Commentaries and Rejoinder to “Trade-Off Aversion as an Explanation for the Attraction Effect: A Functional Magnetic Resonance Imaging Study”

Integrating Neural and Decision Sciences: Convergence and Constraints

SCOTT A. HUETTEL and JOHN W. PAYNE*

One of the most remarkable aspects of modern science is its diversity. A given phenomenon can be studied simultaneously within a range of different disciplines, each with its own methodologies and forms of explanation. Most of the time, these different disciplines proceed apace, without clear avenues of communication. However, in recent years, there has been growing interest in bridging the methods of neuroscience with behavioral studies of decision-making phenomena, leading to the burgeoning interdisciplinary of neuroeconomics.

In their target article, Hedgcock and Rao (2009) describe a novel and integrative neuroeconomic study of the “attraction effect” (i.e., preference for one option in a choice set increases following the introduction of a new option that it dominates). The attraction effect, also known as the “asymmetric dominance effect” (Huber, Payne, and Puto 1982), is one of several important context effects in decision making that can lead to violations of basic economic choice principles, such as regularity (Luce 1977). For example, a consequence of the attraction effect is an increase in the preference for an item in a choice set following the addition of another choice option. Hedgcock and Rao note several potential explanations for this context effect: The dominated option could change decision weights, alter the reference point for choice, or allow the use of a heuristic decision strategy. Each of these could be consistent with the existing behavioral data, at least in the types of simple choice scenarios typically studied, but they evoke different perspectives about the underlying mechanism. In particular, the idea that the attraction effect reflects a transition from a compensatory trade-off-based strategy to a more heuristic strategy that avoids emotional trade-offs (e.g., the lexicographic rule) leads to the strong prediction that there should be a concomitant attenuation of negative emotions, which in turn could be measured using the techniques of neuroscience. The primary goal of this article was to evaluate whether neuroscience data could support or invalidate one postulated explanation for the attraction effect.

Thus far, most neuroeconomic research has taken a relatively brain-centric form. Researchers have used techniques from behavioral economics and the decision sciences to improve the understanding of brain function. Some recent successes include identification of brain systems associated with reward evaluation (Delgado et al. 2000; Knutson et al. 2005; Schultz, Dayan, and Montague 1997), risk and uncertainty (Hsu et al. 2005; Huettel, Song, and McCarthy 2005), and other-regarding preferences (Harbaugh, Mayr, and Burghart 2007; Tankersley, Stowe, and Huettel 2007). Hedgcock and Rao take the opposite approach. They use the techniques of neuroscience to improve the understanding of decision phenomena. As we discuss in this commentary, moving from brain to behavior poses a specific set of challenges, both scientific and sociological, that have limited progress in the past. We consider these challenges in the following sections, ending with some speculations about the future interactions between neuroscience and marketing research.

NEUROESSENTIALISM VERSUS BEHAVIORAL SUFFICIENCY

What does neuroscience data reveal about cognition and behavior? One attitude, which is frequently observed in popular accounts of the latest neuroscience studies, is that measurements of brain function provide access to previously hidden mechanisms. In lay conceptions, measuring the brain is like opening the hood of a car, with previously inferred components now laid bare to the viewer. This attitude has been labeled “neuroessentialism” (Racine, Barlil, and Illes 2005). The rise of neuroessentialism can be attributed to several factors. First, there has been a remarkable growth since the early 1990s in the prevalence and accessibility of neuroimaging techniques, notably functional magnetic resonance imaging (fMRI). Accompanying this growth has been an increased breadth of inquiry, as

*Scott A. Huettel is Associate Professor of Psychology and Neuroscience, Center for Cognitive Neuroscience (e-mail: scott.huettel@duke.edu), and John W. Payne is Joseph J. Ruvane Jr. Professor of Management and Marketing, Fuqua School of Business (e-mail: jpayne@duke.edu), Duke University. The authors thank colleagues in the Duke Center for Neuroeconomic Studies for stimulating discussions about integrating the neural and decision sciences. The Duke Institute for Brain Sciences provided support for those discussions and for our joint research.

© 2009, American Marketing Association
ISSN: 0022-2437 (print), 1547-7193 (electronic)
researchers explore increasingly abstract and subjective topics from optimism bias (Sharot et al. 2007) to mind wandering (Mason et al. 2007). Second, neuroscience techniques seem to be both technological and mysterious; in contrast, behavioral research methods can appear simplistic and dated. Furthermore, neuroscience data (again, particularly fMRI) can be easily misinterpreted. As Hedgcock and Rao note, people often interpret the color maps from an fMRI scan as if they were photographs rather than the output of a complex series of statistical tests.

An alternative perspective, as adopted by commentators decrying the expansion of neuroscience methods into the social sciences, is that neuroscience data are unnecessary for understanding behavior. We refer to this as the “behavioral sufficiency” argument (Clithero, Tankersley, and Huettel 2008). According to this line of thinking, knowing the brain regions activated during performance of some task might be of interest to brain researchers, but such knowledge is of no consequence for behavioral research. Instead, to understand a behavioral phenomenon (e.g., the attraction effect), the context in which the phenomenon occurs should be manipulated, and then changes in the associated behavior should be measured. Proponents of behavioral sufficiency contend that neuroscience data, while interesting to those who study the brain, are irrelevant for studies of behavior.

We believe that these two perspectives are equally misguided. Neuroscience data cannot simply replace behavioral research, despite the breathless descriptions in the popular press, nor should behavioral researchers dismiss neuroscience data as irrelevant to their goals. An encouraging aspect of Hedgcock and Rao’s work is how they tread a careful middle ground between these extremes. Particularly notable is how they motivate their neuroscience research. They do not attempt simply to identify the “neural correlates” of the attraction effect; those kinds of exploratory studies may (rightfully) be viewed as being of interest only to neuroscientists. Instead, they construct specific and well-formed hypotheses about brain function based on prior behavioral theories. In effect, neuroscience data become an operational proxy for the cognitive processes postulated in prior studies of trade-off decisions (e.g., Luce, Bettman, and Payne 2001). In summary, the value of the authors’ approach is that they use neuroscience data to test predictions derived from behavior. This is a major strength of the article.

Yet the authors themselves sound some notes of caution, especially about the relative crudeness of functional neuroimaging methods compared with the vast complexity and interactivity of the human brain. Although we appreciate their intentions, we believe that these points are overly negative. Within just the past few years, functional neuroimaging has become increasingly sophisticated, especially in experimental design and data analysis. There are now many elegant approaches for understanding interactions between brain regions that, together, support complex cognitive processes, such as when monetary rewards shape memories (Adcock et al. 2006) or when people repay another’s trust in an economic game (King-Casas et al. 2005). New analysis methods have already enabled researchers to map out functional connectivity between sets of regions, even at the relatively poor temporal resolution of most fMRI studies, and further technological and algorithmic improvements are likely. However, caution is justified in the kinds of inferences that can be drawn from fMRI research; we consider such inferences in greater detail in the following section.

Forward Versus Reverse Inference

As indicated previously, we are particularly enthusiastic about how Hedgcock and Rao construct their experiment to test hypotheses about brain function. Such hypotheses are too-frequently absent from the neuroscience literature, and thus the authors’ efforts to make specific predictions is laudable. However, this approach has its challenges, one of which reflects an area of active debate within neuroscience: the use of reverse inference.

Most cognitive neuroscience studies, whether using fMRI or another technique, reason from behavior to brain: “Because my experimental conditions differ in their engagement of cognitive process X (e.g., emotional arousal), the observed activation in brain region Y (e.g., the amygdala) can be attributed to that cognitive process.” This style of reasoning, which is called “forward inference,” enables neuroscientists to create maps of cognitive functions within the brain. The hypotheses that Hedgcock and Rao test require reversing the normal direction of inference. Consider their H1, which could be rephrased as follows: “We expect to observe reduced activation in brain region Y (i.e., the amygdala), which is associated with process X (i.e., negative emotion), if our manipulation (i.e., the introduction of a decoy) leads to reductions in that process.” In effect, changes in the brain are to be used as evidence about the cognitive processes associated with the task.

As outlined in a compelling review by Poldrack (2006), reverse inference can be justifiable but only under specific conditions. Consider H1 from Hedgcock and Rao’s article. It assumes that activation in the amygdala can be used as an index of the processing of a negative emotion; represented another way, p(negative emotion|amygdala activation) is assumed to be high. We can evaluate the reasonableness of this assumption on the basis of prior knowledge: The amygdala is a small region, its activation is reliably evoked by negative emotional stimuli, and relatively similar stimuli (e.g., positive emotional stimuli) evoke much less activation. Although a comprehensive review is well outside the scope of this commentary, we believe that the prior specificity of amygdala activation makes their H1 reasonable.

Reverse inference is less justifiable for the other hypotheses. Their H2 argues, in part, that activation in the dorsolateral prefrontal cortex (DLPFC), which comprises approximately 15% of the entire brain, can be used as an index of the use of a rule-based selection process. For this to be valid, p(rule-based selection|DLPFC activation) must be high. Although many studies show rule-based selection evokes activation in this region—and thus, p(DLPFC activation|rule-based selection) is likely—there are also many other putative cognitive processes that lead to DLPFC activation: maintaining and manipulating information in working memory, shifting attention, and inhibiting irrelevant stimuli, among others. Thus, compared with the amygdala, DLPFC is much less selective in its function. (Many researchers are investigating functions associated with specific parts of DLPFC, which could improve the
specificity of associated inferences.) Thus, changes in DLPFC activation provide much weaker evidence for or against particular theories about behavior. For comparison, their H₃ is associated with an intermediate level of specificity: The anterior cingulate cortex (ACC) is indeed often associated with “response conflict,” but whether that label accurately describes the computations supported by ACC remains an area of intense research.

Despite these limitations, which are hardly fatal flaws given the novelty of this research area, we remain enthusiastic about this overall approach. Decision scientists and neuroscientists should collaborate along the lines of Hedgcock and Rao’s study—that is, identifying neural markers of specific brain functions and using those markers to test hypotheses about the mechanisms underlying decision phenomena.

CONVERGENCE OF NEURAL AND BEHAVIORAL RESEARCH

We are particularly excited using converging neuroscience and behavioral data to understand variability in decision making. As any behavioral experimentalist knows, whether a marketing researcher or a behavioral economist, even the most robust decision phenomena are subject to variability. In Hedgcock and Rao’s target article, participants exhibited a robust bias toward the decoyed option but still chose the nondecoyed option approximately 40% of the time. Thus, two questions about variability are possible: What makes a given person choose one option or another, despite an overall strategic bias? And what causes people to differ in their strategic tendencies?

Such questions about variability remain largely outside the purview of behavioral decision science, in part because of methodological factors (e.g., relatively few trials per participant, use of group testing). However, within any fMRI experiment, there is a welter of data specific to each person, as reflected in the magnitude of activation within and across regions of interest. Hedgcock and Rao show not only that the amygdala response is reduced when a decoy is present but also that participants who are more influenced by the presence of the decoy show less amygdala activation. This result provides compelling evidence that the observed neural differences are indeed related to the attraction effect and not to some other aspect of the experimental stimuli or task. To extend these findings, neural data (e.g., activation in regions of interest), along with behavioral data, could be used to predict trial-to-trial choices within a logistic regression model, as has been done for other domains of decision making (Chiu, Lohrenz, and Montague 2008; Kuhnen and Knutson 2005). Further research might want to collect more behavioral individual difference measures that would be related to the hypotheses. Again, one of the strengths (and, indeed, a requirement) of fMRI experiments is the wealth of data that must be collected from each participant.

Consistent with the work of Hedgcock and Rao, we suggest three simple guidelines for integrating neuroscience and behavioral data. First, the neuroscience experiments should be designed so that they test specific hypotheses about brain function; those hypotheses must be made as precise as possible, given the concerns about reverse inference we advanced in the previous section. Second, and relatedly, the research program should attempt to use neuroscience data to identify constraints for behavioral models, so that models that are biologically implausible (and, thus, candidates for revision) can be identified. Third, the results of the neuroscience experiments should motivate new studies of behavior to confirm the revised models. This iterative approach—that is, moving from behavior to brain and back to behavior—recognizes the converging value of both forms of data without privileging either.

CONCLUSIONS

Given the increasing number of studies that bridge the methods of neuroscience with behavioral studies of decision-making phenomena, we conclude with one final point. Robust decision phenomena, such as context-dependent preferences, are likely to be robust precisely because they are multiply determined. Context may matter for a host of reasons—for example, a desire to minimize emotion-laden trade-offs (Luce 1998) and anticipation of justification of decisions (Simonson 1989), among many others. When using an exciting new method, such as fMRI, researchers must be careful not to overgeneralize from one set of brain imaging results to conclude that the one cause of a phenomenon has been found. As we indicate in this commentary, we are enthusiastic about Hedgcock and Rao’s research, which we agree supports the idea that a reduction in negative emotions is an important motivator for choice of a decoyed alternative. However, we suspect that Hedgcock and Rao share with us the view that changes in emotional response may be but one of several contributors to the attraction effect.

REFERENCES

Using fMRI to Inform Marketing Research: Challenges and Opportunities

CAROLYN YOON, RICHARD GONZALEZ, and JAMES R. BETTMAN*

Hedgcock and Rao (2009; hereinafter, HR) make several distinct contributions. First and foremost, they raise and attempt to resolve several issues surrounding the interrelationships of (negative) emotions, task complexity, and consumers’ tendency to avoid cognitively demanding decisions. They also illustrate the power and insight afforded by recent innovations in neuroimaging as well as some of the practical pitfalls in using the methodology to constrain different process-based explanations for empirical phenomena. In this commentary, we focus on several aspects of HR’s findings and methods. In particular, we discuss issues related to the conceptualization of research questions, study design, and interpretation of neural data. However, our primary goal is to use their research as a springboard to inform future studies and “best practices” for research in marketing employing neuroimaging technologies.

RESEARCH CONCEPTUALIZATION

Behavioral research demands tight functional definitions of core constructs, even more so for neuroimaging studies, in which such constructs determine which brain regions will be examined for activations or deactivations in either whole-brain or region-of-interest analyses. Therefore, it is somewhat troubling that several key definitions in HR must be gleaned, if at all, from context. Chief among these are “trade-off choice set” and “decoy.” For example, a key idea running throughout HR is that the presence of a (specially defined) decoy in a trade-off choice set reduces experienced trade-off difficulty. This suggests that a three-item choice set without a decoy would show increased difficulty. Because HR do not examine such sets (i.e., more general three-item choice sets that vary the critical property of the decoy), this idea is not examined directly. An especially useful feature for further research would be the use of explicit protocols, behavioral measures, self-reports, or neuroimaging data to determine the conditions under which three-item choice sets influence task complexity and negative emotion; doing so would shed light on the otherwise perplexing finding that HR’s observed pattern of neural activity was seemingly identical for two types of decoy: “inferior” and “dominated” decoys. Similarly, trade-off-type sets are defined implicitly, and then only toward the end of the article, when not involving a trade-off is taken to mean that the structure of the choice set enables a dominant option to be identified. Yet it remains unclear whether the target is any more psychologically dominant in the three-item choice set than in the two-item one. Resolving such definitional issues seems central to supporting HR’s main claims and determining their range of applicability.

Related to this point, existing behavioral research could be used to postulate specific effects. For example, in a careful set of studies, Wedell and Pettibone (1996) show that a dominating option appears to receive “added value” from being better than the decoy. Do HR find any evidence consistent with such an effect? If not, why?

Task-based neuroimaging research also benefits from an explicit objective function. For example, HR rely on the Stroop task to identify certain region-of-interest targets and related hypotheses. In Stroop tasks, participants know what constitutes a correct answer, which would presumably eliminate post hoc uncertainty, dissonance, or conflict. In contrast, participants in HR’s choice task were asked to select an option on the basis of unstated criteria. Was their activity was seemingly identical for two types of decoy: “consistent with such an effect? If not, why?

---

*Carolyn Yoon is Associate Professor of Marketing, Stephen M. Ross School of Business (e-mail: yoon@umich.edu), and Richard Gonzalez is Professor of Psychology, Department of Psychology (e-mail: gonzo@umich.edu), University of Michigan. James R. Bettman is Burlington Industries Professor, Fuqua School of Business, Duke University (e-mail: jrb12@duke.edu). The authors thank Fred Feinberg for his helpful comments and suggestions on a previous draft of this commentary.
task lacked an externally mandated objective or criterion for correctness, it differs fundamentally from tasks, such as the Stroop, in which the role of the affective system may be more limited. Such considerations may make it difficult for researchers to know whether participants were heterogeneous in the explicit performance standards they chose to bring to the task and, thus, to know which region of interest should be implicated.

When HR state (p. 3) that “consumers will prefer making comparisons between options that do not represent trade-offs,” they implicitly raise a vital issue in distinguishing preference from affect. It may be possible that the kind of (negative) emotion underlying their theory is highly related to or virtually synonymous with (decreased) preference. Neuroimaging studies would need to determine whether patterns of neural activations associated with preference formation and expression were strongly related to those for affective response; HR’s discussion refers almost exclusively to the latter, providing another avenue for future investigation.

STUDY DESIGN

Hedgcock and Rao are to be commended for taking such painstaking account of heterogeneity by using consumer-specific “equivalents.” This effortful phase of the research is critical in neuroimaging studies to customize stimuli across (typically) small samples to which heterogeneous (random coefficient) models will be applied. Wedell and Pettibone’s (1996) method seems especially apt. However, one of HR’s contentions—that “inferior” alternatives were truly inferior—depends (though this is not addressed in the article proper) on estimating heterogeneous trade-off functions, which are strongly model dependent; related methods from conjoint analysis could be brought to bear in future studies for this purpose.

Another of HR’s core assumptions is stated (p. 4) clearly as follows: “To the extent that larger set sizes yield increased computation…. This stands in contrast to the great deal of theorizing in marketing, going back at least to Shugan’s (1980) celebrated “cost of thinking,” in which task complexity or difficulty was measured in terms of this type of computation (specifically, a “tournament” of option-by-option attribute-based comparisons). By that standard, task difficulty is necessarily greater in the larger choice set, regardless of whether it involved a decoy, but in HR, it is posited to be lower. What kinds of (neural) data could be brought to bear to settle this important practical and theoretical issue? One possibility is response latencies: All else being equal, difficult decisions should take longer. It is unfortunate that such latencies are not reported by HR, because they are available in the scanner record for each participant.

Researchers intending to use neuroimaging techniques in marketing need to understand that they are dealing with intrinsically correlational measures: Neural activity can accompany a phenomenon without any causal connection. A possible example in HR involves the following question: Can cognition simply give rise to negative emotion? For example, if participants become frustrated when making so many effortful trade-offs, the appropriate conceptual model may be “cognition → decision and emotion” versus “cognition and emotion → decision.” The data in HR are consistent with either of these directed (i.e., causal) scenarios, but their interpretation favors the latter, which would require a direct causal line from (negative) emotion to decision. It may be possible to approach this central issue using causal modeling, which thus far is relatively rare in neuroimaging studies. However, this problem is at the heart of a long-standing debate in the field of emotions (e.g., Zajonc [1980] versus Lazarus [1984]), so it is not clear that one study will provide a complete solution, let alone a study using fMRI technology that does not offer the necessary time resolution to distinguish such psychological processes.

One of the authors’ operational assumptions highlights the preciousness of participants in fMRI research. They reasoning that any deviation from indifference (a 50% share) suggests an attraction effect. Such reasoning would not be necessary had there been a control condition, one in which exact indifference values—as stated by the participants themselves—were used. Because doing so would have squandered participants as controls, a critical assumption underlying the assessment of effect strengths went untested. Based on the existence of so-called low heuristic processors, there is reason to believe that this assumption may be unfounded. Given the number of participants, perhaps using a continuous measure of the proportion of times the target was selected would yield more power than arbitrarily splitting at 50% and would enable correlations with activation levels to be calculated.

DATA INTERPRETATION

Hedgcock and Rao’s treatment suggests certain “best practices” for research relying on neuroimaging. Perhaps chief among these is the importance of specifying a comprehensive set of associations and dissociations, not merely a few expected activations. That is, when examining a specific theory through region-of-interest analyses, researchers must comprehensively consider all brain regions that should and should not display activity. Otherwise, all else being equal, stimuli affecting more brain regions will appear to support a wider range of theories than those affecting fewer. Another general conundrum affecting much neuroimaging research, including HR’s, is the assumption that if some quality has been implicated in a particular brain region, activation in that region suggests that the quality is in play. However, this could be the converse of the appropriate interpretation. For example, HR claim that activation in the DLPFC and ACC suggests a decreased role of heuristics and reduced conflict. This is possible, but it is equally suggestive of any of the stimuli that have been implicated in activating those regions. We stress again that the only ways around this general problem are comprehensive surveys of (1) all stimuli known to affect a particular region and (2) all associations and dissociations attributable to the specific stimulus in question (see Berman, Jonides, and Nee 2006). Hedgcock and Rao’s conclusions sound a cautionary note for neuroimaging researchers. Specifically, they claim that the avoidance of negative emotion (and trade-off aversion) is a “compelling explanation” for the attraction effect. What they have succeeded in showing, however, is that negative emotion may accompany whatever is causing the attraction effect. That the amygdala effects for the critical trial conditions are deactivations (rather than activations) in
relation to the fixation across conditions suggests an alternative explanation that a larger (three-item) choice task is more cognitively demanding than a two-item choice task. To provide one example, Drevets and Raichle (1998) review several studies to examine the interaction between emotion and cognition and report that more cognitively demanding tasks are associated with deactivations in the amygdala and activations in the ACC and DLPFC, similar to the pattern of results that HR obtain. From HR’s study, it seems difficult to rule out Drevets and Raichle’s explanation. It would be important for further research to examine this issue. Hedgcock and Rao’s interpretation of the ACC activation in terms of response conflict or greater use of controlled processing resources in the three-item choice sets would be a new finding related to asymmetric dominance that could lead to new behavioral studies.

Researchers need to be aware of the difficulty of separating causal and correlational explanations, which can be exacerbated in neuroimaging settings when self-report and other process measures are not collected. It is also necessary to question whether negative emotion is lower when choice sets are enriched with a decoy, because the only comparisons carried out were between two-item and three-item sets. As we mentioned previously, a compelling demonstration would compare three-item sets with a decoy with three-item sets without a decoy (the decoy is a special third option with particular properties) to determine whether the presence of a decoy indeed drives these results. Subsequent analysis splitting the sample in terms of low heuristic usage does not resolve this basic problem. There is no reason to believe that participants who chose the target less than 50% of the time were low heuristic processors. According to the theory, with no decoy, participants should be indifferent. The presence of a decoy should never sway participants to the competitor; such an occurrence is not a case of low heuristic processing but rather a failure to conform to the basic theory of asymmetric dominance. That is, these participants simply failed to respond to the decoy in the posited way. Hedgcock and Rao are wise to caution readers to avoid overinterpreting neuroimaging data and claiming “explanatory depth.” We echo these concerns and suggest that a definitive demonstration of the role of emotions in the attraction effect can be built on the suggestive foundation supplied by HR.

CONCLUSIONS

As HR point out, several aspects of fMRI make it an attractive methodology for the purposes of better understanding conditions and processes related to marketing and consumer domains. It enables the measurement of participants’ psychological processes while they occur and can supplement self-reports, which may reflect biased responses. However, there are several potential pitfalls associated with the use of fMRI. An important consideration is the appropriateness of the fMRI methodology given the research questions. Because neuroimaging techniques are used to identify correlates of underlying processes, it is often difficult to interpret the results without proper conceptualization and operationalization before the fMRI experiment; that is, it is essential that researchers develop an ex ante neurobiological model for the investigated psychological phenomena, along with the expected neural processes and correlates. In addition, behavioral researchers should not lose sight of their behavioral rigor in the rush to try out fMRI techniques on their favorite paradigm. The behavioral portion of the imaging study should stand on its own, with the fMRI component merely adding an additional variable to the paradigm. Finally, fMRI should not be used as a stand-alone methodology. Rather, researchers should seek convergent validity by linking fMRI data to other behavioral measures (e.g., self-report, choice, purchase behavior, reaction time, eye-tracking data) and taking account of the critical roles of careful experimental design, rigorous hypothesis generation, abundant trials, and large-enough samples to ensure sufficient explanatory power.

Again, our goal in this commentary has been to point out some general themes that marketing researchers should be mindful of when they consider bringing fMRI techniques into their behavioral paradigms. We remain optimistic about the future role of brain imaging techniques, and neuroscience more generally, in marketing. As with the addition of any new variable into a research paradigm, there are many considerations. We applaud HR for a solid contribution to the fledging use of brain imaging techniques in standard marketing research.

REFERENCES


Aristotle’s Anxiety: Choosing Among Methods to Study Choice

WILLIAM HEDGCOCK and AKSHAY R. RAO*

According to a thought experiment described by Aristotle, a person “who, though exceedingly hungry and thirsty, and both equally, yet being equidistant from food and drink, is therefore bound to stay where he is” might consequently waste away for want of food and drink (Stocks 1922). The problem of choosing between two equally desirable options is analogous to the trade-off problem a decision maker faces when confronted with two equally attractive options. Choosing from such a choice set is difficult because choosing one option may involve giving up another attractive option, and the trade-off difficulty the decision maker experiences may result in conflict in his or her mind, which can generate negative affect. The introduction of a dominated alternative (the “decoy”) into such a choice set may mitigate the problem by allowing the decision maker to focus on the simpler choice set comprising only the dominating option (the “target”) and the decoy. Therefore, deciding whether the choice set was consistent with this argument.

The two commentaries by Huettel and Payne (2009) and Yoon, Gonzales, and Bettman (2009) offer several thoughtful, provocative, and constructive observations about our work, the particular theoretical issue we examine, and the use of neuroscientific techniques to study consumer behavior. In this rejoinder, we first provide a reprise of our conceptual argument. We then identify three issues raised in the commentaries that we deemed to be particularly noteworthy because they were issues with which we struggled while conceptualizing, designing, and executing our research. Our discussion of these issues allows for an elaboration of theoretical, methodological, and philosophical issues that should inform future studies. We conclude by returning to the metaphor of Aristotle’s anxiety, which is a particularly apt metaphor to employ when choosing research methodologies to study consumer choice.

REPRISE

Huber, Payne, and Puto (1982) demonstrate that the choice shares of an option (the target) in a two-item choice set can increase following the introduction of a third, irrelevant option (the decoy) that is dominated by the target. This attraction effect is a robust finding in the consumer behavior and allied literature streams, but it appears to be multiply determined. Several explanations for the effect have been offered, including a change in the weight attached to the attribute on which the target and decoy perform well (Ariely and Wallsten 1995), the ability to justify choice (Simonson 1989), negative emotion associated with difficult choices (also referred to as “trade-off aversion”; see Luce 1998; Luce, Bettman, and Payne 2001), and so forth. In Hedgcock and Rao (2009), we focused on an examination of the negative emotion explanation for the attraction effect. This explanation was relatively new (having been directly examined in only one study; see Luce 1998), was intuitively appealing, and seemed to be particularly amenable to examination using the cognitive neuroscientific approach that we wanted to employ.

In our study, our core focus was on whether the choice between two equally (un)attractive options might yield cognitive conflict, which in turn might generate negative emotion that would manifest as heightened amygdala activation, compared with a choice problem comprising the same choice set enriched with a third, dominated option. Our empirical setting comprised a complex amalgam of stimuli, and our analyses incorporated both behavioral (i.e., choice share) and cerebral activation data.

In their commentary, Huettel and Payne conclude (p. 16) that we “show not only that the amygdala response is reduced when a decoy is present but also that participants who are more influenced by the presence of the decoy show less amygdala activation. This result provides compelling evidence that the observed neural differences are indeed related to the attraction effect and not to some other aspect of the experimental stimuli or task.” Our principal prediction—that choice sets comprising trade-offs are accompanied by greater negative emotion than choice sets enriched by a decoy—appears to have been supported.

AREAS OF DEBATE

In their commentaries, Huettel and Payne and Yoon, Gonzales, and Bettman raise several observations, three of which we believe are of particular significance. They are (1) the possibility of a rival explanation for our results (Yoon, Gonzales, and Bettman), (2) employing reverse inference to formulate hypotheses (Huettel and Payne), and (3) distinguishing between correlation and causation (Yoon, Gonzales, and Bettman, see also Huettel and Payne).1

Rival Explanation

Yoon, Gonzales, and Bettman are less sanguine about our conclusions than Huettel and Payne, observing that our

1Several other issues were raised in the commentaries, but in the interest of brevity, we do not address them here. For example, Yoon, Gonzales, and Bettman request response latencies (response time for two-item choice sets was shorter for three-item choice sets [8167 milliseconds < 10121 milliseconds, p < .001]). The data and accompanying discussion were removed from Hedgcock and Rao (2009) at the behest of the review team. We concur with that decision because, unlike Yoon, Gonzales, and Bettman, we are not persuaded that more difficult decisions always take more time or that quick but difficult decisions will not generate negative affect. Similarly, our decision not to impose an “objective function” that would specify a correct answer was consistent with the tradition in attraction effect research. Interested readers may contact the authors for further details.
premise that “the presence of a (specially defined) decoy in a trade-off choice set reduces experienced trade-off difficulty” (p. 17) was not directly tested, because we did not compare a three-item choice set that included a dominated decoy with a three-item choice set that included a nondominated decoy. They cite (p. 19) Drevets and Raichle (1998), who “review several studies to examine the interaction between emotion and cognition and report that more cognitively demanding tasks are associated with deactivations in the amygdala…. From [Hedgcock and Rao’s] study, it seems difficult to rule out Drevets and Raichle’s explanation.”

Although Yoon, Gonzales, and Bettman’s assertion that we did not compare three-item choice sets that incorporated or did not incorporate a decoy is correct, this omission does not invalidate our conclusion regarding the implication of amygdalar activation, for two reasons. First, as Huettem and Payne note, we compared participants who examined identically structured choice sets but displayed the attraction effect to a greater or lesser degree, and we observed the predicted differences in amygdalar activation. In the interest of parsimony, any alternative explanation should explain both sets of findings in our study—that amygdalar activation is (1) higher under the trade-off condition and (2) higher for participants who display the attraction effect. Second, a careful reading of Drevets and Raichle (1998) indicates that their conclusions are not inconsistent with our thesis. They note (p. 370) that a reduction in regional cerebral blood flow in the amygdala when engaged in other tasks (e.g., processing a third option, as in our study) “may thus relate to an attentional mechanism in which the neural-processing resources become allocated across systems as attentional demands rise…. Deactivation of emotion-related areas may reflect a relative reduction in the processing resources devoted to emotional evaluation.” To draw a parallel, “in the course of bereavement, grief can be temporarily interrupted by occupation with the attentionally demanding cognitive activities required to solve work-related tasks” (Drevets and Raichle 1998, p. 370). Therefore, our findings of reduced amygdalar activation accompanied by reciprocal activation in other brain regions when a person is engaged in additional processing simply indicate that when the person is distracted by the need to engage in additional cognitive processing, decreased negative emotional processing occurs. However, our core claim that choices based on trade-offs are relatively aversive is not in jeopardy.

Substantively, the negative emotion–driven amygdalar activation observed in the two-item choice set is suppressed in the three-item choice set, perhaps because of the employment of a different choice process (e.g., the employment of a heuristic) or perhaps because resources are drawn away from the amygdala and diverted to other brain regions because of the need to engage in additional cognitive operations. The implication of this reasoning is that though we have evidence that trade-offs generate negative emotion, it is less clear whether the reduction in negative emotion due to the introduction of a decoy is driven by the invocation of an alternative choice process, the redirection of resources, or some other process. This implication is one fruitful consequence of Yoon, Gonzales, and Bettman’s critique, which suggests that the emotional and cognitive systems could share a common set of resources. That is, when higher-order cognitive tasks are engaged in, emotional activation may be suppressed. As Drevets and Raichle (1998, p. 356) observe, “[t]he deactivation of particular (brain) regions may be consistent with a ‘limited capacity’ model of cognitive processing, in which the excessive amounts of information available to the brain necessitate a variety of attentional mechanisms that select among competing mental processes.” This speculation would indicate that the depletion of resources (through the performance of a cognitively demanding task) should diminish amygdalar activation. However, it is feasible that resource depletion due to heightened emotional activation might also increase the likelihood that “decisions are made less prudently” (Drevets and Raichle 1998, p. 376). To the extent that the attraction effect is a “bias” reflecting insufficient attention to all options, attraction effect–based choices may be imprudent (i.e., subject to irregularity), and therefore the impact of resource depletion on the role of emotions in decision making is an issue worthy of empirical scrutiny.

A second useful implication of Yoon, Gonzales, and Bettman’s critique is the implicit question of the nature of the baseline or control condition. That is, they liken the three-item choice set with a decoy to an experimental condition and a three-item choice set without a decoy to a control condition. This is an important methodological and philosophical concern that is particularly relevant to neuroscience research. Unlike paper-and-pencil-based research, in which a control condition is frequently identifiable, in brain scanning, such a baseline or control condition may not be easy to establish. For example, in mood research, it is often the case that a “neutral mood” is generated by asking participants to write or read a story that is expected not to generate a positive or negative mood. Then, responses on the positive and negative affect schedule (PANAS) scale in positive- or negative-mood conditions are compared with the neutral-mood condition to assess the strength of the mood manipulation relative to the baseline. In neuroscientific research, such neutral conditions are elusive. For example, a plausible control condition is the focusing on or anticipation of a fixation point (Hedgcock and Rao 2009). However, such a task may not be neutral from a cerebral activation standpoint. It might produce anxiety. The appearance of the fixation point might startle the participants when it is first encountered, or as Drevets and Raichle (1998, p. 370) note, “while performing a control task that is not very attentionally demanding,… subjects may be actively monitoring their environment,” a problem the authors associate with many PET (positron emission tomography) studies. Thus, it is important to recognize that

---

2 In addition, our purpose in Hedgcock and Rao (2009) was to identify whether negative emotion might explain the attraction effect. Over the past three decades, the standard approach to studying the phenomenon has been to compare two- with three-item choice sets (e.g., Ariely and Wallsten 1995; Huber et al. 1982).
in neuroscientific work, a baseline condition may not be the equivalent of a control condition as conceived of in standard experimental design. The living brain is never completely at rest.

Reverse Inference

Huettel and Payne point out (p. 14) that “most neuroeconomic research has taken a relatively brain-centric form. Researchers have used techniques from behavioral economics and the decision sciences to improve the understanding of brain function.... [Yet] Hedgcock and Rao take the opposite approach. They use the techniques of neuroscience to improve the understanding of decision phenomena.” That is, rather than taking known phenomenon, such as the fear of snakes, and identifying where in the brain such fear is manifested (an exercise in cranial cartography), we engage in the considerably more risky task of “reverse inference,” which assesses whether a known brain function manifests in accordance with a prior behavioral theory. We concur with Huettel and Payne, who observe that we are on relatively safe ground when reaching our core conclusion regarding amygdala activation; our secondary conclusions regarding the use of rule-based selection processes due to activation in the DLPFC and implication of “response conflict” due to activation in the ACC are more speculative in light of current knowledge about the multitude of processes that might occur in these cranial areas.

The issue of forward versus reverse inference (i.e., brain mapping versus theory testing) is a critical one in cognitive neuroscientific research that has far-reaching implications for the study of human behavior. Rather than conceive of these two modes of inquiry as “opposites,” we believe that they are complementary modes of inquiry. On the basis of our own work, we recognize that it would have been infeasible to test our prediction implicating negative emotion in the attraction effect had there not been considerable prior forward inferential research correlating activity in the amygdala with aversive stimuli. Similarly, in prior research, it would have been difficult to identify the cerebral area associated with negative emotion in the absence of some theory about what constitutes negative emotion, such as sadness (Schneider et al. 1995) or aversive olfactory stimulation (Zald and Pardo 1996).

On this topic, a caveat is in order. Although more precise theories that rely on precisely defined constructs are desirable, it does not follow that such precision will always yield the identification of small regions of the brain (through forward inferencing) that can subsequently be implicated in a test of behavioral theories. That is, “emotion” is a precise term that plays a role in several theories that vary in their degree of precision. It is also fortuitous that the cerebral area for negative emotion has been reliably identified. However, other precise constructs, such as loss aversion (Tom et al. 2007), generate activations in several brain regions, including the striatum and the ventromedial prefrontal cortex. It is possible that the simultaneous employment of both forward and reverse inference modes of inquiry will be more productive than a sequential process, according to which reverse inference can occur only after forward inference has precisely and reliably identified a brain area that is implicated in a particular process. In other words, systematic theoretically justifiable reverse inference–based investigations that implicate “rule-based processes” in the DLPFC might eventually confirm or refute the role of that part of the brain in heuristic-based decision making.

Correlation Versus Causation

Yoon, Gonzales, and Bettman observe (p. 18) that “neural activity can accompany a phenomenon without any causal connection” and caution against inferring causation from correlational data. However, Huettel and Payne note (p. 15) that “[Hedgcock and Rao (2009)] do not attempt simply to identify the ‘neural correlates’ of the attraction effect; those kinds of exploratory studies may (rightfully) be viewed as being of interest only to neuroscientists. Instead, they construct specific and well-formed hypotheses about brain function based on prior behavioral theories. In effect, neuroscience data become an operational proxy for the cognitive processes postulated in prior studies.... In summary, the value of the authors’ approach is that they use neuroscience data to test predictions derived from behavior. This is a major strength of the article.” Huettel and Payne then observe that our cautions against overinterpreting fMRI data may be overstated because increasing sophistication in design and analysis has begun to allow for less “conservatism” in the conclusions that can be drawn from imaging work. The difference in positions between the two sets of commentators is informative because it reflects philosophical differences that often typify a nascent science. While Huettel and Payne do not explicitly argue against the claim that statements about neural activity are essentially correlational, Yoon, Gonzales, and Bettman seem to recommend a more conservative rhetorical style.

Our position on this issue is squarely in the middle. Although it is technically correct that fMRI-based data are correlational, it is also philosophically defensible to attribute causation (1) if the underlying theory would support that claim and (2) until the underlying theory’s causal claim is refuted.3 In other words, whether trade-offs cause negative emotion or are accompanied by negative emotion is as much a matter of theoretical sufficiency as it is a matter of conceptual or methodological clarity. To address the issue of causation, future studies could employ other techniques, including neuroscientific methods that examine patients suffering from damage to particular cerebral areas that have been implicated in emotional activation, and behavioral studies that manipulate factors that are theoretically implicated in the generation of emotion.

QUO VADIS?

There are three broad topics that merit further discussion: (1) the development of a better understanding of the processes underlying the attraction effect, (2) the employment of cognitive neuroscientific methods in consumer research, and (3) the role of neuroscientific methods in the broader domain of research on human behavior. We discuss each in turn.

---

3We are cognizant of the philosophical problems associated with empirical refutations of theories (i.e., the discredited philosophy of science referred to as “falsificationism”). A discussion of that topic is beyond the scope of this rejoinder. We refer the interested reader to Anderson (1986).
The Attraction Effect

We are in complete agreement with Huettel and Payne when they say (p. 16), “However, we suspect that Heddock and Rao share with us the view that changes in emotional response may be but one of several contributors to the attraction effect.” We are also intrigued by Yoon, Gonzales, and Bettman’s suggestion that the underlying process may be cognition → decision and emotion rather than cognition and emotion → decision (it is also feasible that cognition and emotion → decision and more emotion). Furthermore, consistent with Huettel and Payne, we believe that context matters. Whether compromise, justification, weight shifting, negative emotion activation, or some other process occurs to generate the attraction effect likely depends on a host of contextual factors, including whether the choice is publicly observable, the hedonic properties of the product, and the emotional and/or long-term significance of the decision. In addition, according to our data, some people are likely to display the attraction effect (and associated amygdala activation) more so than others, which suggests the need for an examination of individual differences. Finally, as we alluded to previously, resource depletion seemingly matters in the manifestation of the attraction effect (Pocheptsova et al. 2009). In light of our findings regarding the role of emotion, neuroscientific examinations of the effect of resource depletion on the magnitude of the attraction effect will be informative.

Employing Neuroscientific Techniques in Consumer Behavior Research

Much has been written about the promise and problems of employing fMRI, MEG (magnetoencephalography), and other neuroscientific techniques in examining human and animal behavior (e.g., Rubinstein 2009). Some researchers view these techniques as a fad, while others are of the opinion that these techniques represent an important step forward in identifying the processes underlying human behavior (Camerer, Loewenstein, and Prelec 2005). Consistent with Huettel and Payne and Yoon, Gonzales, and Bettman, we are simultaneously excited by the promise of the existing and emerging technologies but cautious about raising expectations.

Within the consumer behavior domain, in addition to understanding the underpinnings of choice processes and the role of emotions in economic decisions, there are many derivative questions that are amenable to neuroscientific examination. In particular, as the availability of fMRI technology expands globally, knotty questions of cross-cultural differences in perception, cognition, emotion, decision making, and behavior can be addressed. In that vein, the role of language (ideographic versus phonetic scripts), family structure, religion, and other subtle dimensions of culture, as well as their role on the social psychological drivers of consumer behavior, would benefit from careful scrutiny.

Perhaps the most important prescription that emerges from our own experience and those of others who are schooled in the technique is that imaging methods and traditional approaches should not be viewed as alternative methods but rather as complementary methods. That is, paper-and-pencil approaches, observations of behavior, verbal protocols, and other traditional methods have many advantages, including their nonintrusive nature and cost, while imaging procedures have the advantage of real-time observations of process. We are in agreement with Yoon, Gonzales, and Bettman, who observe (p. 19) that “fMRI should not be used as a stand-alone methodology. Rather, researchers should seek convergent validity by linking fMRI data to other behavioral measures.” In particular, because human participants are an extremely precious resource in neuroscientific studies, it is useful if a neuroscientific study is informed by a theory that has strong behavioral support. This is precisely the reason we chose to examine the attraction effect, a phenomenon that has been examined exhaustively and has yielded a rich set of contextual factors that allow for nonobvious predictions about the underlying cerebral process.

Employing Neuroscientific Techniques in General

We are particularly cognizant of and troubled by the potential misuse of fMRI and associated methods. For example, in a New York Times op-ed, Iacoboni and colleagues (2007) present brain scan evidence regarding presidential candidates and proceed to make claims that are inconsistent with their own prior work. In a stern rebuke, 17 prominent neuroscientists castigated Iacoboni and colleagues for, among other things, “flawed reasoning to draw unfounded conclusions” (Aron et al. 2007). In marketing research, in political science, and in other areas in which practitioners are interested in “holy grail” kinds of answers to human behavior, the potential to misrepresent neuroscientific findings is substantial. Important ethical issues must be considered in the employment of these seemingly powerful techniques.

CONCLUSION

In closing, we would be remiss if we did not acknowledge the many thoughtful and insightful observations offered by Huettel and Payne and Yoon, Gonzales, and Bettman. We are grateful to them for their gracious and generous comments. Like them, we believe that neuroscientific techniques, such as fMRI and MEG, have the potential to revolutionize the conduct of research in human behavior. Yet there are prohibitive costs to conducting such research, as well as setup costs of learning the requisite technical skills. Therefore, as in any hyperspecialty, the number of consumer researchers employing these techniques is likely to remain small, and their need to collaborate with those who employ the technique on a daily basis (e.g., medical scientists, psychologists, physicists) will remain high. In light of this observation, it seems fair to assume that these techniques will not (and should not) replace existing techniques, such as surveys and self-reports on questionnaires. Yet the notable and unique insights that emerge from neuroscientific examinations of consumer behavior phenomena suggest that such research should receive serious publication consideration in marketing journals. In other words, from the standpoint of the consumer behavior discipline, the field is presented with the trade-off between a sophisticated but costly and difficult option and a relatively crude, less expensive, yet highly reliable alternative. Unlike Aristotle’s mythical person, we will doubtless eschew intellectual starvation and make the best of both techniques.
REFERENCES


