior. People who already engage in the behavior being promoted likely already believe that the conclusion is true and are likely to think the findings would replicate regardless of which condition they are in.

We also replicated the findings from Experiment 3.3 that people high in scientific literacy are less likely to inflate their understanding of the phenomenon and tend to give lower ratings of research quality and scientist quality. However, given the choice between neuroscience and psychology evidence, highly scientifically literate people are more likely to think that neuroscience evidence provides better support for the conclusion than psychology evidence. It is possible that people who have higher scientific literacy may be more aware of the shortcomings of psychological research. Regardless, this finding suggests that the preference for neuroscience information might occur not through a heuristic process but instead through deliberate valuation.

General Discussion

In four experiments, we provide evidence that the presence of a neural correlate of a behavioral effect can make participants more likely to infer a causal relationship between the two behavioral variables. Specifically, participants tended to think that they had a greater perceived understanding of the mechanism underlying a cognitive phenomenon, and they were more likely to expect that the phenomenon would replicate in a future study, when a neural pathway was suggested. Similar effects on perceived understanding were found for information about another technical, but not brain-based methodology, suggesting that the technical language in an explanation may increase its credibility. Importantly, the effects of expectations of replication tended to be driven by participants who did not engage in the behavior being promoted. Interestingly, the presence of neural information had no effect on participants' ratings of how well the research was conducted or how convincing it was.

Recent research has investigated the effect of brain images on perceptions of scientific quality, and the majority of the evidence suggests that brain images do not necessarily affect the perceived scientific quality of a research study (Farah & Hook, 2013; Gruber & Dickerson, 2012; Hook & Farah, 2013; Michael, Newman, Vuorre, Cumming, & Garry, 2013). Our work suggests that textual neural information may not necessarily affect the perceived quality of a research study, either; it can, however, influence the perceived validity of the evidence and peoples' perceived understanding of the phenomenon.

One explanation for these findings could relate to the fact that lay people have limited knowledge of neuroscience and gather much of their neuroscience knowledge from the media (Hurculano-Houzel, 2002). Given that media reports of neuroscience tend to be optimistic in tone and focus more on the benefits of neuroscience research rather than its limitations and challenges (Racine, Waldman, Rosenberg, & Illes, 2010), participants' limited knowledge about neuroscience may result in increased trust in the science rather than increased skepticism. This could also help explain the positive correlations between scientific literacy and the tendency to think that neuroscience evidence provided the best support for a conclusion. People who are higher in scientific literacy may have encountered neuroscience information more frequently than people who are lower in scientific literacy, which could result in a familiarity effect (Zajonc, 1968). Moreover, if the primary source of participants' familiarity with neuroscience is through the media, which tends to treat neuroscience with enthusiasm and optimism, they are unlikely to have as deep an understanding of the limitations of neuroscience research as they might for psychology research.

That neural data may be more likely to elicit causal inference for cognitive phenomena than behavioral data demonstrates how influential this evidence can be for lay understanding of the mind. For example, perhaps people would be more willing to spend money on a cognitive training program when this training program is explicitly tied to the brain. Research already suggests that this may be the case (Lindell & Kidd, 2013). Given the widespread public interest in neuroscience and the implications this field has for understanding social and political issues such as mental health and human agency (O'Connor & Joffe, 2013; Lavazza & De Caro, 2010), as well as the fact that products allegedly based on neuroscience are marketed towards clinical populations before robust empirical support is available (Chancellor & Chatterjee, 2011), communicating neuroscience research in a way that allows accurate lay interpretation is paramount. Further evidence for this comes from a recent call for increased empirical research on neuroscience communication (Illes et al., 2010). Judy Illes and colleagues outline a multipronged approach to increase neuroscience literacy, which could involve, for example, training science journalists on how to communicate neuroscience findings and rewarding academicians for communicating their science to the public. Coordination between scientists and journalists could reduce unwarranted causal interpretations and help prevent the lay public from uncritically assuming that neural data always increase the validity of behavioral phenomena.

References

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

Ali, S. S., Lifshitz, M., & Raz, A. (2014). Empirical neuroenchantment: From reading minds to thinking critically. *Frontiers in Human Neuroscience, 8*, 1-4.

Bastian, B., & Haslam, N. (2006). Psychological essentialism and stereotype endorsement. *Journal of Experimental Social Psychology*, *42*, 228-235.

Beck, D. (2010). The appeal of the brain in the popular press. *Perspectives in Psychological Science, 5*, 762-766.

Chancellor, B., & Chatterjee, A. (2011). Brain branding: When neuroscience and commerce collide. *AJOB Neuroscience*, *2*, 18-27.

Dar-Nimrod, I., & Heine, S. J. (2011). Genetic essentialism: On the deceptive determinism of DNA. *Psychological Bulletin*, *137*, 800-818.

Diekmann, J., Konig, C. J., & Alles, J. (in press). The role of neuroscience information in choosing a personality test: Not as seductive as expected. *International Journal of Selection and Assessment*.

Evans, J. S. B. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences*, *7*, 454-459.

Farah, M. J., & Hook, C. J. (2013). The seductive allure of “seductive allure”. *Perspectives on Psychological Science*, *8*, 88-90.

Fernandez-Duque, D., Evans, J., Colton, C., & Hodges, S. D. (in press). Superfluous neuroscience information makes explanations of psychological phenomena more appealing. *Journal of Cognitive Neuroscience*.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*, 25-42.

Gruber, D., & Dickerson, J. A. (2012). Persuasive images in popular science: Testing judgments of scientific reasoning and credibility. *Public Understanding of Science*, *21*, 938-948.

Hurculano-Houzel, S. (2002). Do you know your brain? A survey on public neuroscience literacy at the closing of the decade of the brain. *Neuroscientist*, *8*, 98-110.

Hook, C. J., & Farah, M. J. (2013). Look again: Effects of brain images and mind-brain dualism on lay evaluations of research. *Journal of Cognitive Neuroscience, 25*, 1397-1405.

Illes, J., et al. (2010). NeuroTalk: Improving the communication of neuroscience. *Nature Reviews Neuroscience*, *11,* 61-69.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness*. Cognitive Psychology, 3*, 430-454.

Keehner, M., Mayberry, L., & Fischer, M. H. (2011). Different clues from different views: The role of image format in public perceptions of neuroimaging results. *Psychonomic Bulletin & Review, 18*, 422-428.

Keil, F. C. (2006). Explanation and understanding. *Annual Review of Psychology, 57*, 227-254.

Klaczynski, P. A. (2000). Motivated scientific reasoning biases, epistemological beliefs, and theory polarization: A two-process approach to adolescent cognition. *Child Development, 71*, 1347-1366.

Koslowski, B., & Masnick, A. (2010). Causal Reasoning. In U. Goswami (Ed.), *Handbook of Childhood Cognitive Development*. Oxford: Blackwell Publishers.

Kuhn, D., Phelps, E., & Walters, J. (1985). Correlational reasoning in an everyday context. *Journal of Applied Developmental Psychology, 6*, 85-97.

Lavazza, A., & De Caro, M. (2010). Not so fast. On some bold neuroscientific claims concerning human agency. *Neuroethics*, *3*, 23-41.

Lindell, A. K., & Kidd, E. (2013). Consumers favor "right brain" training: The dangerous lure of neuromarketing. *Mind, Brain, and Education, 7*, 35-39.

McCabe, D. P., & Castel, A. D. (2008). Seeing is believing: The effect of brain images on judgments of scientific reasoning. *Cognition*, *107*, 343-352.

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*(3), 437-455.

Michael, R. B., Newman, E. J., Vuorre, M., Cumming, G., Garry, M. (2013). On the (non)persuasive power of a brain image. *Psychonomic Bulletin & Review, 20*, 720-725.

Miller, J. D. (1998). The measurement of civic scientific literacy. *Public Understanding of Science, 7*, 203-223.

Muller, D., Judd, C. M., & Yzerbyt, V. Y. (2005). When moderation is mediated and mediation is moderated. *Journal of Personality and Social Psychology, 89*(6), 852-863.

O'Connor, C., & Joffe, H. (2013). How has neuroscience affected lay understandings of personhood? A review of the evidence. *Public Understanding of Science*, *22*, 254-268.

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*(4), 867-872.

Racine, E., Bar-Ilan, O., & Illes, J. (2005). fMRI in the public eye. *Nature Reviews Neuroscience, 6*, 159-164.

Racine, E., Waldman, S., Rosenberg, J., & Illes, J. (2010). Contemporary neuroscience in the media. *Social Science & Medicine*, *71*, 725-733.

Rhodes, R. E., Rodriguez, F., Shah, P. (2014). Explaining the alluring influence of neuroscience information on scientific reasoning. *Journal of experimental psychology: learning, memory, and cognition, 40*(5), 1432-1440.

Roskies, A. (2007). Are neuroimages like photographs of the brain? *Philosophy of Science*, *74*, 860-872.

Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science, 26*, 521-562.

Sedlmeier, P., & Gigerenzer, G. (1997). Intuitions about sample size: The empirical law of large numbers. *Journal of Behavioral Decision Making, 10*, 33-51.

Stanovich, K. E., & West, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology, 89*, 342-357.

Trout, J. D. (2007). The psychology of scientific explanation. *Philosophy Compass, 2*, 564-591.

Trout, J. D. (2008). Seduction without cause: Uncovering explanatory neurophilia. *Trends in Cognitive Sciences, 12*, 281-282.

Weisberg, D. S., Keil, F. C., Goodstein, J., Rawson, E., & Gray, J. R. (2008). The seductive allure of neuroscience explanations. *Journal of Cognitive Neuroscience, 20*, 470-477.

White, P. A. (1989). A theory of causal processing. *British Journal of Psychology, 80*, 431-454.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9*(2, Pt.2), 1-27.

Appendix

Science Literacy Scale

(Adapted from Miller, 1998)

True/False

1. Lasers work by focusing sound waves.
2. All radioactivity is man-made.
3. Electrons are smaller than atoms.
4. The center of the Earth is very hot.
5. The continents have been moving their location for millions of years and will continue to move.
6. It is the father's gene that decides whether the baby is a boy or a girl.
7. Antibiotics kill viruses as well as bacteria.

8. Two scientists want to know if a certain drug is effective against high blood pressure. The first scientist wants to give the drug to 1000 people with high blood pressure and see how many of them experience lower blood pressure levels. The second scientist wants to give this drug to 500 people with high blood pressure, and not give this drug to another 500 people with high blood pressure, and see how many in both groups experience lower blood pressure levels.
    a. Which scientist suggests a better way to test this drug?

9. A doctor tells a couple that their genetic makeup means that they've got a one-in-four chance of having a child with an inherited illness. Does this mean that if their first child has the illness, the next three will not?
    a. Does this mean that if their first child has the illness, the next three will not?
    b. Does this mean that each of the couple's children will have the same risk of suffering from the illness?

CHAPTER 4

Preference for Micro-Level Science: Evidence and Implications

Introduction

Previous research has investigated the influence of neuroscience explanations or pictures of brain data in popular media and found that neuroscience can influence scientific reasoning. Specifically, the presence of neuroscience can make explanations more satisfying (Fernandez-Duque, Evans, Colton, & Hodges, in press; Michael, Newman, Vuorre, Cumming, & Garry, 2013; Weisberg, Keil, Goodstein, Rawson, & Gray, 2008), make people feel like they have a better mechanistic understanding of the phenomenon (Rhodes, Rodriguez, & Shah, 2014), and even increase expectations that the study is likely to replicate (Rhodes & Shah, in revision). Although researchers have demonstrated the potential for neuroscience to influence scientific reasoning, many questions remain regarding the necessary conditions required for neuroscience to be influential, the cognitive processes underlying this effect, and the practical implications this effect might have on everyday reasoning. It is also unclear the extent to which similar effects exist for other micro-level, as opposed macro-level, methodologies. Here we present two experiments examining the extent to which micro-level information is preferred for a variety of research scenarios, as well as the influence of various factors such as the way in which the micro-level information is integrated into the research study and the type of subjects used in the research study. Additionally, we examine how the preference for micro-level information affects assumptions about the appropriate sample size needed.

Research studies that are reported in the media often contain evidence collected from a study, an explanation for that evidence, and a conclusion based on that evidence as well as implications. Much research has examined how people evaluate explanations. For example, people often judge explanations based on the sense of understanding they provide and how they fit with their own intuitions, rather than how accurate they are (Trout, 2007). Furthermore, people likely overestimate the extent to which they understand an explanation for a pattern of results, especially if they are low in scientific literacy (Rhodes & Shah, in revision). Despite the challenges in reasoning about explanations, explanations play an important role in scientific reasoning. Specifically, explanations provide a causal framework through which evidence can be understood and evaluated (Koslowski, Marasia, Chelenza, & Dublin, 2008). Koslowski and colleagues even argue that information is not considered evidence unless an explanation is provided. Given the presumed power of explanations, an interesting question is whether micro-level information would still increase perceived understanding if it were only presented as evidence, not as an explanation. In the present work, we define evidence as the raw outcome of a study and explanation as the mechanistic information that connects an outcome to the researchers' conclusion.

Previous research that has examined micro-level information has not systematically manipulated how that information is integrated into the research scenario. In Experiments 2.1-3.4, micro-level information was incorporated into the research study as a suggested mechanistic explanation for the effect described in the fictional research study. In contrast, much of the previous research examining neuroscience information has included it as evidence, not an explanation, and has failed to find significant effects (Gruber & Dickinson, 2012; Hook & Farah, 2013; Michael et al., 2008). Thus, it is important to examine whether the influence of micro-level information depends on the way it is integrated into the research study.

Understanding how the type of integration influences the preference for micro-level information can provide more insight into the assumptions and/or beliefs people have about micro-level information. For example, if people only prefer micro-level information to macro-level information when it is presented as evidence, but not when it is presented as an explanation, it would suggest that people believe micro-level evidence to be more valid than macro-level evidence, perhaps resulting from more sophisticated and precise technology. In contrast, if people only prefer micro-level information when it is presented in an explanation, it would suggest that the appeal of micro-level information relates to its ability to provide a mechanistic framework for the phenomenon at hand. Another possibility is that micro-level information would only be preferred when it is presented as both evidence and explanation. Indeed, the findings from Koslowski et al. (2008) suggest that evidence alone without an explanation leads to weaker causal attributions, and the same might be true when micro- or macro-level explanations are presented alone without specific evidence. Additionally, people value mechanistic information (Ahn, Kalish, Medin, & Gelman, 1995), so people may prefer scenarios that provide explanations rather than just evidence. Given the interdependence of evidence and explanation (Koslowski et al. 2008), we predict that ratings of understanding and quality will be highest when both evidence and explanations are present.

In addition to identifying the conditions under which micro-level information is perceived as more valid, a second goal of this work is to examine the implications that these micro-level assumptions have for expectations about appropriate sample sizes. If micro-level evidence is perceived as more valid, this could have implications for research methodology. For example, people might assume that a research study that involves micro-level information does not require as large of a sample size as a research study that involves macro-level information. Understanding whether beliefs about the validity of micro-level information affects sample size expectations is important for understanding how people reason about scientific evidence in their everyday lives. Recent meta-analyses of neuroscience and animal studies, both of which tend to employ micro-level methodologies, have suggested that many, if not most, of these studies are underpowered (Button et al., 2013; Macleod, 2011). Underpowered studies mean that, if significant effects are found, they are likely to be false positives. Additionally, research suggests that publication biases exist in many fields – that is, studies that show significant effects are more likely to be published than studies that show null effects. For the lay reader, the end result of this process might be a flashy headline that makes a bold claim based on significant effects from an underpowered study. Of course, it is not the case that all studies employing micro-level methodologies are underpowered; however, if people are willing to accept smaller sample sizes for micro-level than macro-level methodologies, this would have implications for how both the lay and scientific community view research studies that utilize different levels of analysis.

Finally, previous research has begun to examine how susceptibility to neuroscience may be moderated by individual differences, such as dualistic beliefs (Hook

& Farah, 2013) and domain-relevant knowledge (Weisberg et al., 2008; Rhodes et al., 2014), as well as prior beliefs and scientific literacy (Rhodes et al., 2014). However, it is unclear how individual differences moderate susceptibility to other types of micro-level information. The present research explores whether individual differences in reasoning ability and scientific knowledge predict preferences for micro-level information.

Overview

We present two experiments examining moderators of the preference for micro-level information and the implications this preference might have for judgments about sample size. In Experiment 4.1, participants read eight research descriptions, half of which contained micro-level information and the other half contained macro-level information, and the way in which the micro- vs. macro-level information was integrated into the research description was manipulated between-subjects. In Experiment 4.2, participants read the same eight research scenarios of one type of integration and estimated the number of subjects that would need to be in the study in order for them to be sufficiently confident in the researchers' conclusions.

Experiment 4.1

Methods

Participants

We recruited 330 participants (49.69% female; mean age = 35.45; range = 19-75) from Amazon Mechanical Turk. Approximately half of the participants (49.68%) had obtained at least a college degree. Fourteen participants who spent longer than 2.5 standard deviations of the mean time to complete the survey (M = 29.83 minutes, SD = 13.17) were excluded. Participants were paid $2.00 for their participation.

Materials

Participants read eight brief descriptions of fictional research studies that tested treatments of four ailments: cold symptoms, depression, canine aggression, and social fear in monkeys. In each fictional research study, there were treatment and control conditions as well as a main outcome of interest (measures of cold symptoms, depression, aggression, or social fear). Additionally, there was information about the potential mechanism through which the treatment acted. We manipulated within-subjects whether the mechanistic information came from a macro-level or a micro-level, and we manipulated between-subjects whether the macro- and micro-level information was presented in the form of evidence, an explanation, or both evidence and explanation.

Type of mechanistic information. For each ailment, two research studies were described: one that included mechanistic information from a macro-level of analysis and one that included mechanistic information from a micro-level analysis. In each case, the mechanistic information was chosen based on existing research studies that suggest that this could be a mechanism for the effect of interest. Two of the ailments pertained to animals and two pertained to humans. For the human studies, macro-level information was typically obtained through surveys. For the animal studies, macro-level information was obtained through behavioral observation. The types of mechanistic information given for each ailment are listed in Table 4.1.1.

Table 4.1.1
Mechanistic Information Provided in Research Descriptions

Scenario	Macro-level: Method of Data Collection	Macro-level: Mechanistic Information	Micro-level: Method of Data Collection	Micro-level: Mechanistic Information
Tea reduces cold symptoms	Survey	Tea decreases anxiety	Blood test	Tea decreases cortisol
Therapy reduces canine aggression	Behavioral observation	Therapy decreases anxiety	Blood test	Therapy decreases catecholamine levels
Therapy decreases depressive symptoms	Survey	Therapy decreases chronic stress	Neuroimaging	Therapy decreases amygdala activity
Therapy reduces social fear in monkeys	Behavioral observation	Therapy reduces psychological stress	Neuroimaging	Therapy reduces pulvinar signaling

Integration of mechanistic information with research study. In addition to the within-subjects manipulation of the level of analysis at which the mechanistic information came from, we also manipulated how this information was integrated within the research description. Specifically, we manipulated whether the mechanistic information was presented in the research study's result (Evidence Only condition), the explanation for the research study's result (Explanation Only condition), or both (Evidence and Explanation condition; see Table 4.1.2). Note that in the Evidence Only condition, the study's result was an effect on the micro/macro evidence, but the study's conclusion was that treatment might also reduce the outcome of interest. The study's result and conclusion were linked together by a statement suggesting that micro/macro evidence collected might also have an effect on the outcome of interest; thus, the conclusion was based on a presumed correlational relationship and was primarily speculative.

Table 4.1.2
Integration Manipulations for Example Research Scenario

	Features of Fictional Research Study: Exotic Tea Reduces Cold Symptoms			
Condition	**Treatment**	**Study Result**	**Explanation for Result**	**Study Conclusion**
Evidence Only	Tea	Decrease in anxiety	None	Tea might decrease cold symptoms
		Decrease in cortisol	None	Tea might decrease cold symptoms
Explanation Only	Tea	Decrease in cold symptoms	Tea reduces anxiety, which impairs immune functioning	Tea decreases cold symptoms
		Decrease in cold symptoms	Tea reduces cortisol, which impairs immune functioning	Tea decreases cold symptoms
Evidence & Explanation	Tea	Decrease in cold symptoms & anxiety	Tea reduces anxiety, which impairs immune functioning	Tea decreases cold symptoms
		Decrease in cold symptoms & cortisol	Tea reduces cortisol, which impairs immune functioning	Tea decreases cold symptoms

Procedure

For each ailment, the micro and macro versions of the study were presented on the same page, with the macro version appearing first and the micro version appearing immediately after it. Participants were instructed to read the two studies and then answer the questions that follow. A block design was used so that participants answered all questions pertaining to a single ailment before moving on to the next ailment.

Preference for micro-level information. The first question asked participants to indicate which study's conclusion was more supported, in a forced choice format. Selection of the micro version was coded as '1' and selection of the macro version was coded as '0'. Participants were also asked to indicate their confidence in their choice, on a scale of 0-100%. The dichotomous choice scores were multiplied by these confidence ratings to create weighted preference scores. For example, a participant whose preference score was ‘50’ meant they preferred the micro version of the study and were 50% confident in this choice. A benefit of this approach is that participants who selected the micro version but were not at all confident in their choice (e.g., confidence was 0) received a weighted preference score of 0. We expected that the preference for micro-level information could be driven by perceptions of quality and/or perceptions of understanding, so we also measured these dimensions.

Perceived understanding. For each of the two research studies, participants rated their perceived understanding on a scale of 1 (not at all) to 6 (completely) of the reason *why* the treatment had the observed effect on the outcome of interest. These questions assessed participants’ perceived mechanistic understanding of the phenomenon. For correlational analyses, we measured individual differences in perceived understanding by subtracting the perceived understanding ratings for the macro-level version from that of the micro-level version. Raw understanding ratings were used in the multilevel models.

Perceived quality. Finally, on a separate page, participants were presented with the two research studies again and were asked to rate the quality of each research study, on a scale of 1 to 5. For correlational analyses, we measured individual differences in quality ratings by subtracting the quality ratings for the macro-level version from that of the micro-level version. Raw quality ratings were used in the multilevel models.

Preference for micro-level scientists. After reading the research scenarios, participants also rated the scientific rigor (how closely its practitioners adhere to the scientific method) of 12 scientists from various subdisciplines on a scale of 1 (not at all) to 10 (absolutely). The question was based on a similar question used in Fernandez-Duque et al. (in press). Order of presentation was random. The 12 subdisciplines were composed of micro- and macro-level versions of six general disciplines: anthropology, economics, psychology, chemistry, physics, and biology. Half of the general disciplines were considered social sciences and the other half were considered physical/life sciences. The subdisciplines that were generated from these general disciplines are listed below in Table 4.1.3.

Table 4.1.3
List of Subdisciplines in Scientific Fields Questionnaire

General Discipline	Subdiscipline	Level
Anthropologist	Cultural Anthropologist	Macro
	Biological Anthropologist	Micro
Economist	Behavioral Economist	Macro
	Neuroeconomist	Micro
Psychologist	Cognitive Psychologist	Macro
	Neuropsychologist	Micro
Chemist	Geochemist	Macro
	Nanochemist	Micro
Physicist	Geophysicist	Macro
	Quantum physicist	Micro
Biologist	Systems biologist	Macro
	Microbiologist	Micro

Individual Differences

Essentialism. Participants completed a 23-item scale from Bastian and Haslam (2008), which measured the extent to which people think in essentialist ways. Participants rated their agreement with each item on a scale of 1 (Strongly Disagree) to 6 (Strongly Agree). Specifically, the scale was composed of three subscales: biological basis, discreteness, and informativeness. The biological basis subscale measured beliefs about the extent to which human attributes are biologically determined (e.g., “The kind of person someone is can be largely attributed to their genetic inheritance”), the discreteness subscale measured beliefs about the extent to which attributes are distinct (e.g., “A person either has a certain attribute or they do not”), and the informativeness subscale measured beliefs about the extent to which attributes can be easily determined (e.g., “When getting to know a person it is possible to get a picture of the kind of person they are very quickly”). We predicted that people who tend to think in essentialist ways would be more influenced by micro-level information than people who do not have strong essentialist beliefs.

Actively Open-Minded Thinking. Participants completed the 41-item Actively Open-Minded Thinking (AOT) Scale from Stanovich and West, 1997. This scale assesses participants’ ability to reason independently of their prior beliefs. Performance on this scale is believed to be an index of cognitive flexibility and reflects sophisticated thinking dispositions. We predicted that people who score highly on the AOT scale would be more critical of the research studies and potentially less influenced by the presence of micro-level information than people who score low on the AOT scale.

Scientific Literacy. Participants completed a civic scientific literacy scale based on work by Miller (1998). The scale consisted of seven declarative knowledge questions about basic physics and biology (e.g., “Antibiotics kill viruses as well as bacteria”), and two questions about probability (e.g., “A doctor tells a couple that their genetic makeup means that they’ve got a one-in-four chance of having a child with an inherited illness. Does this mean that if their first child has the illness, the next three will not?”). Additionally, the scale included one question about the scientific process, shown below.

> Two scientists want to know if a certain drug is effective against high blood pressure. The first scientist wants to give the drug to 1,000 people with high blood pressure and see how many of them experience lower blood pressure levels. The second scientist wants to give this drug to 500 people with high blood pressure, and not give this drug to another 500 people with high blood pressure, and see how many in both groups experience lower blood pressure levels. Which scientist suggests a better way to test this drug?

We predicted that people who have greater scientific literacy would be more critical of the research studies and less influenced by micro-level information than people who are lower in scientific literacy.

Results

The mean weighted preference scores were well above 0 for all scenarios and conditions, indicating a widespread preference for the micro versions of the research studies (Table 4.1.4). Reliability of the four weighted preference scores was high (alpha = .80, CI: .72-.88). In general, people were substantially confident in their choice of the most supportive research study (M = 71.33, SD = 22.12).

Table 4.1.4
Weighted Preference For Micro-Level Version of Each Research Scenario

		Tea	**Dogs**	**Depression**	**Monkeys**	
Type of Manipulation	**N**	Support | Micro M (SD)	Support | Micro M (SD)	Support | Micro M (SD)	Support | Micro M (SD)	**Mean (SD)**
Evidence	121	66.56 (32.65)	51.28 (39.10)	59.52 (37.64)	41.69 (39.76)	54.76 (26.92)
Explanation	107	59.13 (33.51)	44.63 (37.81)	44.42 (38.26)	45.60 (37.06)	48.45 (29.95)
Evidence & Explanation	102	55.17 (35.47)	42.82 (39.61)	46.04 (39.67)	44.23 (37.91)	47.07 (32.35)

Correlations between the three dependent variables (weighted preference scores, differential understanding, and differential quality) and the individual difference measures of interest are listed in Table 4.1.5. The three dependent variables were positively correlated with each other. Additionally, AOT and scientific literacy were positively correlated with both weighted preferences scores and differential quality scores. These positive correlations suggest that individuals who are more sophisticated in their thinking dispositions and scientific knowledge are actually more likely to prefer the micro-level than macro-level versions of the research studies. As expected, preference for micro-level scientific fields was also positively correlated with both weighted preference scores and differential quality scores; it was also positively correlated with scientific literacy. Interestingly, AOT, scientific literacy, and preference for scientific fields were not significantly correlated with differential understanding. Finally, AOT was positively correlated with scientific literacy and negatively correlated with essentialism.

Table 4.1.5
Correlations Between Dependent Variables and Individual Differences

	1	2	3	4	5	6	7
1. Weighted preference scores	--						
2. Differential quality	.68***	--					
3. Differential understanding	.43***	.61***	--				
4. Preference for micro-level fields	.13*	.15**	.09	--			
5.AOT	.19***	.25***	.09	.09	--		
6. Scientific literacy	.17**	.18**	.09	.16**	.43***	--	
7. Essentialism	-.06	-.05	-.11	-.05	-.28***	-.08	--

To determine whether the condition had an effect on the weighted preference scores, we constructed a multilevel model with the type of subject (human vs. animal) modeled as a random effect and type of integration (Evidence, Explanation, Both), AOT, scientific literacy, and essentialism modeled as fixed effects (Table 4.1.6). The type of subject used in the research study had a significant effect on the weighted preference scores, indicating that the preference for micro-level information was stronger when the research study involved humans than when it involved animals. Interestingly, people who scored higher on the AOT scale showed significantly stronger micro-level preferences, and people who had higher than median scientific literacy showed slightly stronger micro-level preferences.

Table 4.1.6
Multilevel Model Predicting Weighted Preference Scores

	B (SE)	p
Integration: Explanation	-2.66 (4.37)	.54
Integration: Both	-3.18 (4.49)	.47
Scientific Literacy	22.39 (12.96)	.08
Subject: Human	16.63 (2.72)	.00
AOT	6.73 (2.72)	.01
Integration: Explanation X Subject: Human	-9.91 (3.95)	.01
Integration: Both X Subject: Human	-9.42 (4.05)	.02

Additionally, although there was not a significant main effect of the type of integration, there was a significant interaction between integration type and the type of subject. As Figure 4.1.1 illustrates, weighted preference scores were similar for all integration types except when the research study involved humans; in these cases, the weighted preference for the micro-level version was significantly stronger in the Evidence Only condition than in the Explanation Only or Evidence & Explanation conditions. There was not a significant main effect of essentialism (p = .94).

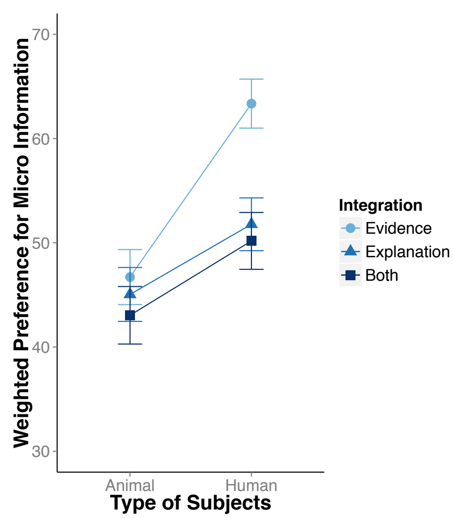


Figure 4.1.1 Weighted preference scores by integration and subject type

We also modeled ratings of quality and perceived understanding for the eight research studies. Table 4.1.7 shows the results of a multilevel model for quality ratings with level of analysis and type of subject modeled as random effects and type of integration, AOT, scientific literacy, and essentialism modeled as fixed effects. There were significant main effects of subject type, AOT, level of analysis, and a marginally significant effect of essentialism, as well as several significant interactions.

Table 4.1.7
Multilevel Model Predicting Quality Ratings

	B (SE)	p
Integration: Explanation	-0.12 (0.09)	.21
Integration: Both	0.08 (0.09)	.39
Subject: Human	-0.33 (0.05)	.00
AOT	-0.21 (0.05)	.00
Level: Micro	-0.97 (0.29)	.00
Essentialism	0.09 (0.29)	.05
Integration: Explanation X Subject: Human	0.29 (0.07)	.00
Integration: Both X Subject: Human	0.22 (0.07)	.00
AOT X Level: Micro	0.30 (0.06)	.00
Subject: Human X Level: Micro	0.50 (0.07)	.00
Integration: Explanation X Level: Micro	-0.01 (0.11)	.86
Integration: Both X Level: Micro	-0.02 (0.12)	.83
Integration: Explanation X Subject: Human X Level: Micro	-0.35 (0.10)	.00
Integration: Both X Subject: Human X Level: Micro	-0.22 (0.10)	.03

The most interesting interaction was the three-way interaction between integration type, subject type, and level of analysis, illustrated in Figure 4.1.2. In general, micro-level information received higher quality ratings and the influence of integration type only mattered for research scenarios involving human subjects. For these scenarios, the relative increase in quality ratings when comparing macro- and micro-level versions of the research scenario was greatest for the evidence only condition. This finding is consistent with the interaction between type of integration and type of subject that was found for the weighted preference scores.

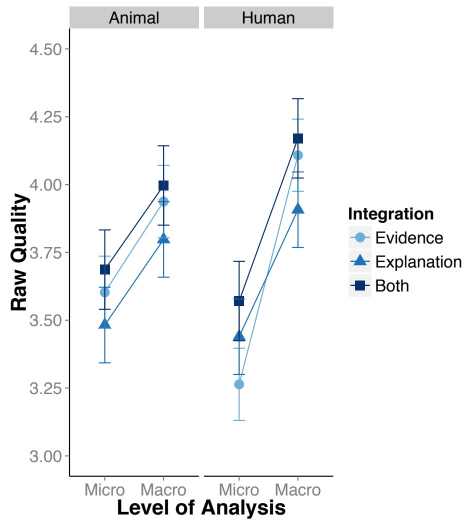


Figure 4.1.2. Quality ratings by level of analysis, subject type, and integration type. Evaluated at mean AOT. Bars represent 95% confidence intervals.

Together, these findings suggest that the evidence only condition might make the weaknesses of human macro-level information and the strengths of human micro-level information particularly salient, perhaps due to the lack of an explanation of how the researchers' results supported their conclusions. There was also a significant interaction between AOT and level of analysis, indicating that people who scored higher on the AOT showed a larger difference in their ratings of quality ($M_{\text{Diff}} = 0.66$) for micro and macro level studies than did people who scored lower on the AOT ($M_{\text{Diff}} = 0.31$). There was not a significant main effect of scientific literacy ($p = .13$), nor was there an interaction between scientific literacy and level of analysis ($p = .10$). There was a marginally significant main effect of essentialism, indicating that people with stronger essentialist beliefs tended to give slightly higher quality ratings, but there was no interaction with level of analysis (p = .66).

Table 4.1.8 shows the results of a multilevel model predicting ratings of perceived understanding with level of analysis and type of subject modeled as random effects and type of integration, AOT, and scientific literacy modeled as fixed effects. There were significant main effects of subject type, AOT, and level of analysis in the predicted directions, but there were no main effects of scientific literacy (p = .52) or essentialism (p = .50) and no interactions between scientific literacy and level of analysis (p = .32) or essentialism and level of analysis (p = .12). Additionally, there was a significant interaction between type of subject and level of analysis, as well as a marginally significant interaction between AOT and level of analysis.

Table 4.1.8
Multilevel Model for Perceived Understanding

	B (SE)	p
Subject: Human	-0.03 (0.04)	.41
AOT	0.02 (0.08)	.85
Level: Micro	-0.84 (0.37)	.02
Integration: Evidence	0.13 (0.11)	.23
Integration: Both	0.31 (0.12)	.01
AOT X Level: Micro	0.14 (0.08)	.09
Subject: Human X Level: Micro	0.42 (0.05)	.00

The interaction between type of subject and level of analysis, controlling for AOT, is illustrated in Figure 4.1.3. Understanding was higher for micro-level versions when the research scenarios involved human subjects, but understanding was higher for macro-level versions when the research scenarios involved animal subjects.

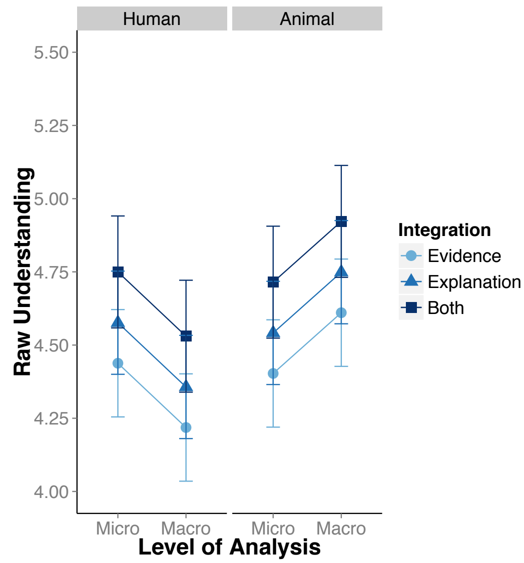


Figure 4.1.3. Perceived understanding by level, subject type, and integration type. Evaluated at mean AOT. Bars represent 95% confidence intervals.

Responses to the scientific fields scale were highly correlated with each other and had high internal consistency (CI alpha = .86 - .92, M = .89). As expected, micro-level fields received significantly higher ratings of scientific rigor than did macro-level fields (M = 8.07, SD = 1.39 and M = 7.32, SD = 1.43, respectively; $t(329) = 15.74$, $p < .001$). We created an individualized preference for micro fields score by subtracting each participant's mean ratings for macro-level fields from their mean ratings for micro-level fields. We conducted a linear regression analysis to examine individuals' preferences for micro-level fields with AOT, essentialism, and scientific literacy as predictors. Scientific literacy was a significant predictor (B = 0.90, $SE(B)$ = 0.39, p = .01), indicating that people who had higher scientific literacy showed stronger preferences for micro-level scientific fields, but AOT and essentialism were not (ps > .50). The model was significant but explained a small proportion of the variance, F (3,311) = 2.87, $p < .05$, R^2 = .02.

Discussion

This experiment demonstrated that, when forced to choose between micro- and macro-level information, people have a consistent preference for micro-level information. Furthermore, this preference was related to scientific literacy and AOT. Specifically, people who were more scientifically literate and higher in AOT were more likely to indicate that research studies that involve micro-level information were of higher quality, and that the conclusions based on these studies were more supportive of the evidence, compared to research studies that involve macro-level information. Additionally, people who were more scientifically literate were more likely to rate micro-level scientific fields as more scientific than their macro-level counterparts. Interestingly, however, we found that scientific literacy and AOT were not significantly correlated with perceived understanding. If the preference for micro-level information were driven by unconscious heuristics, we might expect that higher scores on the scientific literacy and AOT assessments would be negatively associated with the preference for micro-level information. Therefore, our results suggest that the preference for micro-level information is not driven by unconscious heuristics but instead by a deliberative process.

In general, perceived understanding and quality tended to be highest when micro-level information was presented in both the evidence and explanation. Ratings in this condition were significantly higher than ratings in the Explanation Only condition, but not the Evidence Only condition. One interpretation of this finding could be that evidence is valued over explanation, which suggests that, contrary to Koslowski et al. (2008), explanation does not always increase the value of evidence. Weighted preference scores were similar in the Explanation Only and Explanation & Evidence conditions, but significantly higher in the Evidence Only condition when the research study involved humans. This result suggests that either the micro-level information appeared particularly attractive in these conditions, or the macro-level information appeared particularly unattractive. We expect that the latter is the case, since the macro-level research studies that involved humans utilized survey data as the dependent variable, and results from a think aloud pilot study suggested that people are aware of self-report biases that can emerge from surveys. Additionally, the Evidence Only condition provided the least amount of information possible – just a description of what the research study did and concluded, but no explanation linking the two – which means the shortcomings of the human macro-level data may have been particularly salient in this condition.

Our finding that the preference for micro-level information is significantly correlated with scientific literacy and AOT suggests that people may be using a rational strategy when directly comparing macro- and micro-level information. One such strategy may be that people believe that micro-level information is inherently more valid than macro-level information. Experiment 4.2 tests this possibility by examining how micro- vs. macro-level information affects intuitions about the size of the sample needed for a research study. Many micro-level studies are based on significantly smaller sample sizes than macro-level studies, so understanding how people assign weight to the level of analysis and the sample size may shed further light on the types of strategies people use when evaluating research studies as well as the practical implications of these strategies.

Experiment 4.2

Overview

The purpose of this experiment was to examine whether micro-level information affected estimations of the size of the sample needed for a research study. We manipulated the level of analysis (micro vs. macro) and the sample size (large vs. small) within-subjects. If people believe that micro-level information is inherently more valid, they may accept a smaller sample size for these studies than they would for macro-level studies. However, given that previous research suggests that people are well aware that sample size is an important factor, and are likely to pay attention to this factor in a "knee-jerk" way, it is also possible that people may no longer use micro- vs. macro-level information as an evaluation criterion once sample size information is present. Such a finding would suggest that the preference for micro-level information is not necessarily due to an assumption of its superior validity, but instead a result of some other strategy.

Method

Participants

We recruited 200 participants (40.83% female; mean age = 34.45, range = 19-68) from Amazon Mechanical Turk. Approximately half (48.67%) of the sample had obtained at least a college degree. Eight people were excluded for taking longer than 2.5 standard deviations of the mean time to complete the survey (M = 25.56 minutes, SD = 14.47). Participants were paid $2.00 for their participation.

Materials and Procedure

Similar to Experiment 4.1, participants read eight brief descriptions of research studies. The research studies were the same ones used in the Evidence & Explanation condition of Experiment 4.1 with the addition of information about sample size. The sample size described in each research study was small. For each research study, participants were asked to 1) rate their confidence in the researchers' conclusion on a scale of 1 (not at all) to 6 (completely), and 2) indicate how many participants they would need to see in a new study to be sufficiently convinced that the researchers' conclusion is true. Half of the research studies used micro-level evidence and explanations, and the other half used macro-level. Micro- and macro-level versions were organized into blocks, and participants were randomly assigned to see either the micro- or macro- block first. Within each block, the order of presentation of the research studies was randomized. This design allowed for both within-subjects and between-subjects (for the first block only) analyses of the influence of level of analysis. Importantly, for a given research scenario, the sample sizes provided in the micro and macro versions were identical; however, the sample sizes for different research scenarios were not the same, and the sample sizes provided for studies involving humans were slightly higher than those provided for animal studies. To account for the differences in original sample sizes, sample size estimations were divided by the sample size provided in the original study. Additionally, we expected that the variance in sample size estimates would be very high, so all estimates were log transformed to achieve a more normal distribution. Results from a think aloud pilot study (see Appendix) revealed that participants tended to not attend to sample size until later in the experiment. To increase attention to the sample size, we explicitly mentioned that participants would be seeing information about sample size and the type of evidence collected in the research study, and that participants should feel free to make their decisions based on whichever information they felt was most relevant.

Table 4.2.1
Sample Sizes Provided in Research Studies

Research Study	Sample Size Provided
Tea reduces cold symptoms	10
Therapy reduces depression	12
Therapy reduces canine aggression	6
Therapy reduces social fear in monkeys	5

Note. Sample sizes for micro- and macro- versions of each research study were the same.

Individual differences. After evaluating the eight research descriptions, participants completed the scientific fields questionnaire and the essentialism, AOT, and scientific literacy scales used in Experiment 4.1.

Results

A multilevel model for confidence ratings with level of analysis and type of subject as random effects and literacy and condition as fixed effects revealed significant effects of level of analysis (B = 0.25, SE = 0.08, p < .01) and scientific literacy (B = -1.95, SE = 0.57, p < .001). People were more confident in the conclusions based on studies that included micro-level information than macro-level information, and people who were more scientifically literate tended to be less confident in researchers' conclusions. Interestingly, there was also a significant interaction between level of analysis and condition, indicating that that participants tended to be more confident in the level of analysis they saw first (B = -0.35, SE = 0.12, p < .01). Thus, participants in the Micro First condition were more confident in micro- than macro-level studies (M_{Micro} = 2.89, SD_{Micro} = 1.32, M_{Macro} = 2.99, SD_{Macro} = 1.32), and participants in the Macro First condition were more confident in macro- than micro-level studies (M_{Micro} = 3.13, SD_{Micro} = 1.32, M_{Macro} = 2.87, SD_{Macro} = 1.40). There was a marginally significant effect of subject type, indicating that people tended to be slightly more confident in research studies that used humans rather than animals (B = 0.51, SE = 0.29, p = .07), but there was also a marginally significant interaction between type of subject and scientific literacy, indicating that participants who had lower scientific literacy had a slightly larger difference in confidence between human and animal studies (B = -0.57, SE = 0.34, p = .09). Specifically, participants with lower scientific literacy tended to be slightly more confident in human studies than animal studies (M_{Human} = 3.45, SD_{Human} = 1.45, M_{Animal} = 3.26, SD_{Animal} = 1.46), whereas participants with higher scientific literacy gave similar confidence ratings for both types of studies (M_{Human} = 2.78, SD_{Human} = 1.23, M_{Animal} = 2.82, SD_{Animal} = 1.27).

Our primary analysis of interest was to examine predictors of sample size estimations. We constructed a multilevel model with level of analysis and type of subject modeled as random effects and condition (Micro First vs. Macro First), scientific literacy, confidence in the researchers' conclusions, and individual preferences for micro-level fields modeled as fixed effects (Table 4.2.2). There was a main effect of scientific literacy, indicating that people with higher scientific literacy preferred larger sample sizes. There was also a significant main effect of subject type, indicating that participants tended to estimate larger sample sizes for studies that involved humans. There were no significant effects of condition (p = .62) or essentialism (p = .28), and there was no interaction between essentialism and level of analysis (p = .90).

Table 4.2.2
Multilevel Model Predicting Sample Size Estimations

	B (SE)	*p*
Level: Micro	0.21 (0.09)	.19
Subject: Human	0.38 (0.12)	.00
Scientific Literacy	3.70 (0.82)	.00
Confidence	0.08 (0.11)	.45
Preference for micro-level fields	0.19 (0.11)	.07
Level X Confidence	-0.06 (0.02)	.02
Subject X Confidence	-0.12 (0.03)	.00
Literacy X Confidence	-0.34 (0.13)	.01
Level X Fields	-0.11 (0.04)	.00

There were several interesting interactions that emerged, which are illustrated in Figure 4.2.1. The top left figure shows that people tended to estimate smaller sample sizes for micro-level than macro-level studies, but this difference was larger for people who tended to believe that micro-level fields were more scientific than macro-level fields. The top right figure shows the difference in sample sizes estimates for micro- vs. macro-level studies by confidence; the two lines are not parallel, indicating that people who were not very confident in the research study's conclusions gave closer estimates for micro- and macro-level studies, while people who were more confident in the researchers' conclusions were more influenced by the level of analysis of the study. The bottom left figure shows that people who were not confident in the research were the ones who estimated larger sample sizes for the human studies. Finally, the bottom right figure shows that the difference in sample size estimations for people who were low versus high in scientific literacy was larger when people were not confident in the research than when people were confident; in other words, scientific literacy played a larger role in determining sample size estimations when people had low confidence.

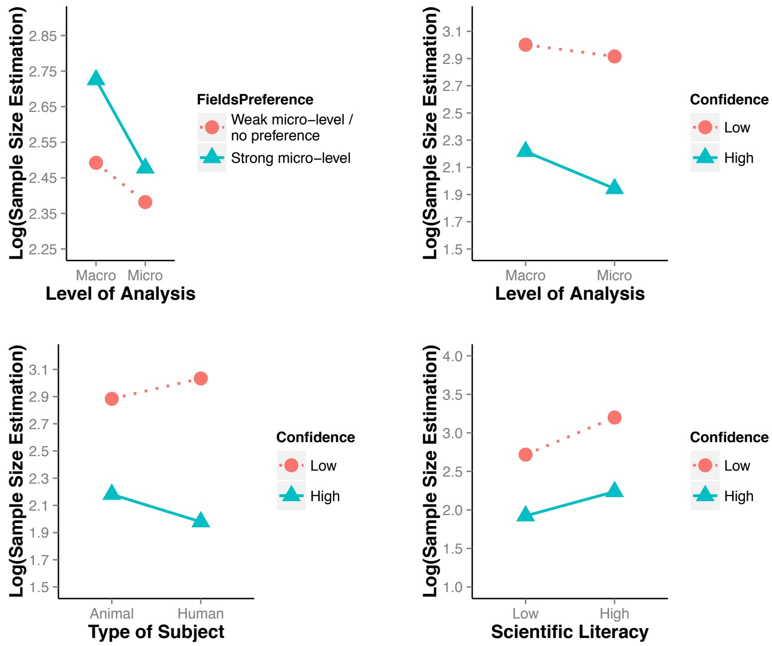


Figure 4.2.1 Significant interactions for model predicting sample size estimations.

We also performed between-subjects analyses for sample size by looking only at the first block of research studies. In these analyses, level of analysis was manipulated between subjects. We constructed a multilevel model for sample size estimations with type of subject as the only random effect and level of analysis, confidence in the researchers' conclusions, and scientific literacy as fixed effects. There was no effect of level of analysis (B = -0.03, SE = 0.18, p = .86), but there were significant effects of scientific literacy (B = 2.34, SE = 0.68, p < .001) type of subject (B = 0.59, SE = 0.15, p < .001), and confidence in the researchers' conclusions (B = -0.19, SE = 0.04, p < .001), indicating that participants who were more scientifically literate tended to estimate larger sample sizes, participants tended to estimate smaller sample sizes for humans than animals, and participants who were more confident in the researchers' conclusions estimated smaller sample sizes. There was a marginal effect of preference for micro-level fields ($B = 0.19$, $SE = 0.11$, $p = .08$), indicating that people thought micro-level fields were more scientific also estimated slightly larger sample sizes than people who had less of a preference for micro-level fields. There was also a significant interaction between confidence and type of subject ($B = -0.18$, $SE = 0.04$, $p < .001$), indicating that people who were less confident in the researchers' conclusions tended to estimate larger sample sizes for humans than animals ($M_{Human} = 3.66$, $SD_{Human} = 1.37$, $M_{Animal} = 3.23$, $SD_{Animal} = 1.28$), while people who were more confident in the researchers' conclusions estimated slightly larger sample sizes for animals than humans ($M_{Human} = 2.23$, $SD_{Human} = 1.74$, $M_{Animal} = 2.42$, $SD_{Animal} = 1.49$).

Discussion

This experiment showed that the preference for micro-level information is not restricted to judgments of quality and evidential support but is also reflected in peoples' expectations about the size of a research sample. One explanation for this finding may be that people believe micro-level information to be more valid and perhaps less variable across subjects. While it is true that studies taking place at the micro-level tend to have smaller sample sizes, given that our sample was primarily non-experts, it is unlikely that they would have known this beforehand. Thus, this experiment suggests that non-experts tend to reach this conclusion on their own.

An interesting finding from this experiment is that the influence of micro- and macro-level information on sample size estimations is only evident when participants are exposed to both conditions and must, therefore, make comparisons between macro- and micro-level research studies. One explanation for this finding may relate to the sparse amount of detail that was provided in the research descriptions. Several participants commented that the task was difficult because there was not much information to base their judgments off of. In this experiment, it was necessary to constrain the research description to contain only the information we were manipulating – level of evidence and sample size. For these stimuli, it seems likely that judgments of sample size would be easier in the within-subjects condition, since participants can at least compare micro-level versions to macro-level versions and use that comparison as a basis for their judgment. In a between-subjects design, however, participants only see the micro-level studies or the macro-level studies, so it is possible their sample size estimations are less informed.

General Discussion

The present work examined the conditions under which micro-level information is likely to be preferred to macro-level information. Specifically, we manipulated how the micro- and macro-level information was integrated into the research study and the type of subjects and method of data collection used in the research study, as well as individual differences related to micro-level preferences. We found that the way that the way micro-level information was integrated into the study did not have an effect on its appeal. The presence of micro-level information in any form tended to increase ratings of perceived quality and understanding. Interestingly, what appeared to matter more was the type of subjects that were used in the research study. It is important to keep in mind that macro-level human data was collected through self-report, while macro-level animal data was collected through observation. As a result, differences in human versus animal studies may simply reflect distrust in self-report data from humans. On the other hand, it could also reflect the belief that humans are more variable than animals and thus macro-level data may appear less valid for these subjects.

The finding that the type of integration did not have an effect on micro-level preference suggests that multiple mechanisms may be working to elicit micro-level preferences. People appear to hold two beliefs: 1) Micro-level explanations provide a stronger mechanistic framework than macro-level explanations, and 2) micro-level evidence is more valid than macro-level evidence. Support for the latter primarily comes from the findings from Experiment 4.2 – the presence of micro-level information reduces the sample size needed for a research study. One rationale people may be using to support this belief is that micro-level processes are less subject to change. For example, people might rationalize that self-report measures can fluctuate depending on the day and time that you collect them, but processes that occur at a biological level might appear less likely to have these random fluctuations. Although this rationale might apply to some micro- and macro-level processes, it certainly does not apply to all instances of these processes. For example, the brain is constantly changing based on one's experiences. Moreover, there are a number of other sources of variance that must be dealt with in micro-level methodologies. Perhaps more importantly, with some micro-level methodologies, especially newer ones that tend to receive a lot of media coverage, the mapping between what we observe and what we are trying to measure is not always straightforward; as a result, large sample sizes are needed in these fields. Additionally, false positives are typically high in fMRI research, which reiterates the need for large sample sizes. Of course, there are often more practical limitations to acquiring large sample sizes for micro-level methodologies – equipment is expensive and it is unethical to waste animal lives. Nevertheless, understanding people's assumptions about the validity of micro- versus macro-level information could help inform researchers about the public perception of their science as well as funding and publication decisions that occur within a researchers' subject area.

Another important goal of these experiments was to examine how individual differences relate to micro-level preferences. Our finding that people who are high in scientific literacy and AOT are more likely to endorse micro-level methodologies suggests that the preference for micro-level information results from a deliberative, as opposed to a heuristic, process. People likely have beliefs about the validity of micro-level information that are activated and influence their evaluations of micro- and macro-level information. Another piece of the explanation could be that people who are less scientifically literate are less familiar with the language used to describe micro-level processes and might be less likely to endorse the macro-level version, basing their judgment on familiarity instead. If this is the case, however, it is interesting that people who are less scientifically literate tend to rate their perceived understanding higher than people who are more scientifically literate; this pattern of results suggests that these individuals are not accurate (or perhaps truthful) in assessing their knowledge.

In conclusion, we have demonstrated that the preference for micro-level information is stable across different kinds of research scenarios and types of micro-level information and has implications for how people reason about methodology in everyday settings. While the valuation of micro-level information might be a rational process, it has implications for how reductionist approaches will be received and influence scientific fields.

References

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

Bastian, B., & Haslam, N. (2006). Psychological essentialism and stereotype endorsement. *Journal of Experimental Social Psychology*, *42*, 228-235.

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., & Robinson, E. S. J. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*, 365-376.

Fernandez-Duque, D., Evans, J., Colton, C., & Hodges, S. D. (in press). Superfluous neuroscience information makes explanations of psychological phenomena more appealing. *Journal of Cognitive Neuroscience*.

Gruber, D., & Dickerson, J. A. (2012). Persuasive images in popular science: Testing judgments of scientific reasoning and credibility. *Public Understanding of Science, 21,* 938-948.

Hook, C. J., & Farah, M. J. (2013). Look again: Effects of brain images and mind-brain dualism on lay evaluations of research. *Journal of Cognitive Neuroscience, 25*, 1397-1405.

Koslowski, B., Marasia, J., Chelenza, M., & Dublin, R. (2008). Information become evidence when an explanation can incorporate it into a causal framework. *Cognitive Development, 23*, 472-487.

Macleod, M. (2011). Why animal research needs to improve. *Nature,* 477, 511.

McCabe, D. P., & Castel, A. D. (2008). Seeing is believing: The effect of brain images on judgments of scientific reasoning. *Cognition, 107,* 343-352.

Michael, R. B., Newman, E. J., Vuorre, M., Cumming, G., Garry, M. (2013). On the (non)persuasive power of a brain image. *Psychonomic Bulletin & Review, 20,* 720-725.

Miller, J. D. (1998). The measurement of civic scientific literacy. *Public Understanding of Science, 7*, 203-223.

Rhodes, R. E., Rodriguez, F., & Shah, P. (2014). Explaining the alluring influence of neuroscience information on scientific reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(5), 1432-1440.

Stanovich, K. E., & West, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology, 89,* 342-357.

Trout, J. D. (2007). The psychology of scientific explanation. *Philosophy Compass*, 2, 564-591.

Weisberg, D. S., Keil, F. C., Goodstein, J., Rawson, E., & Gray, J. R. (2008). The seductive allure of neuroscience explanations. *Journal of Cognitive Neuroscience, 20*, 470-477.

Appendix

Description of Think Aloud Pilot Study

Protocol

Participants read eight research scenarios, all of which used the Evidence Only integration type. The type of ailment (tea vs. monkeys), the level of analysis (micro vs. macro), and the size of the sample size used in the research study (small vs. large) were manipulated within-subjects. Participants were instructed to read each scenario and rate 1) how confident they were in the researchers' conclusions, 2) how well they understood the reason why the researchers found what they did, and 3) how scientific the research seemed. Participants were instructed to think out loud and vocalize their thoughts as they answered each question. Responses were transcribed and categorized into seven themes.

Definitions of Themes

Terminology – Participants commented on how scientific the language/jargon seemed
Mechanism – Participants indicated how well the research description helped them conceptually understand why the researchers found what they found
Results – Participants re-stated the results of the research study
Research Design – Participants praised or critiqued the design of the research, including the use of control groups and operational definitions
Write-up – Participants referred to the level of detail provided in the research study
Data – Participants mentioned the presence or absence of data (e.g., numbers) in the research description
Sample Size – Participants remarked on the sample size used in the study

Frequency of themes, summed across eight research scenarios

	Factors Contributing to Lower Scores		Factors Contributing to Higher Scores	
Evaluation Measure	**Macro-level Studies**	**Micro-level Studies**	**Macro-level Studies**	**Micro-level Studies**
Confidence	research design (5)	sample size (8)	sample size (7)	mechanism (3)
	sample size (4)	mechanism (4)	mechanism (2)	results (3)
	mechanism (4)	research design (4)	results (2)	sample size (2)
	write-up (3)	data (3)	research design (1)	
	results (1)	terminology (2)		
	data (1)			
Understanding	mechanism (8)	mechanism (8)	mechanism (7)	mechanism (11)
	research design (3)		terminology (6)	
	write-up (3)			
Scientific	write-up (6)		write-up (5)	terminology (7)
	mechanism (4)		mechanism (4)	mechanism (3)
	terminology (3)		terminology (3)	
	research design (1)		data (1)	
	sample size (1)			

Key Themes

For ratings of confidence in the researchers' conclusions, participants were likely to critique the research design in macro-level studies and the sample sizes in both macro- and micro-level studies. When sample sizes were larger, participants were likely to mention this as a positive factor for macro-level studies. Participants tended to mention the mechanistic information as a positive factor for micro-level studies.

For ratings of perceived understanding, participants tended to think about whether they understood the reason why the treatment of interest elicited the observed effect. In several instances, participants also remarked that the lack of scientific jargon (terminology) in macro-level studies helped them understand the results better.

For ratings of how scientific the research seemed, participants were likely to mention that the scientific jargon (terminology) made the micro-level studies seem more scientific. Participants tended to mention that there was not much detail or specifics for macro-level studies (write-up), which contributed to lower ratings of "scientificness".

Important Takeaways

The sophistication of the language and the lucidity of the mechanism are important factors for judging perceived understanding and the "scientificness" of the research. Participants attend to how scientific the language seems when deciding how scientific the research is. However, the use of more easily understandable language in macro-level studies may contribute to better mechanistic understanding for some people.

Although participants did attend to sample size, they only attended to it about halfway through the research studies. Once they realized that the sample size changed across studies, they tended to use that as another judgment criterion. For this reason, Experiment 4.2 used instructions that explicitly drew participants' attention to the sample size mentioned in each research description.

CHAPTER 5

Conclusion

The goal of my dissertation was to examine how people reason scientifically about information that comes from different levels of analysis. Specifically, I examined how dispositional and contextual factors influence the extent to which people prefer micro-level information over macro-level information. Across eight studies, I found that micro-level information is perceived as more supportive of a researchers' conclusion, across a variety of scenarios; improves perceived understanding; increases expectations of replication for a future study; and decreases the estimated sample size needed for a research study. Taken together, these findings suggest that people perceive micro-level information in a fundamentally different, and more favorable, way than they perceive macro-level information. However, there are important individual differences and contextual factors that moderate this effect.

Summary of Key Findings

Experiments 2.1 and 2.2 showed that the presence of neuroscience increased ratings of scientist quality and mechanistic understanding of a cognitive phenomenon after controlling for individual differences in thinking dispositions. Experiments 3.1 and 3.2 found similar effects on mechanistic understanding when comparing neuroscience information to comparable psychology information, and also found that neuroscience information made people more likely to expect replication in a future study, especially for people who did not already engage in the behavior being promoted. Experiment 3.3 found that eye tracking information also increased perceived understanding of the cognitive phenomenon, although the effect size was smaller than it was for neuroscience information, and that people tended to think neuroscience evidence provided more support for the researchers' conclusion than did either psychology or eye tracking information. Experiment 3.3 also found a marginally significant interaction between neuroscience information and prior behavior. Experiment 3.4 found that neuroscience information no longer increased understanding of the phenomenon when it was non-cognitive; however, neuroscience still increased expectations of replication for people who did not already engage in the behavior, and people were still more likely to indicate that neuroscience evidence provided more support for the conclusion than did behavioral evidence. Experiment 4.1 found that the preference for micro-level information is stable across different research scenarios involving different data collection techniques and different ways of integrating the micro-level information into the research study; additionally, people who had more sophisticated scientific knowledge and thinking dispositions were more likely to display a preference for micro-level information. Finally, Experiment 4.2 showed that this preference for micro-level information results in a decrease in the perceived sample size needed in a research study.

Theoretical Contributions

It is important to note that many of the previous studies looking at the influence of neuroscience have reached inconsistent conclusions, and one implication of my dissertation is that these previous studies may not have focused on the right dependent variables. Previous research has primarily focused on how the information affects perceived quality of the research and scientific reasoning in the article. In my research, I consistently failed to find effects of neuroscience on perceived quality of the research; however, consistent effects emerged for perceived understanding, when the phenomenon being studied was cognitive, and expectations of replication. Additionally, previous research has neglected to control for the type of phenomena (e.g., cognitive vs. non-cognitive) paired with the neuroscience information, which could also partially explain inconsistencies in the literature. For example, Hook and Farah (2013) failed to replicate McCabe and Castel (2008) which investigated the influence of brain images presented in research articles about cognitive phenomena; however, in Hook and Farah's replication attempts, many of the fictional studies were about non-cognitive phenomena. My research suggests that neuroscience information may be most influential for phenomena that are easily associated with the mind.

The present research improves on this previous body of work in a number of other ways, as well. For example, previous studies have not attended to how the neuroscience or other micro-level information is integrated into the fictional research scenarios. As illustrated in Table 5.1 below, some studies used neuroscience only as an explanation, while others used it only as evidence, and some used it as both. Given the inconsistencies across these studies, it was important to address whether the way in which micro-level information is integrated into the research study moderates its influence on reasoning. Additionally, in studies investigating the influence of neuroscience text, researchers have typically compared the influence of a research study with neuroscience text to the influence of a research study with either no additional text at all or no additional text that could be construed as a mechanism. Thus, it is unclear whether the effects of neuroscience in previous studies have been due to something unique about neuroscience information or, instead, due simply to the fact that a plausible mechanism was suggested in the neuroscience condition but not in the control condition. Creating proper control conditions for the micro-level conditions was a significant focus of my dissertation and allowed for a purer assessment of the influence of level of analysis on causal understanding and beliefs about validity.

Table 5.1
Stimulus Features in Previous Neuroscience Studies (re-presented from Chapter 1)

	Features of Fictional Research study				
Paper	Type of Integration	Outcome Variable	Subjects	Comparison of Interest	Outcome Variable(s) of Interest
Weisberg et al. (2008)	Explanation	16 cognitive, 2 non-cognitive	Human	Neuroscience text vs. no additional text	Quality of explanations
McCabe & Castel (2008)	Evidence & Explanation	Cognitive	Human	Brain image vs. bar graph vs. text; brain image vs. topographical maps; Brain image vs. text	Credibility & reasoning
Gruber & Dickerson (2008)	Evidence	Pseudo-cognitive	Human	No image vs. several different types of images, including brain	Credibility & reasoning
Hook & Farah (2013)	Evidence	2 cognitive, 4 non-cognitive	Human	Brain image vs. bar graph vs. control photo	Credibility & reasoning
Diekmann et al. (in press)	Evidence	Non-cognitive	Human	Neuroscience text vs. no additional text	Interestingness
Fernandez-Duque et al. (2014)	Explanation	16 cognitive, 2 non-cognitive (used stimuli from Weisberg, et al. (2008)	Human	Brain image + neuroscience text vs. neuroscience text only vs. no additional text ; neuroscience text vs. social science text vs. no additional text; neuroscience text vs. hard science text vs. social science text	Quality of explanations

Why do people tend to think micro-level is superior to macro-level information? As discussed in previous chapters, one explanation could be that micro-level information sounds technical and scientific, and people could use these superficial features as a cue for quality. However, the fact that people were more likely to think that neuroscience, rather than eye tracking, evidence provided the best support for the researchers' conclusions in Experiment 3.3 suggests that technical jargon cannot be the only explanation. Additional support for this claim comes from Fernandez-Duque, Evans, Colton, and Hodges (in press), who found that the presence of jargon from hard science disciplines – such as biochemistry, biology, and mathematics – in an explanation did not make the explanation seem better than if social science information was present; however, the presence of neuroscience information made the explanation seem significantly better than explanations with social science information.

Another explanation could be that micro-level information appears to provide a better mechanistic explanation for a phenomenon. Support for this possibility comes from the dramatic effect that micro-level information had on perceived understanding in Experiments 2.1 through 3.3. However, as suggested in Experiment 3.4, the perceived mechanistic advantage of micro-level information may only hold when the micro-level methodology is judged to be an appropriate tool for studying the construct of interest. For example, neuroscience, a brain-based methodology, had significant effects on understanding cognitive phenomena, but not an emotional phenomenon. These results suggest that it may be less straightforward for lay readers to understand how neural activity can influence emotions, which folk wisdom tends to ascribe to the heart or the "soul", than how it influences cognitive performance.

Another process that might be going on is that people may have fixed beliefs about what science is supposed to look like. Beginning at an early age, children tend to think that science involves sophisticated equipment. Chambers (1983) asked children to draw a picture of a scientist and found that the stereotypical image of a scientist involves lab coats, eyeglasses, and laboratory equipment such as microscopes, telescopes, and computers. Children were much less likely to draw notebooks and reports, which would depict science more generally as a process of recording knowledge. Chambers found that the stereotypical image of a scientist begins in second and third grade, with more intelligent children forming the image sooner. One factor contributing to these stereotypical representations of science could be the way in which children are taught science. Beginning in pre-school, teachers often set aside a fixed amount of time per day to “do science”, which typically consists of equipment and disjoint activities that do not get integrated into the rest of the day’s activities (Conezio & French, 2002; French & Woodring, 2012). Thus, people tend to think of science as a collection of facts resulting from sophisticated equipment, rather than an approach we use to think about the world, and may fail to appreciate how the scientific process can be applied in everyday lives at multiple levels of analysis. These fixed beliefs about science might result in a bias towards reductionist approaches, such that information that is collected from a lower level of analysis will be perceived as more valid than information that is collected from a higher level of analysis. The deficiency in science process skills has affected the Next Generation Science Standards, which seek to move beyond just content knowledge and place a greater emphasis on the mastery of transferable scientific thinking skills that can be incorporated across disciplines (NGSS Lead States, 2013).

One possibility that can be ruled out is that people are distracted by micro-level evidence. We consistently found that, in Studies 1-4, the presence of neuroscience information did not interfere with participants’ ability to recall important methodological details from the article they read. Although this finding does not rule out the possibility that there is some process below the threshold of awareness that is influencing evaluations, such as a heuristic process, it does suggest that neuroscience is not affecting attentional processes.

An intriguing finding from this research is that, even among scientifically literate samples, people whose scientific literacy was higher than the median were even more likely to show a preference for micro-level information. Such a finding could still be explained in part by the aforementioned "prior beliefs about science" hypothesis, given that these prior beliefs are not necessarily limited to people who are low in scientific literacy. In fact, it is possible that these prior beliefs are even stronger among the scientifically literate. For example, perhaps scientifically literate people have had more education and more exposure to these beliefs about what science is and is not, resulting in a kind of confirmation bias (Nickerson, 1998), while people with less scientific knowledge might approach the task with more naïveté. It is also possible that people with more scientific literacy were more familiar with the terminology used and scientific domains represented, resulting in increased ratings based on familiarity, compared to people with lower scientific literacy whose lack of knowledge or familiarity may have resulted in lower ratings. However, this explanation would not necessarily be consistent with our finding that people who are lower in scientific literacy rate their perceived understanding higher than do those who are higher in scientific literacy, although possibly this latter finding could be explained by a self-enhancement effect (Paulhus, 1998).

Although distinguishing between these potential processes is outside the scope of this work, this research has provided information on which processes are more or less likely candidates and which ones should be explored in future work. Additionally, this latter set of findings strongly suggests that the preference for micro-level information is deliberative – that is, people are consciously deciding that micro-level information appears more valid than macro-level information. It would not be correct to say that this conscious valuation of micro-level over macro-level information is irrational; as previously mentioned, there may be well-founded reasons why people believe micro-level information to be superior. However, the fact that people are expressing this value provides insight into lay people's assumptions about what a scientific process is and what makes evidence valid.

Practical Implications

Given that the lay public encounters scientific evidence online, when a number of contextual factors may be present, an important implication of my work pertains to the influence of these contextual factors on the preference for micro-level information. The first implication is that one's current or prior behavior plays a significant role in determining how people are influenced by micro-level information that is relevant to that behavior. Much research has examined how prior beliefs affect the way people reason about evidence and found that people tend to be more critical of evidence that contradicts their prior beliefs, and less critical of evidence that confirms their prior beliefs (Lord, Ross, & Lepper, 1979; Klaczynski, 2000). Similarly, my research consistently demonstrated that people with congruent prior beliefs evaluated the fictional research studies more favorably than did people with neutral or incongruent beliefs (Experiments 2.1 through 3.4). Interestingly, interactions between the condition (neuroscience vs. no neuroscience) and prior behavior showed that people who already engage in the behavior were less influenced by neuroscience information, whereas people who did not yet engage in the behavior were influenced by the neuroscience information (Experiments 3.1, 3.2, and 3.4). These results suggest that when a newspaper headline touts a new behavior or product unfamiliar to people, coupled with neuroscience or other micro-level information, people may be especially likely to be influenced by it, particularly if the micro-level process is seen as directly relevant to the outcome variable.

Another implication from my research is that we have a better understanding of the ways in which micro-level preferences can manifest themselves in everyday reasoning and decision-making scenarios. Specifically, people are more likely to expect that a study will replicate when the neuroscience information is included (Experiments 3.1, 3.2, and 3.4), and people are willing to accept smaller sample sizes when micro-level information is present (Experiment 4.2). Taken together, people tend to believe conclusions more when they are based on micro-level evidence. This finding may help explain the tension that is often created when reductionist approaches seem to be overemphasized, typically by funding agencies, in fields where a complex systems approach is often more beneficial. As noted in Breckler (2006), a clear example of the implications of micro-level preferences has played out at the National Institute of Mental Health, who previously championed the biopsychosocial model of mental health but has recently shifted its focus to neural and genetic approaches, even though, for many mental health issues, balanced approaches are more likely to lead to significant advances. For example, research on interpersonal relationships and social support is critical for understanding how to manage many mental health issues, and it would be inappropriate to assume that micro-level approaches could provide more insight into the management of mental health issues than macro-level approaches. In other words, the assumption that micro-level information is superior to macro-level information could lead to the dismissal of important environmental factors that could moderate, or even give rise to, physiological processes. Another, less direct, consequence of an overemphasis on reductionism is that many reductionist techniques are expensive, which often limits the number of subjects that can be included in a study, which can lead to an overabundance of false positives in the scientific literature (Button et al., 2013; Macleod, 2011). However, given the perceived validity of micro-level evidence demonstrated in the present work, people may not immediately appreciate these caveats to the extent they would for macro-level evidence.

Final Remarks

I have provided evidence that people show consistent preferences for micro-level information across a variety of behavioral phenomena, and it appears that people arrive at these preferences through deliberate strategies. These findings suggest that people think about macro- and micro-level evidence in fundamentally different ways, with a tendency to value micro-level evidence over macro-level evidence. While there may be some rational basis for this differential valuation, I have also illustrated some of the implications this may pose for scientific fields, particularly ones that seek to understand complex systems in which important factors may operate at many different levels of analysis. As suggested by the Next Generation Science Standards, one strategy for improving the way that people reason about scientific evidence may be to place a larger emphasis on the *process* of science, and encourage children from a young age to see science as strategic approach for understanding the phenomena they encounter in their everyday lives, regardless of the level of analysis.

References

Breckler, S. J. (2006). The newest age of reductionism. *Monitor on Psychology*, *27*(8), 23.

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., & Robinson, E. S. J. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*, 365-376.

Chambers, D. W. (1983). Stereotypic images of the scientist: The Draw-A-Scientist Test. *Science Education, 67*(2), 255-265.

Conezio, K., & French, L. (2002). Science in the preschool classroom. *Young Children*, *57*(5), 12-18.

Fernandez-Duque, D., Evans, J., Colton, C., & Hodges, S. D. (in press). Superfluous neuroscience information makes explanations of psychological phenomena more appealing. *Journal of Cognitive Neuroscience*.

French, L. A., & Woodring, S. D. (2013). Science education in the early years. In O. N. Saracho & B. Spodek (Eds.), *Handbook of research on the education of young children* (3rd ed., pp. 179-196). New York, NY: Routledge.

Hook, C. J., & Farah, M. J. (2013). Look again: Effects of brain images and mind-brain dualism on lay evaluations of research. *Journal of Cognitive Neuroscience, 25*, 1397-1405.

Klaczynski, P. A. (2000). Motivated scientific reasoning biases, epistemological beliefs, and theory polarization: A two-process approach to adolescent cognition. *Child Development, 71,* 1347-1366.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, *37*, 2098-2109.

Macleod, M. (2011). Why animal research needs to improve. *Nature,* 477, 511.

McCabe, D. P., & Castel, A. D. (2008). Seeing is believing: The effect of brain images on judgments of scientific reasoning. *Cognition, 107,* 343-352.

NGSS Lead States. (2013). Next Generation Science Standards: For States, By States.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175.

Paulhus, D. L. (1998). Intrapersonal and intrapsychic adaptiveness of trait self-enhancement: A mixed blessing? *Journal of Personality and Social Psychology, 75*, 1197-1208.