JCP JOURNAL OF CONSUMER PSYCHOLOGY

SCP SOCIETY FOR CONSUMER PSYCHOLOGY

This article is part of a Methods Dialogue:
Huber (2021): https://doi.org/10.1002/jcpy.1203
Schwarz (2021): https://doi.org/10.1002/jcpy.1206

Distinguishing Constructs from Variables in Designing Research

Bobby J. Calder
Northwestern University

C. Miguel Brendl
University of Basel

Alice M. Tybout
Northwestern University

Brian Sternthal
Northwestern University

Accepted by Lauren Block, Editor; Associate Editor, Joel Huber

Current research practices often conflate theoretical constructs and explanatory hypotheses with variables and predicted effects, to the detriment of research progress. This has led to the use of procedures such as manipulation checks, mediation analysis, and boundary conditions predicated on the idea that matching constructs to variables is necessary to validate that a theory corresponds to an effect . An alternative perspective, Inference to the Best Explanation (IBE), calls for designing research to exploit the power of distinguishing constructs from variables, hypotheses from predictions, and theory from effects. IBE calls for stating hypotheses (Hs) about construct-to-construct relationships and, separately, the predictions (Ps) about variable-to-variable effects that are explained by the hypotheses. In addition, articles should include disparate effects, a single explanation covering all studies, and a discussion of the use of the research in specific problem contexts. The application of IBE is illustrated with research investigating when judgments are based on a feeling about the ease of information retrieval versus the information content itself.

Keywords Applying research; Boundary conditions; Constructs versus variables; Manipulation checks; Mediation analysis; Theory versus effects

Designing empirical research necessarily requires thinking about variables; things that can be experimentally manipulated or measured. Predictions about the effect of one variable on another are the basis for any study. Yet, ultimately, the history of science teaches that research progress also comes through the development of constructs, concepts that are not themselves variables but form hypotheses that explain the relationships between variables. In practice, this means distinguishing constructs from variables so that construct-to-construct causal relationships explain why observed effects occur.

By contrast, the prevailing approach in much research is to develop explanations that emphasize a close fit between constructs and variables or even treat constructs as measured variables. The consequence of this is that an explanation is highly "likely," in the sense that it is an accurate description of an effect, perhaps in slightly more abstract terms, but the explanation provides limited insight into why the effect occurred. To the extent that variables (e.g., expertise) logically entail constructs (source credibility), constructs do not make observed effects more understandable. The alternative is to distinguish "best" explanations from explanations that are only "likely" because they depend on closely fitting constructs to variables at the expense of explanatory power.

Received 29 May 2020; accepted 8 October 2020
Available online 20 November 2020
The authors thank participants in the Kellogg on Designing Studies for Research Progress Conference held June 25-26, 2020, and the JCP editorial team, reviewers, and colleagues for feedback on earlier versions of this article.

Correspondence concerning this article should be addressed to Bobby J. Calder, Kellogg School of Management, Northwestern University, Evanston, IL 60208, USA. Electronic mail may be sent to calder@northwestern.edu

 1057-7408/2021/1532-7663/31(1)/188–208
DOI: 10.1002/jcpy.1204

We first consider a research example, at enough remove to avoid preconceptions, illustrating the joint roles of variables and constructs in designing research. Following this, we specifically address how the conflation of variables with constructs leads to questionable research practices in consumer psychology and related fields, elucidate the distinction between best explanations and only likely ones, and discuss implications for the application of research.

The Roles of Variables and Constructs in Research

A classic historical illustration of the interplay of variables and constructs is research on food consumption and scurvy (Carpenter, 1988; Pearl & Mackenzie, 2018). James Lind conducted one of the first controlled clinical trials in 1747. At the time, crews on sailing ships routinely suffered from scurvy. Symptoms of the disease included corkscrew shaped hair, skin discoloration, aching joints, and bleeding gums. Common remedies, such as bloodletting, were mostly ineffective, but there were indications that eating lemons or the ship's rats might be helpful. In his study, Lind treated different sailors with different remedies; some for instance drank seawater daily and others consumed lemons and oranges. Eating lemons and oranges had the effect of reducing the symptoms of scurvy whereas drinking seawater did not. Lind's conclusion was that fruits of the genus citrus prevented scurvy. Based on this, he recommended that ships carry a syrup made with boiled citrus fruit.

Applying Lind's research, the British Navy used limes to prevent scurvy. West Indian limes were used because their cost was less than Mediterranean citrus fruits. The limes were boiled, as recommended, to preserve them during long sea voyages. Unfortunately, this use of limes was ineffective in combatting scurvy. It was not until 1912 that Casimir Funk proposed the idea of vitamins (originally "vital amines" and then "vitamines") and used "C" to designate a hypothesized vitamin that prevented scurvy. It was not until 1930 that Albert Szent-Gyorgyi, who received the Nobel Prize in 1937, proposed that the key to vitamin C was ascorbic acid, so named because ascorbic means antiscurvy. From there the emerging explanation was that humans, unlike animals such as rats, do not naturally produce vitamin C, which is necessary to produce the protein collagen in body tissue. Today, scurvy can be treated effectively by consuming anything that is high in vitamin C, including broccoli, kale, and even rats. However, boiling significantly reduces vitamin C, which accounts for the ineffectiveness of limes, along with the fact that the cheaper West Indian limes used were low in vitamin C to start with.

We raise this story here because it illustrates the nature of research in a way that clearly delineates the roles of variables and constructs and points out the consequences of failing to do so. Research is not only about studying the effects of things, lemons, and oranges, but also about hypothesizing concepts, "vitamines" and the collagen formation process, as well. Progress depends on going back and forth between empirical results (variables) and theoretical explanations (constructs) to solve consequential problems. The observation that citrus fruits were not always effective in treating scurvy, and foods that did not contain citrus were effective, eventually led to the abandonment of the citrus prediction and the search for a better explanation of scurvy's symptoms. It led to the hypothesis that a vitamin C deficiency could interfere with the collagen synthesis process, thereby explaining the effect of diet on symptomatic disease.

Whereas most researchers agree that developing explanations at the construct level for observable phenomena at the variable level is important, the connection between constructs and variables presents a problem in designing research. How are constructs and variables related? Presently, the prevailing view is that researchers should seek a close fit between a construct and a variable. A construct should "correspond" to a variable, a variable should "represent" a construct, or a variable should be "typical" of a construct. The idea is that the distance between a variable and a construct should be minimized or that their similarity should be maximized. Sometimes researchers even think of constructs as measured variables. In this case, it makes sense to refer to the correlation of constructs and variables. Note however, that the term construct is used here to refer to unobservable explanatory concepts, so there can be no issue of statistical correlation. (To refer to the correlation of a construct with a variable is to treat constructs, by definition, as variables.) Whatever the language used, we label this view the Verification Approach (VA) because it assumes the close fit of constructs to variables and dictates the use of methodological procedures that are assumed to verify this fit.

The problem with VA to be examined here is that a close fit between a construct and a variable is in logical conflict with the explanatory power of a construct. In the extreme case, a variable may

simply be renamed and labeled a construct. The proffered construct seems a likely fit to the variable, yet it lacks explanatory power. In some cases, the term for the construct may be more general than the variable (e.g., citrus versus lemons/oranges), but the explanation still borders on tautology.

Even when the constructs and variables are not so obviously conflated, the general problem of logical entailment remains. The principle is that something cannot logically be the cause of itself. For example, suppose a study manipulated the number of jokes in a persuasive message, say none or three. Following VA, the construct might be taken to be "humor." The problem is that the use of jokes implies humor. The construct does not explain the variable because it is logically implied by the variable. At somewhat further remove, the hypothesized construct might be "positive mood," but there is still a clear logically implied link between the variable and the construct. To see the difference, suppose the study used a candy manipulation and the construct of positive mood to explain an effect. There is no logical entailment in this case; candy neither necessarily implies a positive mood or a negative one (though mood could explain the effect). With VA, however, the candy manipulation would seem to represent the mood construct less clearly and add to the need for a verification procedure such as a manipulation check. VA treats these procedures as especially necessary when logical entailment is less clear. As will be discussed in detail, VA procedures are predicated on linking constructs directly to variables. Thus, the procedures VA employs to confirm fit can actually reduce the quality of the ensuing explanation. Prioritizing the fit of constructs to variables contradicts the goal of explanation rather than fulfilling it.

Our contention is that a second approach, Inference to the Best Explanation (IBE), represents an alternative that has greater promise for enhancing research progress. It emphasizes that explanatory power is more than the apparent, or seemingly likely, fit of construct to variable and that the present overriding concern with demonstrating fit is counterproductive. IBE differs from VA in two important ways to be elucidated later: (1) the conceptualization of the relationship between variables and constructs, and (2) the criteria for judging the rigor of a theoretical explanation. The relationship is not a matter of fit, or correspondence, or representation, or operationalization. Moreover, the criteria for judging an explanation do not depend on conforming to procedural requirements aimed at verifying fit.

Variable and Construct Relationships Under VA and IBE

Before turning to current consumer research practices, we highlight the differences in how VA and IBE represent constructs and variables, returning to the history of efforts to treat scurvy. Lind tested the hypothesis that consumption of citrus alleviated scurvy by manipulating what sailors ate: oranges and lemons in one condition versus things such as seawater in other conditions. These conditions served as the independent variable and the construct was citrus. VA calls for a close correspondence between a construct (citrus) and a variable (lemons and oranges) as illustrated by the short vertical distance between them in Figure 1a. Critical to VA is the premise that the closeness of a construct to a variable enhances the likelihood that the independent variable represents the construct. Note that the citrus construct adds little explanatory power: citrus is all but synonymous with lemons and oranges. The citrus construct fails to explain why things such as oranges affect scurvy. Nor can it explain why noncitrus foods such as kale are as effective as oranges in mitigating the symptoms of scurvy and why boiled limes are not effective.

The close correspondence of constructs and variables in VA is even more apparent in the case of dependent variables. The construct of the dependent variable usually receives little attention as indicated by the gray color shading in Figure 1a. If considered at all, the construct is often a mere relabeling of the variable (scurvy the disease corresponds to the symptoms of scurvy). Most often, the construct of the independent variable is viewed as directly causing the dependent variable, as it is in mediation analysis (dotted line in Figure 1a). As will be discussed, this assumption presents its own explanatory issues.

Although the IBE is like VA in that both specify a relationship between theory and effects, the characterization of this relationship differs. In contrast to VA, IBE does not directly link constructs to variables. Instead, variable-to-variable relationships are the basis for developing hypotheses about construct-to-construct relationships, which in turn guide subsequent predictions about variables, as depicted in Figure 1b (Lipton, 2004). For IBE, finding that foods such as kale could prevent scurvy would not only rule out citrus as an explanatory construct, but also prompt consideration of more abstract constructs, abstract enough to explain the effects of lemons and kale. In the case of scurvy, the construct of a vitamin C deficiency emerged as

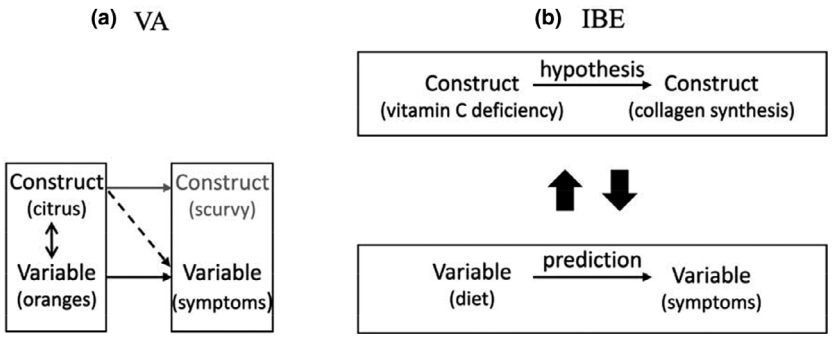

Figure 1. Differences in the VA and IBE conceptualization of constructs and variables.

an explanation of the diet-to-symptoms (i.e., variable-to-variable) relationship via the hypothesis that the deficiency interfered with the collagen synthesis process.

This was an inference to the best explanation: the uptake of vitamin C is necessary to produce the protein collagen required for healthy connective tissue because humans do not produce vitamin C naturally. This inference is not based on a direct link between a construct and a specific variable. In principle, any one construct could be part of explaining the effects of many different variables. IBE specifies that a construct-to-construct causal relationship represents a hypothesis that, in turn, can be tested by making a prediction about a variable-to-variable empirical relationship. What is important is going back and forth between hypotheses (construct level) and predictions (variable level), not between individual constructs and variables.

Thus, VA seeks to resolve the tension between constructs and variables by using constructs that closely correspond to variables, where the validity of an explanation follows from using criteria in the form of procedures that demonstrate this fit. By contrast, IBE seeks to exploit the tension between constructs and variables by constructing hypotheses, where validity follows from evaluating hypotheses according to the criteria of good explanation discussed below. Although there is certainly a logic to both approaches, we contend that, as illustrated by the scurvy example, IBE ultimately offers greater promise. It is one thing to show that citrus affects symptoms of scurvy, but it is risky to act on this without a good explanation of why. It is true that one might accept the risk of taking a drug that had "passed" a clinical trial even though no good explanation for the effect of the drug exists. But this does not mean that the absence of a good explanation is preferable to its presence.

VA seeks explanations; however, it implicitly favors resolving the logical conflict between fit and explanatory power in favor of fit. VA may not explicitly recognize it, but this tradeoff is clearly manifested in how research is designed. VA is no straw man. Imagine reviewer reactions to a methods article arguing that researchers should closely tie variables to constructs and confirm this fit with manipulation checks and other procedures. The article would be rejected out of hand as adding nothing to what researchers routinely do. IBE emphasizes explanation over fit but it does not merely contend that having a good explanation is important. It calls for going beyond explanations that seem obviously likely because the construct (citrus) fits the variable (lemons/oranges). IBE does not consider these to be explanations. But IBE does not value explanations that merely distinguish between constructs and variables. In his treatise on scurvy, James Lind went on to advance a theory employing the construct of waste that he and many others found very plausible. He hypothesized that food intake involves a digestive process in which waste is secreted through, among other things, perspiration. However, if the pores in the skin became blocked by improper diet and humidity, such blockage could cause waste to build up and produce the symptoms of scurvy. Some foods, such as oranges, were assumed to act as a "detergent" and eliminate such blockage. For the time, Lind's detergent theory seemed reasonable, but the problem was that it did not explain how citrus would be able to unblock the system. The theory led many readers of his treatise not to see that the best explanation lay in hypothesizing, as some others were beginning to, that another process was operating, one that required chemicals for balance rather than for digestion and that this theory explained why citrus was effective (De Vreese, 2008). IBE, as we will show, calls for theories that answer the why question.

Our discussion of research on scurvy serves to highlight the issue facing researchers in designing

studies. The issue is how to use both constructs and variables to the best advantage. In current research practice, many accept VA as the default for thinking about this issue. We contend that VA should be assessed more critically and that the potential contribution of IBE should receive greater recognition as an alternative.

The Criteria for Research Design Employed by VA and IBE

VA and IBE do not merely reflect different ways of thinking about the relationship between constructs and variables. They also specify distinct criteria for designing rigorous theory testing research. VA evaluates a theory test in terms of whether a series of experimental procedures are followed. By contrast, IBE focuses attention on whether a set of criteria necessary for theoretical progress is met rather than on whether specific procedures are followed. Applying these criteria, which entails making an inference about whether a postulated theory is superior to its rivals, may include procedures that are used in following VA, but adherence to these procedures is not necessary or sufficient in making an inference that a theory is the best explanation. We begin this analysis by describing and evaluating the empirical procedures prescribed by VA, followed by an analysis of the criteria for IBE.

Verification Approach Procedures

VA has long historical roots and is well established in current consumer research practice. Indeed, its criteria have been practically codified as a series of steps that are necessary and sufficient to verify the fit between variables and constructs. We suspect that in many cases researchers feel pressured by the review process to conform to these steps even though they may wish for more flexibility.

VA entails implementing a set of procedures intended to confirm that a variable is tied to a particular construct. The insistence that the rigor of a theory test depends on the inclusion of certain procedures has at this point become so ingrained that Lynch, Alba, Krishna, Morwitz, and Gürhan-Canli (2012) refer to it as the "modal scientific approach in consumer research" (p. 473). In a similar vein, Peracchio et al. (2014) were prompted to comment that there is "the sense that every paper needs to mediate, then moderate, then moderate the mediation or mediate the moderation" (p. vi).

Prototypically, VA calls for executing the following five procedures over a set of studies:

1. Demonstrate an effect predicted by theory.
2. Document the robustness of the effect using a range of settings and participant samples.
3. Assess the theoretical explanation for the effect by using manipulation checks and mediation analysis.
4. Identify moderators and boundary conditions for the effect.
5. Verify that the effect occurs in a real-world setting.

We describe each of these procedures and the assumptions on which they are based. We also discuss some important cautions to keep in mind in relying on these VA procedures. To illustrate the procedures specified by VA as they are practiced in many investigations, we selected a specific research area, one that has been prominent in the Journal of Consumer Psychology and other journals. This research originated with the question of when judgments are based on a feeling about the ease of information retrieval versus information content itself (Schwarz, 1998, 2004; Tversky & Kahneman, 1973) and was followed up by many different investigators. We regard this as an important area of research because it deals with how sometimes more information is more effective and how sometimes less information is more effective. But the main reason we selected this area is that the research procedures used are typical of those employed in many other areas. The ease of retrieval literature allows us to present concrete examples of the five VA procedures and their assumptions about how research progress is achieved. It also enables us to discuss some limitations to the inferences that can be made based on following these procedures. Importantly, this same larger body of research includes experiments that we use to illustrate the effective pursuit of IBE (Schwarz, 2012).

The ease of retrieval research area can be traced to the availability heuristic. It states that judgments of frequency, probability, and typicality are based on "the ease with which instances or associations could be brought to mind" (Tversky & Kahneman, 1973, p. 208). As Schwarz (1998) observes, whereas the availability heuristic refers to how easy or difficult it feels to bring content to mind, the effects of the variables used in some studies can be explained with another construct, the amount of information content that is readily accessible in memory. In other words, two distinct theoretical processes may

be operating, one metacognitive (i.e., ease of retrieval), the other cognitive (i.e., content).

In their seminal investigation, Schwarz et al. (1991) sought to disentangle the two explanations. They introduced what we shall refer to as "the number of exemplars effect." In general, the effect entails asking research participants to list from memory either a few or many exemplars (e.g., favorable brand attributes) regarding some belief (e.g., quality of a brand), and then to make a favorableness judgment regarding that belief (e.g., judge brand quality). The predicted effect is that requesting fewer exemplars is associated with more favorable judgments. More information does not necessarily lead to a more favorable evaluation. The effort required for information retrieval can lead to an inference about the accessibility of that information. The experience of easier retrieval when there are fewer exemplars implies greater accessibility of favorable exemplars and thus a more favorable evaluation. Other investigations have gone beyond favorableness judgments to include frequency or probability judgments and choice (Schwarz, 1998, 2004). Early work on the number of exemplars effect spawned research that illustrates the five VA procedures.

Demonstrate an Effect

Schwarz et al. (1991) asked research participants to describe either six or 12 examples of situations in which they behaved assertively. Participants in the six-exemplar condition evaluated themselves as more assertive than those in the 12-exemplar condition. Other participants were asked to describe six exemplars of situations in which they behaved unassertively. These participants evaluated themselves as less assertive than a group asked to describe 12 such situations. Thus, the effect is that instructions to list six rather than 12 exemplars lead to more exemplar behavior-consistent judgments.

Note that the effect seems important or even "surprising" due to the intuition, consistent with the results of previous studies in the memory literature, that considering more exemplars leads to more exemplar consistent judgments. Building on Tversky and Kahneman's (1973) work, Schwarz et al. demonstrate the opposite effect. They also proffer a construct, ease of retrieval, to explain the effect.

Document the Robustness of the Effect

A substantial number of subsequent investigations have examined the robustness of the number of exemplars effect (see Schwarz, 1998, 2004 for reviews; Weingarten & Hutchinson, 2018). For instance, Wänke et al. (1997) exposed participants to ads for a BMW car that stated: "There are many reasons to choose a BMW. Can you name one (10)?" (p. 172). The results indicated that prompting a consideration of fewer reasons increased the favorableness of BMW evaluations. Thus, the number of exemplars effect replicates even when participants are merely being prompted to think of reasons in response to seeing an ad, but are not required to list them. This finding, along with other studies using different manipulations, is taken as evidence that the number of exemplars effect is robust. In accord with VA, demonstrating the robustness of an effect provides assurance of its reliability. The documentation of the number of exemplars effect in a variety of different contexts verifies that the effect is not particular to a narrow set of participants, circumstances, and measures.

Although there can be value in repeated demonstrations of an effect, it is important to remember that no number of demonstrations can ensure that the effect will continue to be observed (Calder, Phillips, & Tybout, 1981, 1983). Such generalization is not logical; it rests on the fallacy of induction, a point on which we elaborate upon in our discussion of IBE. A good example of the problem with such induction is the attraction effect, a phenomenon where the addition of an inferior alternative to a choice set increases the attractiveness of the alternative to which it is most similar (Huber et al., 1982). Many studies demonstrated this effect using a variety of numerical stimuli. However, more recent investigations have shown that the effect does not always occur. For instance, the attraction effect did not occur with visual stimuli such as gambles represented in pie charts or pictorial images of the choice set (Frederick et al., 2014), although when examining judgments of perceptual properties rather than preferences others do find the effect (Spektor et al., 2018; Trueblood et al., 2013). As Huber et al., (2014) point out, their original demonstration study should not be interpreted as implying that the attraction effect would always be obtained.

Nor do repeated demonstrations of an effect increase the probability of its future occurrence. A Bayesian analysis is not the same as thinking that, say, five demonstrations versus two should necessarily increase confidence in an effect. Moreover, the procedural criterion of requiring multiple demonstrations "to be sure" an effect exists is no doubt abetted by the long-standing but now

controversial practice of assuming the statistical significance of an effect should ensure that future effects will be significant. The assumption is that verifying significance of an effect with a few more tests means that it will be reliably repeated. There are strong arguments against this view (Amrhein et al., 2019; McShane et al., 2019). It is not possible to reach a conclusion based solely on the number of observed instances given the open-ended number of possible future studies.

In general, the problem is not with multiple studies. The problem is with the interpretation VA places on multiple studies. It puts undue weight on multiple studies demonstrating an effect in a few different ways, which can only lead to a false sense of confidence that the effect is in and of itself reliable and that an important procedural criterion for further consideration is confirmed.

Assess the Explanation Using Manipulation Checks and Mediation Analysis

The assertion that a valid test of theory should include a manipulation check and mediation analysis is grounded in the VA view that there should be a close correspondence between constructs and manipulated variables, where each construct is directly related to a variable. Indicative of this, studies of the number of exemplars effect have tended to conflate the construct of ease of retrieval with variables that seek to measure indirectly or directly the feeling of ease. The emphasis on the experience of ease of retrieval in the Schwarz et al. (1991) theoretical formulation probably suggested to subsequent investigators the likely fit of the construct to variables measuring the experience of ease.

For VA, a manipulation check provides a variable-level indication of whether participants processed a manipulation in a way that fits a corresponding construct, and a mediator provides a variable-level measure of the corresponding theoretical "process" by which the independent variable influences the dependent variable. In practice, the distinction between the two is somewhat arbitrary, as revealed by the use of a measure of task difficulty in the number of exemplars research discussed below.

Researchers examining the number of exemplars effect have assessed the effectiveness of their manipulations using a variety of measures. Some investigations have included a manipulation check, sometimes referred to as an attention check, to document that participants accurately recalled the number of exemplars they were asked to list (Danziger et al., 2006; Tybout et al., 2005). Other investigations have employed a measure of task difficulty as a manipulation check to demonstrate that increasing the number of exemplars indeed increases the perceived difficulty of the task (e.g., Lammers et al., 2017; Novemsky et al., 2007; Schwarz et al., 1991; Tan & Agnew, 2016; Wänke et al., 1997; Winkielman et al., 1998). It warrants mention that using task difficulty is perhaps based on the premise that research participants are more able to report what "difficult" feels like than the feeling of "ease," thereby underscoring the status of the manipulation check as a variable and not a check on the construct. Still other studies have measured feelings of confidence in the dependent variable judgment to assess the effectiveness of a number of exemplars manipulation, with the assumption being that retrieving more exemplars should reduce feelings of confidence (Wänke et al., 1997).

Both task difficulty, which some researchers have treated as a manipulation check (see above), as well as additional process measures have been used as mediators in the number of exemplars research. For example, Tormala et al. (2007, Experiment 3) had participants recall either two or 10 situations in which they acted assertively, after which they assessed the difficulty of the task and rated the extent to which unrequested situations (i.e., unassertive actions) came to mind while attempting to recall assertive actions. They find that both measures mediate the effect of the number of situations recalled on participants' self-assessment of assertiveness: Recalling 10 situations produces a greater perception of task difficulty and more reports of recalling unrequested actions than recalling two situations. Moreover, the correlation between the measures of difficulty and unrequested actions is not significant, suggesting that they reflect independent mediating effects on judgment.

Thus, mediation analysis is used to confirm the construct ease of retrieval as the explanatory mechanism and to identify a new explanatory construct that is different from ease of retrieval. Along the same lines, the absence of any observed effect of a potential mediator is often interpreted as ruling out a rival construct. For example, observing that the number of exemplars requested, but not a measure of participants' mood, has a significant effect on judgment might be construed as evidence that mood does not play a mediating role in the effect (Ling et al., 2017).

In summary, the VA assumption that manipulation checks and mediation analysis enhance validity follows from the view that there should be a close,

if not one-to-one correspondence between constructs and variables, where each construct is directly related to a variable. Thus, at a minimum, research participants should be able to recall the number of exemplars they were asked to list (manipulation check) and report greater feelings of difficulty when asked to list more exemplars (mediator). A mediation analysis should show that these feelings, and not some alternative factor such as mood, mediate the effect of the number of exemplars listed on judgments.

The VA perspective on manipulation checks and mediation raises several concerns beyond the statistical issues detailed by Fiedler et al. (2011). A manipulation check and measures of a mediator should not be regarded as privileged variables that "capture" the construct of ease of retrieval directly. Both the number of exemplars requested and ratings of feelings of difficulty are observable variables, not unobservable constructs. The use of a rating of "difficulty," rather than a rating of "ease," underscores that this is a variable, though even the latter would still be a measured variable. At best, the correlation of a manipulation check or mediator with a dependent variable is an additional effect to be explained rather than being the explanation of why the independent variable affected the dependent variable. Finding that the feelings of difficulty measure is a significant mediator of the relationship between the number of exemplars listed and judgment does not, per se, imply that the underlying mechanism is the construct "ease of retrieval." The effect of the feelings of difficulty variable on the judgment measure could also be interpreted in terms of alternative constructs, such as resource depletion or self-esteem. Contrary to VA, the feelings of difficulty variable does not have a one-to-one relationship with any one construct.

A further concern with VA is that mediation analysis assumes that the independent variable affects the mediator variable and the latter affects the dependent variable. However, a parallel statement in terms of constructs makes no conceptual sense because both the independent variable and the mediator would refer to the same construct—ease of retrieval. Furthermore, there is the possibility that, rather than representing the process by which the constructs are related, the mediator is simply an alternative dependent variable. To illustrate, consider a study by Tan and Agnew (2016, study 3) in which individuals listed either two or 10 plans for their romantic relationship and then indicated the difficulty of this task (manipulation check), their doubts about commitment (mediator), and relationship commitment (dependent variable). Although the data analysis supported role commitment doubt as a mediator, an alternative interpretation is that commitment doubt and relationship commitment are two measures of the same thing that are phrased slightly differently. Similarly, the finding that a one-item measure of ease is just as predictive of a dependent measure as a multi-item measure of ease (Graf et al., 2018) may be attributable to the fact that these measures are not mediators but instead measures of the dependent variable that correlate highly with other measures of the effect.

It is also routinely the case that the mediator is assumed not only to be a construct, but the dependent variable is treated as a construct as well. Mediation is thus thought to show the connection of the causal construct to the construct of the variable to be explained. This mistakes not only the mediator for a construct, but also the dependent variable (see Figure 1a). For IBE, both are variables and it is the construct-to-construct hypothesis that must explain the relationship between them. The relationship between the mediator and dependent variable is something to be explained rather than an explanation.

In addition to the difficulty of interpreting findings related to manipulation checks and mediators according to VA, the requirement of these measures raises the issue of when during the experiment they should be administered. If these measures are taken prior to measures of the dependent variable, they may influence the dependent variable in a way that would not have occurred in the absence of this measurement (Hauser et al., 2018). However, if these measures are taken after measures of the dependent variable, even by the logic of VA (assuming that the variable is closely tied to the construct) it is reasonable to question whether they indeed played a causal role given their order of measurement. This concern has led some research to vary the order of administering these measures before or after the dependent variable. However, finding that the variation in order has an effect itself requires explanation (Kühnen, 2010). For example, individuals who had the proclivity to base their judgments on content were more favorable toward an educational proposal when they listed more reasons for adopting it, but only if they reported their feelings about the difficulty in listing reasons after reporting their judgment. When they reported their feelings of difficulty prior to judgment, the opposite effect was observed (Danziger et al., 2006). A possible explanation is that

administering the measure of task difficulty prior to judgment drew attention to feelings, which offset these individuals' natural tendency to rely on content.

VA also seeks support for an explanation by including measures that are designed to rule out rival explanations. However, finding that the number of exemplars has no effect on one measure of mood, such as the PANAS scale (Watson et al., 1988), does not rule out the possibility that mood influences judgment. The number of exemplars variable might affect some alternative measure of mood, such as the BMIS scale (Mayer & Gaschke, 1988) or judgments of mood based on facial expressions. This might occur if the alternative measures are more reliable than the one employed. However, even if the measure employed is highly reliable, there is a fundamental problem with the VA practice of using null effects to eliminate rival explanations. It rests on the assumption that the construct of mood can be represented by a specific measured variable, and therefore, a null effect on that variable is sufficient to rule out mood as a construct in an alternative explanation. This conflates mood as a construct with mood as a variable. Contrary to VA, the construct should not be tied to the variable. Hence, the absence of an effect on a variable only means there is nothing to be explained, not that the construct is inoperable and could not explain the obtained effects. Relatedly, showing that the number of exemplars affects a measure of mood, but mood does not mediate the effect of the number of exemplars on judgment, runs into the problem of treating mediating variables as constructs, as discussed earlier.

In summary, it is common practice to treat manipulation checks and mediators, which are variables, as if they were direct measures of a construct. However, by definition, there is no such thing as a direct measure of a construct because constructs are abstract entities. Manipulation checks and mediators should be treated as variables that may be related to independent and dependent variables in such a way as to need explanation, rather than as providing an explanation.

Identify Moderators and Boundary Conditions for the Effect

Using moderators is a strategy for gaining insight about a construct, changing the research question from whether an effect is observed to when it occurs. However, VA often advocates a special kind of moderator, boundary conditions; that is, circumstances in which the effect fails to obtain. A boundary condition is thought to verify that the variable represents the construct because the construct is inoperative when the effect is absent.

If an independent variable produces an effect on a dependent variable and it is then shown that in the absence or negation of the independent variable the effect does not occur, for VA this is often taken as evidence that the dependent variable depends on the independent variable and its construct (leaving aside the problematic interpretation of the null hypothesis). However, a boundary condition at the variable level does not necessarily say anything about why the effect occurred in the first place or why it did not occur in the boundary condition. A moderator variable that merely "kills" an effect without being theoretically informative about the original effect makes no contribution. In the case of the attraction effect, showing that the effect occurs with numerical stimuli but not with visual stimuli establishes a boundary condition for the effect (Frederick et al., 2014). But as the authors correctly imply, this demonstration per se does not identify the construct that explains finding the attraction effect with numerical stimuli or the construct that explains why the effect does not occur with visual stimuli.

As with other procedures, the problem is not with moderation but with VA. As an example of the potentially informative use of moderators, consider Aarts and Dijksterhuis (1999). They instructed participants either to be highly accurate or provided no accuracy instruction when listing either three or eight destinations to which they rode their bike. In the absence of the accuracy instruction, participants' subsequent reports of bike usage showed the number of exemplars effect: greater bike usage was reported when three rather than eight destinations were listed. However, in the high accuracy condition the opposite outcome occurred—listing more reasons resulted in reports of greater bike usage. Note that this moderation does not involve a null effect.

The introduction of a moderator is useful when it produces an interaction effect that narrows the set of potential explanatory constructs. One interpretation of the moderation in the Aarts and Dijksterhuis study is that the accuracy instruction directed participants' attention away from their ease of retrieval and toward the content of the exemplars listed. However, even when the moderator does not merely eliminate an effect (i.e., create a null effect), multiple interpretations may be

possible. Perhaps the accuracy instructions created experimental demand, prompting participants to believe that they should expend significant effort. This inference would increase the estimated number of bike destinations as the requested number of exemplars increased.

Regarding moderators and interaction effects, it is also worth noting that some statistical interactions are conceptual main effects. This is relevant because a statistical interaction does not necessarily mean that moderation has been shown. Consider the disordinal interaction reported by Schwarz et al. (1991) described earlier where listing fewer exemplars of assertive behaviors increased the self-perception of assertiveness and fewer exemplars of nonassertiveness decreased the self-perception of assertiveness. Although the statistical interaction between the number of exemplars and valence of the inference is significant, the result is a conceptual main effect: Generating fewer exemplars of one type of behavior results in a greater self-perception of this behavior. Thus, it can be misleading to rely automatically on whether an interaction effect is obtained. We shall return to the Schwarz et al. (1991) investigation in our discussion of IBE to describe how the researchers addressed this issue.

Moderation is a valuable means of testing theory. Our point is that moderation is of great value only when theory hinges on the explanation of the moderation. However, a boundary condition that eliminates or modifies an effect is by itself not sufficient to support a particular explanation. We elaborate on how moderation can enhance explanation in our discussion of IBE.

Verify That the Effect Occurs in a Real-World Setting

A final step in VA is to demonstrate the robustness of effects in a setting where consumers make consequential judgments. VA assumes that the variables in a field study are closer to variables in real-world settings and, at least implicitly, that this consequently ties constructs to these real-world variables. For example, Janssen et al. (2011) investigated the number of exemplars effect by conducting a field study in Germany. Participants were applicants for positions in a large aviation company who were solicited during the application process under the guise of obtaining feedback about the application experience. They were randomly assigned to list either one or four aspects of the application procedures during their contact with the company they considered fair. They then reported their uncertainty about the process (e.g., "Missing information made me feel uncertain during the online process"), and their experience in applying for positions online. Finally, participants completed a manipulation check (i.e., ease or difficulty of listing the requested fair aspects of the application procedure) and judgments of the procedural justice of the application process.

The number of exemplars effect was observed: job applicants' evaluations of the company's online application process were more favorable in the one versus the four fair aspects condition—but only when they felt relatively certain about the application process and had prior experience applying for positions online. This is a measured moderation effect. In the absence of these conditions, participants appeared to rely on the content of the fair aspects they listed: Evaluations were more favorable in the four fair aspect condition than in the one fair aspect condition. Thus, both the robustness of the number of exemplars effect and its boundary conditions were demonstrated in a field setting, thereby, according to VA, verifying that the ease of retrieval construct corresponds to a real-world variable.

It is often further assumed that a field study, such as the Janssen et al. (2011) investigation, is informative because the independent variable produces the effect of interest despite the noise inherent in the field setting. However, a field setting per se does not make the effect more "real" or "true," as is often assumed. Indeed, Loken and Gelman (2017) demonstrate that statistical significance is less informative with noisy measurements, which may occur in field settings if the sample size is relatively small. Thus, field studies run the risk of inflating significance for small effects.

Contrary to VA, the failure to observe an established effect when research is conducted in the field should not be regarded as definitive in disconfirming the explanation for an effect found in controlled settings because, by definition, other variables may be at work. And even when the effect is found in the field, this outcome does not necessarily contribute to the quality of the theoretical explanation. In fact, it is more that the explanation can contribute context to the effect. Ease of retrieval has implications for the kind of contexts in which the number of exemplars effect might occur. Beyond this, "coming through in the field" does not imply that an effect is strong enough to obviate the need to explain the effect. The scurvy and attraction effect examples are cases in point.

VA procedures, their underlying assumptions, and concerns related to strict adherence to these procedures are summarized in Table 1.

Table 1
VA procedures, rationale, and cautions

Procedures	Rationale	Cautions
1) Demonstrate an effect predicted by theory	Theories and effects go hand in hand. Theories lead to hypotheses about variable relationships and observed effects test the adequacy of the explanation	Explanations that are merely a slightly abstracted description of an effect provide limited insight
2) Document the robustness of the effect across contexts and participants	Only robust effects warrant explanation	Repeated observation of an effect does not ensure that it will be observed in future settings nor does it increase the probability of its future occurrence
3) Assess the explanation of the effect using manipulation checks and mediation analysis	Confidence in the explanation is increased by verifying that the manipulation had the intended effect and showing that the hypothesized process and not some alternate process occurred	The distinction between a manipulation check of a variable and a mediator is often arbitrary; both are simply additional measures and have no special status in capturing constructs. There are no direct measures of constructs The timing of the administration of these measures may confound their interpretation Null effects on these measures are inadequate to rule out rival explanations
4) Identify boundary conditions for the effect by introducing a moderator	Potential explanations for an effect should account for circumstances when the effect will not occur, as well as when it will occur (i.e., the scope of the effect.)	Moderators that merely eliminate an effect are open to multiple interpretations. Boundary conditions apply to variables, not constructs Statistical interactions may be conceptual main effects and thereby offer little insight about the explanation for the effect
5) Verify that the effect occurs in a real-world setting	Theories should be relevant to consequential events that occur in natural settings	Real-world settings do not warrant special status in the case of theory testing. Seeming relevance does not make variables explanatory constructs If the sample size is small, the significance of an effect in a noisy real-world setting may be less rather than more informative than in a controlled setting Null effects in real-world settings should not be interpreted as damaging to an explanation per se due to the increased variation

IBE Criteria and Implications for Procedures

To this point, we have examined progress in theory-oriented research in terms of the research procedures commonly accepted as necessary for ensuring rigorous tests as prescribed by VA. We have raised issues that call into doubt an uncritical acceptance of these procedures and the underlying premise that accepts logical entailment in an effort to yoke constructs and variables. We now turn to an analysis of how research progress could be enhanced by adopting the approach implied by the IBE perspective (Lipton, 2004).

IBE is consistent with a falsificationist view of research. Progress comes from advancing theories with highly informative content that expose them to refutation by data (Popper, 1954). Such theories are preferable to theories that may seem probable but would be difficult or impossible to falsify. A theory in which variables logically entail constructs is a case in point. IBE goes beyond this in emphasizing that not only should theory be judged according to whether its content is falsifiable, but also by whether its content is highly explanatory. In the end, we never know whether a theory is true or false; future falsification is always possible. IBE evaluates how well a theory increases understanding of observed effects. Truth is more than understanding, but understanding, and not truth, is the only goal possible. By contrast, the VA approach to constructs and variables is rooted not in falsificationism but in empirical justificationism (Calder et al., 2019). This is the view that truth is possible and can be justified by empirical confirmation.

Thus, IBE focuses on the criteria for explanatory value. Specifically, IBE prioritizes explanatoriness

(Schupbach, 2017). A theory with high explanatoriness is one with constructs that explain why an effect occurs in terms of the causal connections among its constructs. If a theory is the best possible explanation for the effect, the research is valid in that the theory renders the effect understandable and, in turn, the existence of the effect supports the theory. As we noted earlier in describing the history of treating scurvy, effects are explained by hypotheses based on theory that generates predictions about effects. The key question is what are the criteria for evaluating explanatoriness?

A Fundamental Criterion for IBE Explanatoriness: Explanatory Power

Several criteria have been proposed to assess explanatoriness. Perhaps the most fundamental is explanatory power. The explanatory power of a theory is derived from the extent to which it goes beyond observed effects to increase understanding of why the effect occurred (Lipton, 2004). This understanding is at the level of theory, not at the level of effects. Variables do not explain; they are simply observables. Construct-to-construct theory explains the observed relationships between variables.

Defining explanatory power in this way distinguishes IBE from VA. Whereas VA advocates for a close correspondence between variables and constructs, IBE adopts the view that requiring constructs to correspond closely to variables does not lead to the best explanation. The theory may seem likely because of such correspondence, but explanatory power is in fact sacrificed. In many cases, the constructs become little more than a general label for the variable. In the historical example described at the outset, the effect level was the observation that oranges and lemons prevented scurvy. At the theory level, the initial hypothesis was that this was due to citrus. As oranges and lemons were thought to correspond to citrus, this seemed highly likely. Eventually the role of vitamin C in collagen production provided the best explanation of the effects, though few at the time would have thought of it as "likely." Lind's study was viewed through a VA lens, with citrus being the most likely explanation, and even the detergent food theory being seen as more plausibly likely than a deficiency theory. Above all else, IBE rejects being obviously likely or plausibly likely as the criterion for judging how good a theory is. Put another way, seeming "likely" should not be equated with "best." The history of science records many examples of poor explanations being accepted because they seemed likely and others being initially rejected because they seemed unlikely at the time. These include the acceptance of the theories that waves are propagated in the air through an invisible ether, that the universe is static rather than expanding, and that nuclear reactions can occur at low temperatures (cold fusion). Rejected but later accepted theories include Avogadro's law that equal volume gases include an equal number of molecules, continents drift, and handwashing prevents disease. Seeming likely is not a good criterion for judging theory.

According to IBE, a construct only has high explanatory power to the extent that it offers an understanding that is not inherent in the variable itself. As Lipton (2004) puts it, the best explanation is not the most likely in the trivial sense of having constructs that best describe variables. The best explanation occurs when the constructs go beyond the variables to add the most understanding. One way of stating this is to say that the best explanation is the one that takes the most "surprise" out of the effect (Schupbach, 2017). Ease of retrieval is the best explanation to the extent that it makes the number of exemplars effect unsurprising. Progress lies in, among other things, extending the theory to make the effect even more unsurprising. Lipton refers to the best explanation as the "loveliest" to distinguish an explanation that is trivially "likely" based on construct-to-variable correspondence, from one that renders the relationship among variables understandable. Research on the number of exemplars effect illustrates such progress. The finding that fewer favorable exemplars induce more favorable judgments than many favorable exemplars is surprising. The explanation in terms of the construct ease of retrieval renders the effect less surprising. Although there is realization that surprise is relevant (e. g., Huber, 2008), there is pressure in the journals to present theoretical explanations as what was always anticipated and effects as what was expected.

Explanatory power clarifies why an effect occurs. And the fact that it occurs supports the theory. It is not that the variable, the number of exemplars, has been verified to represent the ease of retrieval construct. Explanatory power arises at the level of theory—the explanation in terms of ease of retrieval renders the effect more understandable. However, the "Hs" commonly stated in articles often obscure the issue of explanatory power. Sometimes Hs refer to hypotheses, more often to predictions, and sometimes to both. We believe that this confusion not only reflects the VA perspective but also reinforces

it. We propose, as a practical IBE criterion of explanatory power, that empirical articles include an explicit statement clearly delineating hypotheses from predictions. The format would follow the common practice of stating "Hs," but take the form of "PHs" that clearly separate the predicted effect level (P) from the theory level of explanation (H) as illustrated in the following PH statement.

> PH: Listing a lower (higher) number of favorable exemplars for buying an automobile brand increases (decreases) ratings of favorability toward the brand (P), because ease of processing information about a target product causes an inference that the target's properties are determined by the degree of retrieval ease such that greater ease causes the attribution of positive properties and a lack of ease causes the attribution of negative properties (H).

If, unlike the illustration, the PH statement displays logical entailment of variables and constructs, then explanatory power should be questioned.

Three Additional Criteria for IBE Explanatoriness

Beyond explanatory power, a theory's scope or unification is a second criterion specified by IBE as a basis for inferring that a theoretical account is superior to its rivals (Friedman, 1974; Kitcher, 1989). In our view, if a theory provides an understanding of a broader range of effects than its rivals, this superiority warrants consideration in making a best explanation inference. Thus, scope is intimately linked to abstraction (Van Lange, 2013): It is abstraction that gives scope to a theory. Such abstraction is not ambiguity. Scope derives from unambiguously explaining different effects. As a practical IBE criterion of scope, multiple studies should include studies with disparate effects that allow a theory to display scope.

A third criterion for IBE is parsimony. A theory that has greater parsimony is preferred to its rivals. As Trope (2004) points out, parsimony is closely tied to explanatory power. A theory gains explanatory power over competing theories if it is simpler but similar in scope. This implies that parsimony is not the same as a preference for simple theories. Indeed, a complex theory that provides a better causal mechanism is a better theory. Parsimony occurs when an explanation is irreducibly complex. In other words, an explanation should be as simple as possible, but no simpler. This statement of parsimony is often incorrectly attributed to Einstein, but what he said was even more informative (Robinson, 2018, p. 30): "It can scarcely be denied that the supreme goal of all theory is to make the irreducible basic elements as simple and as few as possible without having to surrender the adequate representation of a single datum of experience." As a practical IBE criterion of parsimony, multiple studies should be explained by one theory, not explained separately. This includes previously published studies to the degree that the theory can make predictions for these.

The final criterion is context. Explanatoriness, scope, and parsimony must be considered relative to context. The contention is that theories should be evaluated contextually or pragmatically in terms of whether they are the best explanations for a set of intended users (van Fraassen, 1980; Pham, 2013). Our view, however, is that this should not be a matter of agreement among one group of users, but part of the explanatory intent in selecting causes of explanatory interest. Moreover, the causes selected must still be evaluated according to explanatoriness, scope, and parsimony. A good explanation is one that can become part of a set of practices. We return to this point in the discussion of theory application. As a practical IBE criterion of context, rather than broad managerial implications, articles should discuss how the research can be used in specific problem situations.

To sum up, in the historical scurvy example, the vitamin C explanation had high explanatoriness. It had the explanatory power to say why oranges and lemons, as well as broccoli and kale would be effective. And the role of vitamin C in collagen production explains why the symptoms of scurvy include corkscrew shaped hair, aching joints, and bleeding gums; these are related to tissue formation. Thus, the Vitamin C hypothesis has more scope than the citrus hypothesis. It explains all the observations that citrus can explain, and even more (scope), with irreducible complexity (parsimony). And the explanation was well-suited to the context, or domain, of dietary interventions.

The Role of Explanatory Convergence

IBE shifts research emphasis from trying to trace a theoretical process by directly linking variables to constructs to uncovering the best explanation for a pattern of effects created by disparate variables. This involves examining different independent and dependent variables that produce disparate effects, where the effects (e.g., of lemons and kale) are all

explained by the same constructs (e.g., vitamin C) as a means of achieving the best explanation. We refer to this approach as explanatory convergence.

Explanatory Convergence Across Disparate Effects

Although we have used the number of exemplars effect to illustrate the limitations imposed by VA in some studies, in a broader context, research on this effect reflects significant progress consistent with IBE. This entails theoretically motivated convergence across disparate effects.

To illustrate explanatory convergence, consider the theorizing that people regard their feelings like any other kind of information when making a judgment, using feelings to the extent they regard them as useful (for a review of this theorizing see Schwarz, 2012). In the context of the number of exemplars effect, the feeling prompted by ease of retrieval in listing exemplars is perceived as relevant and attributed to the target of judgment. This contention implies that increasing the number of favorable exemplars reduces the favorableness of judgments or estimates of frequency. However, an awareness that it is not appropriate to attribute feelings to the target of judgment renders feelings as irrelevant, and other information such as the content of the exemplars will be sought. This theorizing implies that increasing the number of favorable exemplars will increase the favorableness of judgment or the estimate of frequency.

Schwarz et al. (1991) tested this theorizing by asking respondents to describe either six or 12 exemplars regarding their self-assertiveness, a task that was performed in the context of background music (experiment 3). In the control condition, this was the only information respondents were given, whereas in the experimental condition they were also informed that the difficulty in retrieving exemplars was caused by the background music. As predicted, increasing the number of exemplars reduced respondents' judgments of their assertiveness in the control condition, whereas increasing the number of exemplars increased judgments of assertiveness in the experimental condition. Thus, the explanatory convergence provided by moderation provides a test of theory about when judgments are likely to be based on feelings of ease and when they are based on content.

Other investigations using explanatory convergence procedures offer a means of abstracting the nature of the situational requirements that determine whether judgment is based on feelings pertaining to ease of retrieval or content itself. Consider the proposal that conditions favoring judgmental heuristics in general (e.g., low relevance, limited cognitive resources, happy mood) would also favor ease of retrieval, whereas conditions favoring systematic processing would favor information content (Schwarz, 2012). Support for this contention is reported by Rothman and Schwarz (1998) who found that individuals with no family history of heart disease, for whom this issue was likely to be of low relevance, perceived greater vulnerability when they recalled three rather than eight risky behaviors, implying that their judgments were based on ease of retrieval. By contrast, those with a history of heart disease, for whom this issue was of high relevance, perceived themselves to be more vulnerable after recalling eight rather than three behaviors that might increase their risk of heart disease, suggesting they based their judgments on the information content itself.

To this point our illustrations of explanatory convergence have relied on variation in the number of exemplars listed. Explanatory convergence can, and ideally should, also entail the use of maximally different independent variables. Stepper and Strack (1993, Experiment 2) applied the insight that the feelings of mental difficulty attributed to listing many exemplars also accompany forehead contractions (i.e., frowning). They predicted that frowning would increase the effort experienced in performing a mental task, such as the difficulty of retrieving an episode from memory. Although recalling six self-assertive episodes may feel easy, whereas doing so while frowning should render the recall experience more difficult. Consistent with this prediction, participants recalling instances of high self-assurance evaluated themselves as less self-assured when frowning than when smiling, and the opposite occurred when they recalled instances of low self-assurance. Thus, frowning while listing six episodes had the same effect as listing twelve episodes in the Schwarz et al. experiment (1991), where no facial instruction was provided. Both effects can be explained by the hypothesis that people typically attribute their feelings to the judgmental target (e.g., self-assertiveness), which in this case involved a misattribution of bodily feelings to feelings of ease of retrieval (Schwarz, 2012).

The convergent effects of frowning and more exemplars result in high explanatoriness around feelings as information as a construct, which is more abstract than ease of retrieval and supported by the two different kinds of feelings having the same effect.

Feelings as information has more scope, and parsimony through convergence. The context in which people use feelings as information (and ease of retrieval) becomes clearer as well, when they attribute them to the target, as also shown by the above-mentioned music manipulation.

Explanatory convergence across experiments conducted by different investigators that employ different moderators can also enhance explanatoriness. Recall the finding described earlier that the reversal of the number of exemplars effects occurs when there was a need for accuracy (Aarts & Dijksterhuis, 1999). We suggested this effect might be dismissed because it is subject to a rival explanation of experimental demand. However, this concern might be mitigated by Janssen et al.'s (2011) finding that feelings of uncertainty and lacking prior experience with the job application process result in the same data pattern as accuracy instructions. Assuming that accuracy instructions, feelings of uncertainty, and a lack of experience all relate to an elevated need for accuracy, the entire data pattern not only reduces the plausibility that experimental demand was responsible for Aarts and Dijksterhuis' results, but also elaborates on the nature of situational requirements influenced by feelings.

The inference that emerged from these findings is that when an elevated need for accuracy prompts participants to view outcomes as consequential and thus motivates them to exert effort, judgments are based on the information content itself (Schwarz, 2012). In the absence of such situations, judgments reflect the goal of limiting effort. This theorizing not only provides a hypothesis that can be tested in disparate contexts, but also links investigations of the number of exemplars to research that pertains to an accuracy-effort framework (Bettman et al., 1998; Payne, 1982). Explanatory convergence over different effects from different domains thus increases the explanatoriness, scope, and parsimony of theory.

These illustrations of explanatory convergence clarify a key difference between VA and IBE. Because VA emphasizes construct-to-variable fit, it often reduces theory to a "likely" restatement of a specific effect and resists multiple-experiment studies with dissimilar variables. VA avoids such studies because dissimilar variables cannot all closely fit a single construct. By contrast, IBE values developing theory to the point that it can explain diverse effects, ideally across seemingly different domains. The guiding principle is that an inference to the best explanation should make a broad range of observations unsurprising.

Explanatory Convergence and Mediation

IBE views mediation analysis (and manipulation checks) as an alternative or an additional means of achieving explanatory convergence but does not accord them the weight that VA does. Mediation analysis can be particularly useful when there are two outcomes, such as when listing more exemplars decreases the favorableness of evaluations under one condition and increases the favorableness of evaluations in another condition, where the mediators differ for each outcome. For instance, participants asked to list their own arguments in favor of using public transportation judged this means of transportation less favorably when listing more arguments. By contrast, those reading arguments in favor of using public transportation generated by others were more favorable when reading more arguments (Wänke et al., 1996). One explanation for these two findings is that listing arguments focuses attention on the difficulty of the task, so that listing more arguments is more difficult and therefore less persuasive. By contrast, reading others' arguments focuses attention on the content of the arguments, so that reading more arguments is more persuasive. Support for this explanation would be enhanced if the experiment had included two other measures: (1) a measure of the difficulty experienced in listing arguments that mediated the effect of listing arguments but not of reading arguments, and (2) a measure of the time taken to make a judgment that mediated the effect of reading but not listing arguments.

Documenting the role of distinct mediators for different effects at the variable level enhances the explanatoriness of the theory test at the construct level in two ways. It provides evidence that evaluations based on feelings of ease of retrieval and those based on content listed by others involve unique processes. In addition, observing distinct mediators for different outcomes reduces the possibility that the mediator and dependent variable represent the same construct. This follows from the fact that the feeling of difficulty is not always related to the dependent variable of interest. In the hypothetical example above, it was the time to make a decision variable and not the difficulty experienced in listing variable that mediated the outcome when evaluations were content based. At the same time, it should be recognized that mediation analysis is not necessary. An inference to the best explanation can be made by using independent variables to achieve explanatory convergence, or by using mediation, or by both.

It is also the case that convergence over dependent variables can aid inference to the best explanation. Ruder and Bless (2003, experiment 4), prompted participants to adopt a positive or negative mood by describing either a happy or sad life event and then asked participants to generate either two or five arguments advocating against introducing a toll on the German freeway system. Participants in a positive mood exhibited more favorable evaluations of the position advocated in their arguments when they generated two rather than five arguments, whereas the reverse occurred for those in a negative mood. One interpretation of this finding is that those in a positive mood based their evaluations on feelings, and those in a negative mood on content. An alternative interpretation is that those in both moods based their judgments on feelings: Generating more arguments felt more difficult for those in a positive mood, whereas generating more arguments implied that there were many reasons for their advocacy for those in the negative mood. Ruder and Bless addressed this alternative explanation by including a latency measure. They found that the time taken to respond to the evaluative questions was relatively short and invariant for those in a positive mood and longer for those in a negative mood, particularly when five rather than two arguments had been generated. The inclusion of the latency measure provides convergent evidence supporting the inference that evaluations of those in a positive mood were based on feelings, whereas those in a negative mood were based on content.

We have emphasized that IBE is an inferential process. Figure 2 summarizes the goal of employing studies of different independent and dependent variables across divergent effects to achieve explanatory convergence across these effects to meet the four criteria of explanatoriness. A frequently asked question is whether explanations that converge across disparate effects as illustrated in Figure 2 might in fact be inferior to lower level explanations at the level of specific effects. In this regard, note that IBE calls for theories that satisfy the criteria for explanatoriness. Any convergent theory might, of course, fail to meet these criteria, in which case more local theories of specific effects might be better. There is no reason, however, to think that research offering more local theories is necessarily superior. To the extent these local theories conform to VA, their contribution is limited ipso facto.

IBE and Research Applications

As previously discussed, VA privileges field studies for connecting constructs to real- world variables to ensure that constructs are applicable. MacInnis et al. (2020) refer to this as "phenomenon-construct mapping," defined this way: "researchers start with observations of real-world marketing-relevant phenomena and then identify constructs and relationships that can explain them" (p. 3). This is consistent with the view of some researchers that theory testing studies have become too insular. Thus, Pham (2013) distinguishes between "theories of studies" that pass for "studies of theories." Theories of studies focus on "very narrow phenomena that are created by the studies themselves" (p. 10). The adherence to VA procedural prescriptions probably contributes to why many find such studies formulaic. Others call for more studies emphasizing effects over theory (Alba, 2012; Lynch et al., 2012).

IBE, however, leads to a different perspective: Effects are important, but strong theory is necessary to know why an effect is important. What if an initial investigation had simply reported the number of exemplars effect with little or no theorizing? Why would the effect have been seen as important? The reason would have to be that it was surprising. But why would it be surprising? One must have

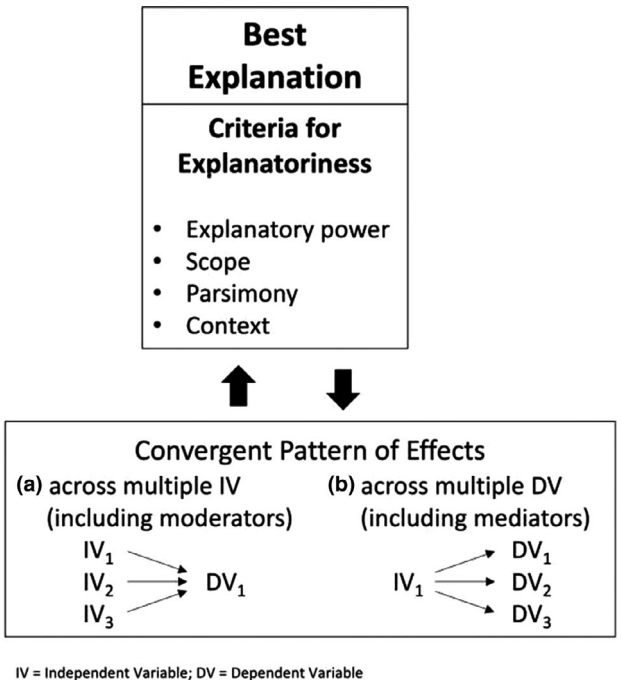

Figure 2. IBE criteria and process.

had a low prior that the effect would occur. To have a high prior would imply that the effect is understandable, that is, that there is a theoretical rationale for it. But having a low prior and finding the effect can only mean one of two things: Either the effect is an artifact, or there is an unknown explanation for it.

Thus, the appropriate stance according to IBE is to view an effect as an invitation to develop not more theories but more explanatory theory. IBE is meant to apply to an entire research program or to an entire field, not to a single research article. The inference evolves across empirical findings. So-called effect articles do not stand in contradiction with IBE, and IBE should not be used to reject them. Their publication should serve as a call for explanation. We agree with Lynch et al. (2012) that "Such 'atheoretical' findings-first papers can be of exceptional heuristic value in alerting us to topics of true importance" (p. 480). With areas such as the attraction effect (Fredrick et al., 2014) and the mere-measurement effect (Morwitz & Fitzsimmons, 2004), where according to Lynch et al. explanations remain "elusive," this value cannot be taken for granted. Until an effect is truly rendered unsurprising by theory, it may or may not be important. Moreover, this is the case whether the effect is large or small.

IBE and Effect Sizes

It may seem intuitive that larger effects are inherently more important, both theoretically and practically, than smaller ones. To illustrate our contention that an effect's size should not be equated with its importance, consider the following example. We adapt the example from Funder and Ozer (2019) and Darlington (1990). Suppose two studies are conducted. Study 1 examines the effect of the size of a product (small or large) on how much a consumer is willing to pay for it. Study 2 looks at the effect of asking for a few versus many reasons for buying a product on willingness to pay. For purposes of the illustration, there are four participants in each study, and the dependent variable values are the same across the two studies. The results of both studies are shown in Table 2. The mean for the effect in Study 1 is 12.5 for the large size product versus 2.5 for the small size. The correlation between the IV and DV is .89. The effect in Study 2 is 10 for few reasons and 5 for many reasons, with a correlation of .45. The effect in Study 1 is thus much greater. The variance accounted for, though a dubious statistic (Funder & Ozer, 2019), indicates that the former effect is four times as large. But does this make it more important?

Although the effect of the size of the product on willingness to pay is large, effect size does not equate to importance. According to IBE, there must be some surprise in the effect that calls out for explanation. Importance will be a function of the understanding created by the explanation. There is no surprise with Study 1 and not much need for explanation: Consumers will pay more for more product. Study 2, on the other hand, is surprising. Why would thinking of fewer reasons make consumers willing to pay more? Though the effect is much smaller, it is more important from the standpoint of IBE. The intuition that focusing on finding large effects is worthwhile in and of itself is counterproductive. Note that if both the size and the number of reasons variables were part of the same 2 x 2 design the research becomes more interesting. This design reveals that consumers are willing to pay more for a larger size especially if they do not have to exert much effort to think about their reasons. It is the smaller effect that creates more interest. Yet theory-oriented research is often criticized for dwelling on small effects rather than seeking large ones that are "obviously" important.

Real-World Interventions and the Criterion of Context

Theory is crucial to real-world interventions. The scurvy example is a case in point. How would you protect an expedition if you did not have citrus fruit? Because the mechanism by which citrus fruit prevented scurvy was not understood, the British Navy tragically used limes (low in vitamin C) for decades, resulting in many deaths (Carpenter, 1988; Pearl & Mackenzie, 2018). The design of an intervention requires an understanding of how it brings about a desired effect. The theory selected may

Table 2
Treatment effects

	Study 1		Study 2	
Participant	IV	DV	IV	DV
1	1	15	1	15
2	0	5	1	5
3	1	15	0	15
4	0	5	0	5

IV Study 1: 1 = large, 0 = small; IV Study 2: 1 = few reasons, 0 = many reasons.
DV is price paid by a participant (0 = no purchase).

depend on the application context, but some theory is necessary to think through what the intervention should be (Calder & Tybout, 2016). Conversely, being applicable to a user context is an IBE criterion for explanatoriness.

Consider an application context in which the problem is to stimulate consumers to think of the benefits of a product. Take the case of Air Miles, a leading consumer loyalty program in Canada, with 11 million consumers and 300 partner brands. Consumers can redeem their miles for desirable prizes. An intervention Air Miles employs is asking consumers to enter an online contest to win extra miles. The contest requires consumers to state what prize they would redeem with their reward miles and their reason for this choice. The impact of this user-generated content intervention was evaluated in terms of its effect on consumers' subsequent use of the loyalty card. A larger effect on card usage was found when consumers not only named the prize they would redeem their miles for, but also reported their reason for this choice (Malthouse et al., 2016; Malthouse et al., 2016). The intervention had a clear effect.

While using this intervention to affect usage of the loyalty card, the need for better theoretical understanding of the problem also became clear. Asking consumers to elaborate on their reasons for wanting a prize made them more favorable. Would asking them to generate more reasons make them even more favorable? Theory provides a basis for addressing this issue. Consumers may have been relying not on the information content of their reasons but on the feeling of ease. Unless consumers are induced not to rely on feelings of ease, asking for more reasons would be counterproductive. In the absence of theory, Air Miles would have no way to design the intervention with any confidence in the outcome, despite having field study evidence of an effect. Merely obtaining an effect in the real world should not in itself be equated with applicability. Theory makes for applicability, and applicability makes for good theory.

Conclusion

VA focuses on matching constructs to variables to validate that a theory corresponds to an effect. Hence, it treats manipulation check, mediator, and moderator variables as constructs. This focus on the apparent or seemingly likely fit of constructs to variables leads not to improved explanation but to reduced explanatoriness. Contrary to VA, manipulation checks are variables to be explained, not measures of constructs. Mediator variables do not confirm a theoretical process; they require construct-to-construct explanations. When boundary conditions consist of theoretically uninformative moderator variables, they add nothing to explanation. Following VA procedures may seem to make theoretical explanations more likely. But, in so doing, these procedures actually impoverish explanation. Contrary to the view of many reviewers, these procedures, as often employed, should not be considered either necessary or sufficient for identifying the best explanation for an effect.

IBE offers a broader perspective on designing research by offering a rationale for why both theory and effects are necessary. A theory test requires the demonstration of effects, and effects ultimately require theoretical explanation. What is generalized are deductions from the construct-to-construct relations specified by a theory.

In testing theory, IBE emphasizes the criteria of explanatoriness and the use of explanatory convergence procedures across divergent effects. Although such convergence often entails moderation, which demonstrates when independent variables produce a particular effect, IBE can also make use of manipulation checks, mediation analysis, and manipulating mediators. IBE, however, does not prescribe the automatic use of these procedures as a way of reducing constructs to a "likely" restatement of a specific effect. IBE specifies a goal of increasing explanatoriness, which can be achieved by using explanatory convergence procedures and assessed in terms of explanatory power, scope, parsimony, and context.

The current reliance on confirmatory procedures results in explanations that are low in risk of being falsified because the explanatory constructs are yoked to the observed effects at the variable level. However, the reduction in risk due to this logical entailment in which variable and construct are inseparable comes at the cost of low explanatory power, scope, and parsimony. IBE provides explanations with greater explanatory power, but it does so at a higher risk of being wrong. This risk, however, is warranted given that theories can never be proven and that research involves a self-correcting process of advancing bold hypotheses that may be accepted as the best explanation available at some point but are, nonetheless, subject to future falsification and theory development.

IBE provides a bridge between theory and designing consequential real-world interventions. Field studies do not automatically generalize. Large

effects should not equate with importance. Only an understanding of the process by which effects occur can truly increase confidence that an intervention will be effective in a consequential setting.

Proposed IBE Guides for Research Design

In place of relying on the procedural prescriptions of VA, we propose four ways of fostering attention to the IBE criteria of explanatory power, scope, parsimony, and context in assessing theoretical explanations. The following recommendations are not intended to define IBE but to serve as concrete actions authors and reviewers can take to enrich understanding of consumer research phenomena.

- Explanatory Power: Research articles should include "PHs" that clearly distinguish between the predicted effect(s) and theoretical explanation of those effect(s) using the format: PH: _____(Prediction), because _____(Hypothesis).
- Scope: The use of very disparate variables across studies examining a single hypothesis should be encouraged.
- Parsimony: Single explanations over a set multiple studies should be preferred to multiple explanations, provided that the explanation has explanatory power (i.e., imparts meaningful understanding.)
- Context: Articles should replace broad managerial implications with more specific discussion of how the research might be applied to solving specific practical problems.

We proffer these IBE guidelines as an aide to fully realizing the value of both constructs and variables in designing research.

References

Aarts, H., & Dijksterhuis, A. (1999). How often did I do it? Experienced ease of retrieval and frequency estimates of past behavior. *Acta Psychologica*, *103*, 77–89.

Alba, J.W. (2012). In defense of bumbling. *Journal of Consumer Research*, *38*, 981–987.

Amrhein, V., Greenland, S., & McShane, B. (2019). Retire statistical significance. *Nature*, *567*, 305–307.

Bettman, J.R., Luce, M.F., & Payne, J.W. (1998). Constructive consumer choice processes. *Journal of Consumer Research*, *25*, 187–217.

Calder, B.J., Brendl, C.M., & Tybout, A.M. (2019). Integrating effects and theory in research and application. In F. Kardes, P. Herr, & N. Schwarz (Eds.), *Handbook of Research Methods in Consumer Psychology* (pp. 419–437). Routledge.

Calder, B.J., Phillips, L.W., & Tybout, A.M. (1981). Designing research for application. *Journal of Consumer Research*, *8*, 197–207.

Calder, B.J., Phillips, L.W., & Tybout, A.M. (1983). Beyond external validity. *Journal of Consumer Research*, *10*, 112–114.

Calder, B.J., & Tybout, A.M. (2016). What makes a good theory practical? *AMS Review*, *6*, 116–124.

Carpenter, K.J. (1988). *The history of scurvy and vitamin C*. Cambridge: Cambridge University Press.

Danziger, S., Moran, S., & Rafaely, V. (2006). The influence of ease of retrieval on judgment as a function of attention to subjective experience. *Journal of Consumer Psychology*, *16*, 191–195.

Darlington, R. (1990). *Regression and linear models*. McGraw-Hill.

De Vreese, L. (2008). Causal (mis)understanding and the search for scientific explanations: A case study from the history of medicine. *Studies in the History and Philosophy of Biological and Biomedical Sciences*, *39*, 14–24.

Fiedler, K., Schott, M., & Meiser, T. (2011). What mediation analysis can (not) do. *Journal of Experimental Social Psychology*, *47*, 1231–1236.

Frederick, S., Lee, L., & Baskin, E. (2014). The limits of attraction. *Journal of Marketing Research*, *51*, 487–507.

Friedman, M. (1974). Explanation and scientific understanding. *Journal of Philosophy*, *71*, 5–19.

Funder, D.C., & Ozer, D.J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, *2*, 156–168.

Graf, L.K.M., Mayer, S., & Landwehr, J.R. (2018). Measuring processing fluency: One versus five items. *Journal of Consumer Psychology*, *28*, 393–411.

Hauser, D., Ellsworth, P., & Gonzalez, R. (2018). Are manipulation checks necessary? *Frontiers in Psychology*, *21*, https://doi.org/10.3389/psyg.2018.00998

Huber, J. (2008). The value of sticky articles. *Journal of Marketing Research*, *45*, 257–260.

Huber, J., Payne, J.W., & Puto, C.P. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Marketing Research*, *9*, 90–98.

Huber, J., Payne, J.W., & Puto, C.P. (2014). Let's be honest about the attraction effect. *Journal of Marketing Research*, *51*, 520–525.

Janssen, J., Müller, P., & Greifeneder, R. (2011). Cognitive processes in procedural justice judgments: The role of ease-of-retrieval, uncertainty, and experience. *Journal of Organizational Behavior*, *32*, 726–750.

Kitcher, P. (1989). Explanatory unification and the causal structure of the world. In P. Kitcher, & W. Salmon (Eds.), *Scientific explanation* (pp. 410–505). University of Minnesota Press.

Kühnen, U. (2010). Manipulation checks as manipulation: Another look at the ease-of retrieval heuristic. *Personality and Social Psychology Bulletin*, *36*, 47–58.

Lammers, J., Dubois, D., Rucker, D.D., & Galinsky, A.D. (2017). Ease of retrieval moderates the effects of power: implications for the replicability of power recall effects. *Social Cognition*, *35*, 1–17.

Ling, B., Sun, L., & Gu, J. (2017). Effects of ease of retrieval on confirmatory information search. *Social Behavior and Personality*, *45*, 309–320.

Lipton, P. (2004). *Inference to the best explanation* (2nd ed). London: Routledge.

Loken, E., & Gelman, A. (2017). Measurement error and the replication crisis: The assumption that measurement error always reduces effect sizes is false. *Science*, *355* (3625), 584–585.

Lynch, J.G., Alba, J.W., Krishna, A., Morwitz, V.G., & Gürhan-Canali, Z. (2012). Knowledge creation in consumer research: multiple routes, multiple criteria. *Journal of Consumer Psychology*, *22*, 473–485.

MacInnis, D., Morwitz, V.G., Botti, S., Hoffman, D.L., Kozinets, R.V., Lehmann, D.R., Lynch, J.G., & Pechmann, C. (2020). Creating boundary-breaking, marketing-relevant consumer research. *Journal of Marketing*, *84*, 1–23.

Malthouse, E., Calder, B., Kim, S.J., & Vandenbosch, M. (2016). Evidence that user-generated content that produces engagement increases purchase behaviors. *Journal of Marketing Management*, *32*, 427–444.

Malthouse, E., Calder, B., & Vandenbosch, M. (2016). Creating brand engagement on digital, social and mobile media. In R. Brodie, L. Hollebeek, & J. Conduit (Eds.), *Customer engagement* (pp. 85–101). Routledge.

Mayer, J.D., & Gaschke, Y.N. (1988). The experience and meta-experience of mood. *Journal of Personality and Social Psychology*, *55*, 102–111.

McShane, B., Gal, D., Gelman, A., Robert, C., & Tackett, J. (2019). Abandon statistical significance. *The American Statistician*, *73*, 235–245.

Morwitz, V.G., & Fitzsimmons, G.J. (2004). The mere-measurement effect: Why does measuring intentions change actual behavior? *Journal of Consumer Psychology*, *14*, 64–73.

Novemsky, N., Dhar, R., Schwarz, N., & Simonson, I. (2007). Preference fluency in choice. *Journal of Marketing Research*, *44*, 347–356.

Payne, J.W. (1982). Contingent decision behavior. *Psychological Bulletin*, *92*, 382–402.

Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. New York: Basic Books.

Peracchio, L.A., Luce, M.F., & McGill, A.L. (2014). Building bridges for an interconnected field of consumer research. *Journal of Consumer Research*, *40*, v–viii.

Pham, M.P. (2013). The seven sins of consumer psychology. *Journal of Consumer Psychology*, *23*, 411–423.

Popper, K.R. (1954). Degree of Confirmation. *The British Journal of Philosophy of Science*, *5*, 143–149.

Robinson, A. (2018). Did Einstein really say that? *Nature*, *557*, 30.

Rothman, A.J., & Schwarz, N. (1998). Constructing perceptions of vulnerability: Personal relevance and the use of experiential information in health judgments. *Personality and Social Psychology Bulletin*, *24*, 1053–1064.

Ruder, M., & Bless, H. (2003). Mood and the reliance on the ease of retrieval heuristic. *Journal of Personality and Social Psychology*, *85*, 20–32.

Schupbach, J.N. (2017). Inference to the best explanation: Cleaned up and made respectable. In K. McCain, & T. Poston (Eds.), *Best explanations: New essays on Inference to the best explanation* (pp. 39–64). Oxford University Press.

Schwarz, N. (1998). Accessible content and accessibility experiences: the interplay of declarative and experiential information in judgment. *Personality and Social Psychology Review*, *2*, 87–99.

Schwarz, N. (2004). Metacognitive experiences in consumer judgment and decision making. *Journal of Consumer Psychology*, *14*, 332–348.

Schwarz, N. (2012). Feelings-as-information theory. In P.A. Van Lange, A.W. Kruglanski, & E. Higgins (Eds.), *Handbook of theories of social psychology* (pp. 289–308). Sage.

Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social Psychology*, *61*, 195–202.

Spektor, M.S., Kellen, D., & Hotaling, J.M. (2018). When the good looks bad: An experimental exploration of the repulsion effect. *Psychological Science*, *29*, 1309–1320.

Stepper, S., & Strack, F. (1993). Proprioceptive determinants of emotional and nonemotional feelings. *Journal of Personality and Social Psychology*, *64*, 211–220.

Tan, K., & Agnew, C.R. (2016). Ease of retrieval effects on relationship commitment: the role of future plans. *Personality and Social Psychology Bulletin*, *42*, 161–171.

Tormala, Z.L., Falces, C., Briñol, P., & Petty, R.E. (2007). Ease of retrieval effects in social judgment: The role of unrequested cognitions. *Journal of Personality and Social Psychology*, *93*, 143–157.

Trope, Y. (2004). Theory in social psychology: Seeing the forest and the trees. *Personality and Social Psychology Review*, *8*, 193–200.

Trueblood, J.S., Brown, S.D., Heathcote, A., & Busemeyer, J.R. (2013). Not just for consumers: context effects are fundamental to decision making. *Psychological Science*, *24*, 901–908.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*, 207–232.

Tybout, A.M., Sternthal, B., Malaviya, P., Bakamitsos, G.A., & Park, S. (2005). Information accessibility as a moderator of judgments: The role of content versus retrieval ease. *Journal of Consumer Research*, *32*, 76–85.

van Fraassen, B. C. (1980). *The Scientific Image*. Oxford, UK: Oxford University Press.

Van Lange, P.A. (2013). What we should expect from theories in social psychology: truth, abstraction, progress, and applicability. *Personality and Social Psychology Review*, *17*, 40–55.

Wänke, M., Bless, H., & Biller, B. (1996). Subjective experience versus content of information in the construction of attitude judgments. *Personality and Social Psychology Bulletin*, *22*, 1105–1113.

Wänke, M., Bohner, G., & Jurkowitsch, A. (1997). There are many reasons to drive a BMW: Does imagined ease of argument generation influence attitudes? *Journal of Consumer Research*, *24*, 170–177.

Watson, D., Clark, L.A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070.

Weingarten, E., & Hutchinson, J.W. (2018). Does ease mediate the ease-of-retrieval effect? A meta-analysis. *Psychological Bulletin*, *144*, 227–283.

Winkielman, P., Schwarz, N., & Belli, R.R. (1998). The role of ease of retrieval and attribution in memory judgments: Judging your memory as worse despite recalling more events. *Psychological Science*, *9*, 124–126.