

What warrants our claims? A methodological evaluation of argument structure

Mikko Ketokivi¹  | Saku Mantere²

¹IE Business School, IE University, Madrid, Spain

²McGill University, Montreal, Quebec, Canada

Correspondence

Mikko Ketokivi, IE Business School, IE University, Maria de Molina 12-5, 28006 Madrid, Spain.
Email: mikko.ketokivi@ie.edu

Handling Editor: Mikko Rönkkö

Abstract

The process of justifying a generalized theoretical conclusion from a specific empirical analysis continues to elude us. In this article, we suggest that this stems from an incomplete understanding and specification of how arguments are structured. Most importantly, in addition to empirical data, a generalized conclusion hinges on the application of various rules and principles of reasoning that British philosopher Stephen Toulmin labeled *warrants*. In this article, we apply Toulmin's model of argument structure to empirical management research by examining in particular the roles of four types of warrants: theoretical, inferential, procedural, and contextual. Based on our analysis, we suggest that making warrants and their backings explicit paves the way toward a more comprehensive understanding of how arguments are structured and how claims are justified. Importantly, an examination of warrants reveals that the choices researchers make are not limited to matters such as choosing the research topic or a particular research design, but they also extend to how we produce our claims. If we wish to understand argument structure, we must understand these choices.

KEYWORDS

argumentation, reasoning, warrant, methodology, theory development

1 | INTRODUCTION

Should factories be focused (Skinner, 1974)? Should transactions supported by specific assets be internalized (Williamson, 1985)? It is well established that generalized theoretical claims relating to questions such as these cannot rest on empirical data alone; in the philosophy of science literature, this is known as *underdetermination of theories by data* (Laudan, 1990; Mill, [1843] 1882; Quine, 1951; Stanford, 2017). This means that the same empirical data can be interpreted as evidence for several, even conflicting theoretical explanations (Ketokivi & Mantere, 2010).

Underdetermination is significantly driven by the fact that when scholars use empirical material to draw

conclusions, they do not simply *derive* the conclusion from the data as much as they *reason* it from the data. This is not a matter of semantics: Derivation of a conclusion is typically algorithmic and computational, but reasoning from data is an essentially cognitive process that involves much more than just the data. In the context of scientific arguments, these cognitive processes further involve multiple parties, such as authors and peer reviewers. If we wish to understand how knowledge claims are made, evaluated, and accepted (or rejected), we must understand these processes.

In the field of management and organization studies, one sometimes encounters claims that empirical data are indeed sufficient to produce a knowledge claim, which directly clashes with the underdetermination thesis. A

case in point, some case researchers suggest that theories “emerge” inductively from empirical material: “Case study theory building is a bottom up approach such that the specifics of data *produce* the generalizations of theory” (Eisenhardt, 1989, p. 547, our emphasis). Indeed, Eisenhardt and Graebner (2007, p. 25) suggest that multiple-case research is “surprisingly *objective*.” Objectivity rests on the idea that the data “provide the discipline that mathematics does in formal analytic modeling” (Eisenhardt & Graebner, 2007, p. 25).

If Eisenhardt and Graebner are correct and the data indeed *produce* theoretical generalizations, management theories do not in fact suffer from underdetermination: In order to make a theoretical claim, all we would have to do is to examine the data. This is an extraordinary claim, which, to invoke the old maxim, requires extraordinary evidence. The claim that the data provide not only the raw material but indeed “the discipline” to arrive at the claim requires much elaboration, particularly since it has been well established, in both mundane and scientific contexts, that there are always multiple ways of accounting for what one has observed. Indeed, this was recognized already by early 19th-century empiricists, such as John Stuart Mill ([1843] 1882, p. 617): “Most thinkers of any degree of sobriety allow, that an hypothesis [...] is not to be received as probably true because it accounts for all the known phenomena, since this is a condition sometimes fulfilled tolerably well by two conflicting hypotheses [...] while there are probably a thousand more which are equally possible, but which, for want of anything analogous in our experience, our minds are unfitted to conceive.” This is the essence of underdetermination.

We understand the appeal in both the idea of obtaining objective knowledge and the promotion of empirical data as the foundation of this knowledge. However, calls to seek objectivity and discipline *in one's data* are ultimately a cause for concern, because they effectively invite us to ignore the choices researchers make in their reasoning. The existence of alternative explanations and interpretations of the very same data has profound implications both for the authoring and the evaluation of claims. Specifically, if the data are insufficient to provide the foundation of a knowledge claim, we must ask: What are all the elements required to warrant the claim? This is the central question examined in this article.

It is astounding to us that as fundamental as this question is to the practice of scientific research, it is hard to come by explicit, systematic evaluations of how management scientists warrant their claims. Conspicuously missing are, in particular, *methodological* evaluations that examine the author's reasoning, that is, the cognitive processes in which the author combines and refines ideas, and ultimately, arrives at the central knowledge claims.

What are the roles of empirical data, theoretical logic, and background assumptions in these processes? What role do various *paradigms* have? Further, because we are evaluating scientific claims, the focus should be specifically on forms of *scientific reasoning*, that is, deduction, induction, and abduction (Mantere & Ketokivi, 2013). But how are these forms of reasoning used, where, and to what ends? In which instances is reasoning algorithmic, and in which instances is it cognitive? These questions are ultimately empirical, but we propose that an effective way of developing an understanding of how knowledge claims are produced and where disagreement may originate is to explicate and evaluate the structure of our arguments.

This article is structured as follows. We start by arguing for the need for explicit *warrants* (Toulmin, [1958] 2003) as an integral part of argument structure. We do this both by a general examination of the claims researchers make as well as by an empirical example. We then present a general framework for argument structure, with emphasis on the types of warrants management researchers use as they are seeking to convince their audiences. We conclude by examining two examples from empirical operations management research and by discussing the implications for research practice.

2 | WHAT ARE WARRANTS AND WHY DO WE NEED THEM?

In the management sciences, authors seek to make generalized theoretical claims based on empirical inquiry in a specific context. A sense of generality is crucial, because in order for an empirical inquiry into Honda's supply-chain practices to be academically relevant, the conclusions and claims made cannot be just about Honda's supply-chain practices; the central claims must transcend the empirical context examined. Here are a few examples:

1. Factories should be focused in their manufacturing tasks and operations strategies must align with competitive strategy (Skinner, 1969);
2. Innovative products require a responsive and functional products an efficient supply chain (Fisher, 1997); and
3. Transactions associated with high asset specificity, high frequency of transacting, and high uncertainty should be conducted within the firm, not across firms (Grover & Malhotra, 2003; Williamson, 1985).

Both the central terms and the logic in all these claims are theoretical. For example, Skinner's notion of *focus* is a theoretical, not an empirical concept.¹ While we

ultimately want to examine focus empirically by operationalizing it as an empirical measure, the measure is not so much *derived* from the theoretical concept as it is *chosen* based on contextual considerations, availability of data, and other conspicuously atheoretical concerns. Substantive theories seldom tell us what an empirical inquiry of the theory should entail. The key question then becomes: What kind of an argument structure supports a generalized conclusion from a specific empirical investigation? It seems methodologically implausible that going from the specific (empirical) to the general (theoretical), or vice versa, is possible without facing multiple choice situations.

To clarify, it is obvious that researchers make choices, such as which topics to study, what literature to incorporate, and so on. But these are choices related more to preference and policy than to methodology. In this article, we focus on the choices we make as we engage in reasoning that bridges the specific (the data) with the general (the theoretical claim). We argue that choice is an essential part of the process of justifying the claim and that this choice must be both informed and explicit. In order to establish this position, we draw on British philosopher Stephen Toulmin's ([1958] 2003, p. 12) concept of *warrant*.

What are warrants, where do they originate, and what is their role in the argument? We begin tackling this central question by recounting one of the best-known examples of competing warrants in the history of management research: What are the general theoretical accounts that explain why firms integrate, and what conclusions are warranted in specific empirical instances? The example should establish warrants in particular and argument structure in general to be matters of both academic and practical relevance.

2.1 | The case of Schwinn bicycles

In the 1960s, bicycle manufacturer Schwinn focused on the design and production of bicycles, and outsourced retail sales to companies such as B.F. Goodrich. This arrangement made sense, because retail sales could easily be organizationally and operationally decoupled from design and production. As part of this decision, Schwinn chose its retail partners carefully by exercising considerable influence over who could sell its products, in what territories, and at what price. The question of how these actions should be interpreted became a topic of intense debate, and ultimately, gave birth to Transaction Cost Economics, one of the most influential economic theories of organization. Most importantly for the purposes of this article, the Schwinn case serves as an introduction to the structure of arguments and the use of warrants.

The U.S. Supreme Court was concerned that Schwinn's restrictions were aimed at limiting competition, and consequently, were in violation of the Sherman Antitrust Act of 1890. Indeed, in 1967, the U.S. Supreme Court ruled Schwinn's supply-chain organization illegal: "[Schwinn was] charged by the Government with a continuing conspiracy, with others, to fix prices, to allocate exclusive territories to wholesalers and jobbers, and to confine merchandise to franchised dealers" (from *United States v. Arnold, Schwinn & Co.*, 388 U.S. 365).

At the time of the Schwinn trial, Oliver Williamson, an emerging organization economist (who would later become an Economics Nobel Laureate), worked as a Special Economic Assistant to the Head of the Antitrust Division of the U.S. Department of Justice. Williamson retrospectively called the Supreme Court's Schwinn ruling "deeply confused" (Williamson, 2002, p. 9): While Schwinn's actions could be interpreted as anticompetitive, they could also plausibly be interpreted as attempts at *efficient* transacting (for a review of efficient contracting in supply chains, see Ketokivi & Mahoney, 2020). In sum, two parties looking at the same evidence drew not only different but more or less opposite, mutually exclusive conclusions.

How did the two parties arrive at their respective conclusions? To understand the Supreme Court's position, one must first understand that much of the thinking in post-World War II antitrust cases in the United States was based on theories of *industrial organization (IO) economics* which—oddly enough for the management researcher—did not incorporate the idea of firm-level strategies; the focus was on industry-level phenomena and industry-level analysis. In these analyses, the firm was typically considered a *production function* (Bronfenbrenner, 1944), a mathematical equation that expresses the relationship between economic inputs (capital, labor) and outputs (products). The idea that two firms serving the same market would offer different kinds of products was something that economic theories could not explain: In the IO economics perspective, all firms serving a specific market would be expected to be identical, save for differences in scale (Porter, 1981, p. 612). Differentiation strategies were considered both anomalous and suspicious. Consequently, when Schwinn did something qualitatively different from others, the Supreme Court, struggling to make sense of Schwinn's idiosyncratic strategy, concluded its actions were questionable (Coase, 1988, p. 67).

Williamson took a fundamentally different position by arguing that Schwinn plausibly sought to organize its supply chain as efficiently as possible. Instead of speaking of *restraints*, it would be more reasonable to use the benign term *supply-chain coordination* to describe

Schwinn's actions (cf. Cachon, 2003; Kouvelis et al., 2009). In order to secure the delivery of a high-quality bicycle and an efficient after-sales service, the supply chain would have to remain under Schwinn's supervision, even if Schwinn no longer owned the bicycles. This is because a Schwinn bicycle would always represent the Schwinn brand, no matter who held title to it: "[Q]uality reputation may be preserved only if goods and services are sold under conditions of constraint" (Williamson, 1985, p. 186).

Who is right, the Supreme Court or Williamson? If we adopt the premise of viewing the firm as a production function and the associated IO economics perspective, the Supreme Court's position has more merit. Specifically, the IO economics perspective permits us to dismiss the notion of firm-specific quality-reputation effects as irrelevant. However, if we adopt the efficient-contracting perspective, conduct an explicit firm-level analysis, and acknowledge the possibility of a variety of legitimate firm-level strategies, Williamson's position gains traction. In Williamson's analysis, the firm is not viewed as a production function, but as a *governance structure*, a fundamentally different core analogy than the production function analogy (Ketokivi, Mantere, & Cornelissen, 2017).

For the purposes of this article, the key insight is that neither position can be justified without invoking something other than the data. This *something* is a theoretical lens that is chosen to provide structure to how observations are interpreted. Williamson's central point in the Schwinn case is that the reasoning embedded in these choices be made explicit: "[I]n the occasional case where efficiency and market power consequences exist, can economies be dismissed on the grounds that market power effects invariably dominate? If they cannot, then a rational treatment of [the question] requires that an effort be made to establish the [economic efficiency] implications" (Williamson, 1968, pp. 18–19). In reasoning terminology, Williamson suggested that we must not accept a claim without explicating and critically evaluating the warrants that support it.

The Schwinn case illustrates one of the four types of warrants researchers use: In presenting theoretical conclusions, the researcher adopts a specific *theoretical warrant* that provides the interpretive lens through which evidence is examined. The two competing theoretical warrants in the Schwinn case are *IO economics* (Bain, 1968) and *organization-economics* (Williamson, 1985). Choosing which warrant to adopt is fundamentally just that—a matter of choice.

In addition to theoretical warrants, we offer three others: procedural, contextual, and inferential. Like theoretical warrants, these other three function in a similar role in arguments in that they help to bridge empirical

grounds to theoretical claims. Our goal in this article is to elucidate argument structure by exploring the use of the four types of warrants. We start by introducing Toulmin's general model of argument, as it offers a useful tool for analyzing argument structure and for showing how warrants operate.

3 | MAPPING ARGUMENT STRUCTURE: THE TOULMIN MODEL

In two ground-breaking books *The Uses of Argument* ([1958] 2003) and *An Introduction to Reasoning* (1979), Toulmin and his colleagues presented a general structure of arguments that can be applied to various contexts, including scientific research (Toulmin et al., 1979, chapter 27). Toulmin's general argument structure consists of six main elements:

1. *Grounds* are the data, facts, and evidence "on which the merits of [the claim] are to depend" (Toulmin, [1958] 2003, p. 12).
2. *Claims* are the ultimate conclusions and assertions made.
3. *Warrants* are "the practical standards or canons of argument" (Toulmin, [1958] 2003, p. 12) that grant the "license" (Toulmin et al., 1979, p. 48) to assert the claim in the specific case.
4. *Backings* are the general principles that establish why the "warrant should be accepted as having authority" (Toulmin, [1958] 2003, p. 95). This conveys the important idea that "warrants are not self-validating" (Toulmin et al., 1979, p. 62).
5. *Qualifiers* (or *Modalities*) are "phrases that show what kind and degree of reliance is to be based on the conclusions, given the arguments available to support them" (Toulmin et al., 1979, p. 85). For example, is the conclusion a *certain* or merely a *plausible* implication of the grounds and the warrants? Qualifiers typically appear embedded in the claim, as in "it is *evidently* the case that..."
6. *Rebuttals* are the "extraordinary or exceptional circumstances that might undermine the force of [the argument]" (Toulmin et al., 1979, p. 95). Toulmin posited that rebuttals are often aimed not at the claim in its entirety, but rather, at the qualifier (Toulmin et al., 1979, p. 98). For example, one might argue that the claim made is merely *plausible* as opposed to *evident*.

Toulmin's framework highlights the fact that arguments have both idiosyncratic and more general characteristics.

For example, researchers often seek *novel* claims by using *original* datasets, which suggests that grounds and claims are idiosyncratic. Similarly, rebuttals are idiosyncratic in that they pertain to *extraordinary* circumstances. In contrast, warrants and backings are general *rules* and *principles* applied across arguments. Moreover, using qualifiers such as *certain*, *likely*, or *probable* links directly to the use of probabilities (Toulmin, [1958] 2003, pp. 41–86), which is based on mathematics (deduction) and statistics (induction); qualifiers thus belong to the category of more general principles. Toulmin further argued that these qualifiers must be used rigorously: Stating something is *probable* instead of *plausible* is not a matter of semantics, it is essential to the content of the argument (Toulmin, [1958] 2003, p. 69). Here, we see a straightforward link from the use of qualifiers to various forms of scientific reasoning: Deduction leads to *certain*, induction to *probabilistic*, and abduction to *plausible* claims (Mantere & Ketokivi, 2013).

The relationships among the six elements are depicted in Figure 1 (Toulmin et al., 1979, p. 98). The logic that links the six elements to one another is the following: Given the *grounds*, we appeal to a *warrant* (with the proper *backing*) to assert a *qualified claim* (subject to *rebuttal*) to our audience (Toulmin et al., 1979, p. 98). From an epistemological perspective, it is crucial to highlight that the only element of the argument which we *know* are the grounds. For example, we do not know the claim, we *assert* it; we do not know the warrant, we *appeal* to it.

The grounds of the argument can be thought of as the facts of the case, or, “the common ground” (Toulmin et al., 1979, p. 38) that must be established before argument evaluation can begin. In empirical research, this involves establishing the validity of one’s data, a step that is logically prior to any further analysis or test of

hypotheses: If the author and the audience cannot agree the data to be valid, evaluation of the claim becomes a moot point. The claim, presented as the conclusion of the analysis, may be a generalization, an explanation, an interpretation, or a policy prescription, depending on the research interest.

Let us apply Toulmin’s framework to the Schwinn case. Figure 2 depicts the structure of the antitrust and the efficiency arguments. To be sure, these depictions are stylized, but our point is more illustrative than substantive. Note that the grounds indeed constitute “the common ground” in the two arguments, which makes the comparison of the arguments meaningful.

As Figure 2 shows, there is no disagreement on the facts of the case: Both sides of the argument acknowledged that Schwinn implemented vertical restraints. But, the arguments differ in that different actors adopted different warrants and asserted different claims. The differences stem from the warrants resting on different intellectual and institutional bases, which effectively illustrates the importance of understanding *backings*. The backings in the antitrust argument are found both in theories of industrial organization as well as legal interpretation, history and precedent. In contrast, the warrant in the efficiency argument receives its backing from firm-level theories of governance, and in particular, Transaction Cost Economics. It is also worthwhile to note that the central warrants used are both based on reasoning by analogy: Whereas the Supreme Court thought of the firm *as if it were a production function*, Williamson’s choice was to think of the firm *as if it were a governance structure*. Importantly, both warrants were based on the premise of viewing the firm *as if it were something comparatively simpler*. This is understandable, but at the same time, such simplification never does full justice to the complexity of organizations (Ketokivi, Mantere, & Cornelissen, 2017).

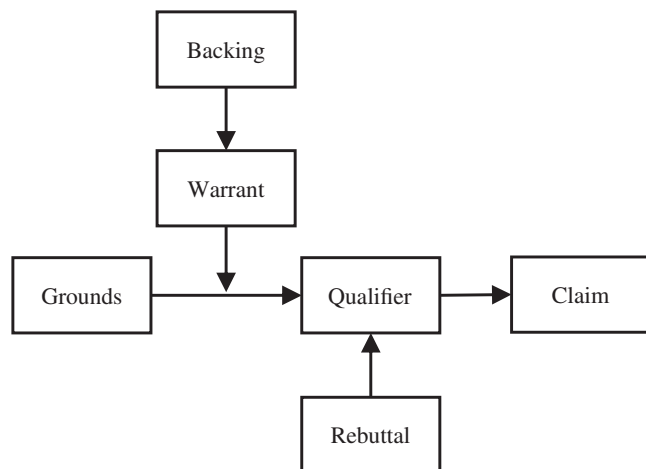
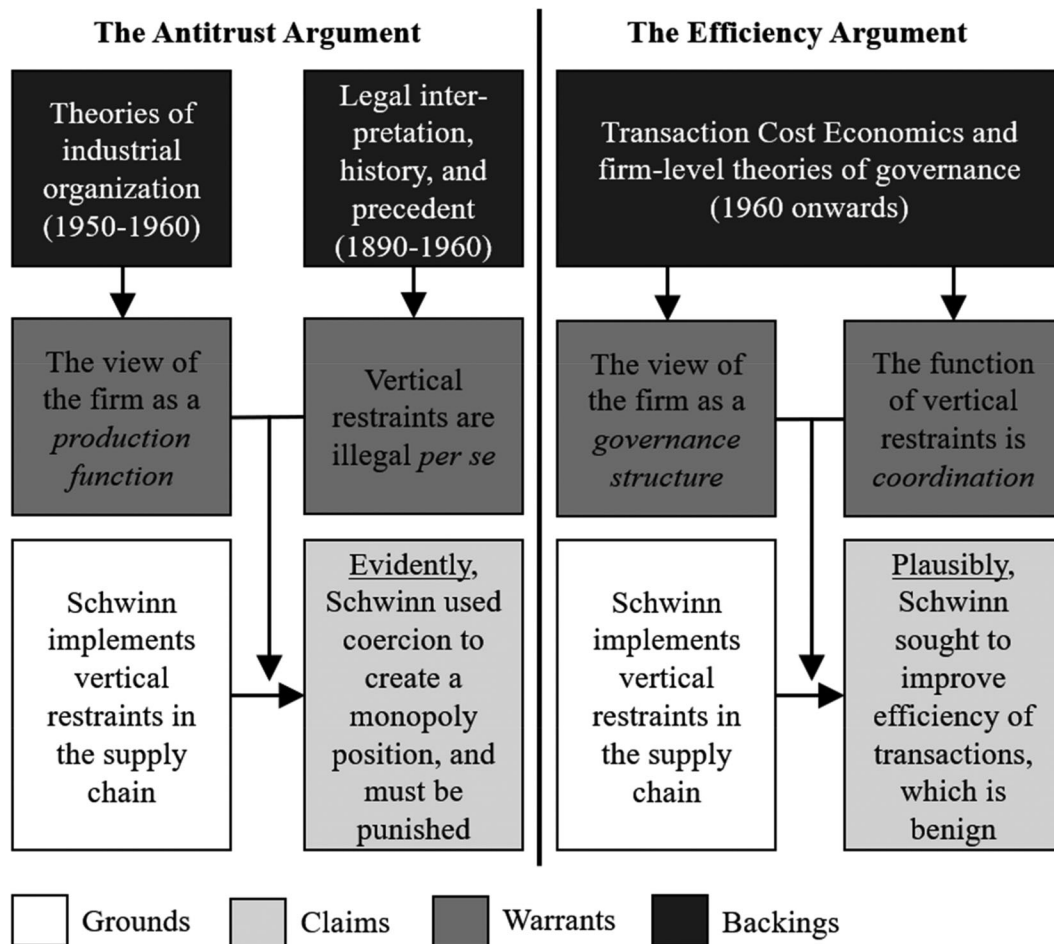


FIGURE 1 Toulmin’s model of argument

3.1 | Why do we not explicate argument structure?

There are very few references to Toulmin, or any other expositions of argument structure, in the management literature. Why is argument structure simply “left understood,” as Toulmin ([1958] 2003, p. 98) put it? We identify two reasons, both of which are understandable, but neither of which is methodologically compelling.

One reason is *practical*, expressed by Toulmin ([1958] 2003, p. 98) himself: “[I]f we demanded the credentials of all warrants at sight and never let one pass unchallenged, argument could scarcely begin.” This may be descriptively accurate in that the desire to expedite the



Notes: (1) Qualifiers are the underlined words in the claims; (2) The Efficiency Argument constitutes a rebuttal of the unconditional qualifier of the Antitrust Argument

FIGURE 2 The structure of the antitrust and efficiency arguments in the Schwinn case

process of proceeding from the grounds to the claims is understandable (e.g., Gehman et al., 2017, p. 10). At the same time, a community of scientists must not accept convenience and expedience as its intellectual basis; in research, we seek to construct credible knowledge claims. Anyone who has been involved in a peer-review process knows that there is nothing convenient or expedient about it. The problem is that without an explicit account of warrants and their backings, our understanding of how the grounds give rise to the claims remains incomplete.

The other reason for omitting warrants can only be described as a *methodological fallacy*: assuming that the grounds (data) alone are ultimately sufficient to support the (theoretical) claim. This assumption rests on the notion that knowledge is readily embedded in the data from which it emerges by proper application of the scientific method. However, it is well established that this is an overly romanticized view of how science works. In

reality, there is no such thing as *the* scientific method (Bird, 1998, p. 156); this can be thought of as an alternative way of arguing underdetermination. A look at research practice quickly reveals that convincing an audience of the credibility of a claim is a matter of justification that requires not only data but also an in-depth understanding of the context in which the knowledge claims are produced and presented (Feyerabend, 1978; Longino, 1990).

One of the central characteristics of claims made in management research is the use of unobservable theoretical terms. It should be obvious that unobservables cannot simply be derived from observables, or vice versa, because this would make unobservables ultimately observable. But unobservables cannot be derived from the observables, they must be actively *constructed* by the researcher. Of many notable 20th-century philosophers of science who wrote about the structure and construction of theoretical claims, Willard Van Orman

Quine (1960) noted that the process of moving from empirical data to a theoretical conclusions involves interpretation and *translation*, not algorithmic reasoning. Further, in this translation—just like in any translation—some meaning is unavoidably lost and other meaning is created.

3.2 | It is the researcher, not the algorithm, that reasons: From induction to abduction

If there is no single, agreed-upon scientific method to guide inquiry, what is the form of reasoning used in inferring the theoretical claim from the empirical grounds? Both management researchers (Mantere & Ketokivi, 2013) and philosophers (Lipton, 2004) have pointed out that the construction of explanations and interpretations involves *abductive reasoning*, which, unlike deduction or induction, is not an algorithmic or computational but a *cognitive* activity that typically extends significantly beyond the evidence. When applied to scientific reasoning, abduction often involves a choice between alternative explanations or interpretations (Harman, 1965). Importantly, this choice is not driven by the data, but rather, the principles of the local scientific community in which the argument is presented. This choice is further justified by appealing to the *epistemic virtues* (Lycan, 1998) embraced by the community. For example, organization-economic communities endorse formalization, parsimony, and quantitative analysis, whereas anthropologists and ethnographers promote thick description, nuance, and context-specific interpretation. Differences aside, all researchers engage in abductive reasoning in crucial phases of research (Mantere & Ketokivi, 2013).

3.2.1 | Abductive reasoning in the Schwinn case

In the Schwinn case, the antitrust warrant provides an argumentation context in which it is legitimate to interpret Schwinn's vertical restraints as anticompetitive; the efficiency warrant in turn provides a context for the efficient-contracting claim. To see why contextualization is essentially abductive reasoning, consider the general form of abduction (Niiniluoto, 1999, p. S439): (a) The surprising fact C is observed; (b) but if A were true, C would be a matter of course; (c) hence, there is a reason to suspect that A is true.

The general structure of abductive reasoning can be illustrated by restating the antitrust argument in the abductive reasoning form (the efficiency argument could

be similarly cast in the same form): (a) The (surprising) fact that Schwinn implements vertical restraints is observed; (b) but if Schwinn were seeking a monopoly position, then the restraints would be a matter of course; (c) hence, there is a reason to conclude that Schwinn has implemented vertical restraints to establish a monopoly position. The final step in the antitrust argument is to invoke what is known as the *per se illegality* warrant (Bork, 1965) to assert that Schwinn *evidently* (the qualifier in the argument) broke the law, and therefore, should be punished. The *per se illegality* principle holds that certain actions—in the Schwinn case, territorial restrictions and price-fixing—are presumptively illegal. In addition to Schwinn's actions (of territorial restrictions and price-fixing), the *per se* warrant is required to arrive at the unconditional qualifier *evidently*.

It is important to explicate the abductive reasoning process, because just like all scientific reasoning, it must be critically evaluated for rigor. This is crucial, because no matter how compelling the assertion may be on its face, the form of abductive reasoning is—from a strictly logical point of view—an instance of the *fallacy of affirming the consequent* (Niiniluoto, 1999, p. S442). Further, Niiniluoto (1999, p. S439, emphasis added) reminded that in abduction, “the conclusion is not A itself, but the *assertion* that there is a reason to suspect that A is true.” Again, this is not just semantics: In evaluating an argument in peer review, for example, the task of the reviewer is not to try to determine whether the claim made is true, but rather, whether it is justified in light of the grounds and the warrants. It should be obvious that the peer review offers no guarantees that the claims will not be disputed later; the aim is to ensure that the reasoning that produces the claim is explicit and acceptable (Peirce, 1955).

In order for an assertion to be accepted by one's audience, its author must achieve two distinct objectives: (a) establish the validity of the grounds (the data); and (b) justify the adoption of the specific warrant. In research practice, we tend to focus on the former but leave the latter implicit, or simply, understood. This is problematic, because it is specifically the warrant that provides the context in which the abductive process enables the author to proceed from the grounds to the claim. Note further that the differences in the two explanations for Schwinn's actions are not about inter-*individual* differences, rather, they stem from differences in the warrants. Warrants in scientific debates in particular do not, therefore, have a psychological but an institutional basis: They constitute the agreed-upon *standards* by which the grounds are evaluated in the specific context.

In this article, we deliberately avoid using the word “objective,” because it is easily interpreted as implying a universal standard. It is more accurate to describe

warrants as the *inter-subjective* rules and principles that researchers in a local community share. Indeed, the entire subjective-objective distinction is irrelevant here, because warrants are strictly speaking neither. We propose that the descriptively accurate term is *paradigmatic*, and accordingly, the focus should be on inter-paradigmatic differences in reasoning.

3.2.2 | Warrants are not Kuhnian paradigms

We use the word *paradigm* with an important caveat, which should be made explicit to avoid confusion. Specifically, we do not use the concept in the same meaning as Kuhn (1962). For Kuhn, inter-paradigmatic differences were so fundamental that they involved differences at the ontological level. For instance, in Newton's paradigm, mass is considered a property of a physical object, whereas in Einstein's paradigm, mass is relative to contextual considerations such as velocity.

Treating IO economics and organizational economics as two different Kuhnian paradigms would be hyperbolic and misguided. It would be hyperbolic, because the two theories are ontologically commensurate. And it would be misguided for this same reason: Even though the two theories lead us to different interpretations in the Schwinn case, they are invoked precisely for the purpose of providing alternative explanations of the *same phenomenon*, that is, Schwinn's decision to impose vertical restraints. It is specifically paradigm commensurability that makes the use of several paradigms in a point-counterpoint argumentation possible. If the paradigms were ontologically incommensurate (as in Kuhn's formulation), we do not see how they could be invoked to address the same phenomenon. If two theories do not address the same phenomenon, how can argumentation across paradigms even begin?

To use Kuhn's (1962) terminology, Figure 2 is an example of *normal science* in action, not a juxtaposition of competing paradigms. Similarly, it would be misleading to describe the shift from an IO economics to an organization economics perspective as a *Kuhnian revolution*. A Kuhnian revolution involves a change so radical that the old and the new paradigms simply cannot be compared to one another, resulting in a communication breakdown (Kuhn, 1962: chapter 10). No such breakdown is evident in the Schwinn example.

With these caveats, we choose to use the term *paradigm* in this article, because we find it a useful descriptive label, and because Kuhn certainly cannot be afforded monopoly for using the term. Our use of the term is closer to Hungarian philosopher Imre Lakatos' (1970) (methodological) notion of *research program* than Kuhn's (historical) notion of *paradigm*. IO economics and

organizational economics can in our view accurately be described as two different Lakatosian research programs.

3.3 | Underdetermination of theory: The case of the focused factory

Let us now turn to an examination of underdetermination in the context of a familiar operations management example by considering empirical research on the focused factory as an instance of translating between the theoretical and the empirical languages. Skinner's central prescription is perhaps the most conspicuous generalized theoretical claim: "Centralize the factory's focus on relative competitive ability" (Skinner, 1974, p. 119). Other claims are less conspicuous, but they still clearly go beyond simply reporting findings or empirical generalizations. For example, when Vokurka and Davis (2000, p. 54) claimed that "the focused plant concept can be very beneficial to organizations," they clearly stated more than an empirical result. Specifically, the conclusion is in its essence a generalized theoretical proposition that uses theoretical language. If Vokurka and Davis had presented an empirical proposition, they would have used the language of the empirical analysis adopted in their study. In order to understand the empirical language used, we must delve deep into the specifics of the context, the dataset, the variables, and their interrelationships.

In the Vokurka and Davis study, the central empirical finding was that when a statistical sample of 294 U.S. manufacturing plants in diverse industries is partitioned into focused and unfocused plants based on managerial self-classification, there are statistically significant differences in a total of 20 self-reported performance metrics. Specifically, managers of self-reportedly focused plants perceive, on average, their plants to be more productive than those with whom they directly compete. This is the central empirical finding. Bridging this empirical generalization to the claim that "the focused plant concept can be very beneficial to organizations" requires translation, which must be made explicit.

One of the reasons translation is required is because the theoretically central concept *focus* does not have a one-to-one correspondence with any empirical measure. To understand this, let us look at some of the issues related to its operationalization. Empirically, the most salient operationalization of focus would be product variety, and indeed, focused plants tend to have fewer final products (Pesch & Schroeder, 1996; Vokurka & Davis, 2000). However, as Skinner argued on several occasions, focus is not about the scope of outputs, it is about the scope of "the task" (Skinner, 1996, p. 3). While there may be a connection between a focused

(manufacturing) task and product mix, a narrow product mix is neither necessary nor sufficient for the task to be focused. Therefore, we simply cannot derive the empirical concept of narrow product mix from the theoretical concept of focus. However, we can see how in a specific context, focus might plausibly translate into product mix, and conversely, narrow product mix could plausibly be interpreted as focus. Importantly, *translating* focus into narrow product mix or *interpreting* narrow product mix as focus are methodologically categorically distinct from *deriving* one from the other.

As an example, in trying to think of what the notion of focus meant in the process industry context, Ketokivi and Jokinen (2006, p. 252) reasoned that because in the process industry, a given production line is typically dedicated to a specific product (this is not true in many other contexts), narrow product mix tends to imply technological focus. Technological focus can further be reasoned to operationalize Skinner's notion of focus in a way that maintains its theoretical meaning. Ketokivi and Jokinen's reasoning process is conspicuously abductive.

4 | THE DIVERSITY OF WARRANTS

The bridge from grounds to claims consists of an intricate set of distinct, mutually complementary warrants that rest on various theoretical, methodological, epistemological, contextual, and practical backings. In order for argument structure to be maximally transparent, we must explicate these warrants and their backings. Or put in more practical terms, even though we may not wish to

critically evaluate all warrants and their backings in every single instance of argument, we must have an agreed-upon understanding of what these warrants and their backings are, how they operate, and what kinds of boundary conditions they set for the argument.

The Schwinn example illustrates the use of warrants that are *theoretical*, because they specifically enable the construction of *explanations* that account for the grounds. In addition to theoretical warrants, we can see at least three other categories of warrants: inferential, procedural, and contextual. We discuss all four categories in the following, starting with theoretical warrants. The categories are illustrated in Figure 3 and Table 1.

The purpose of the following is to offer a conceptual typology of warrants, not to prescribe how they should be used in research practice; the use of warrants always involves exercise of judgment in a specific context. Further, the four types of warrants are not used in arguments in any specific order, nor are all four necessarily invoked in each argument. But in order to present a credible claim based on some empirical grounds, at least some warrants will be required. This is significantly a matter of choice, and can also be expected to be influenced by the negotiation by scientific audiences; in publishing results, this occurs in the review process.

4.1 | Theoretical warrants: Invoking theoretical paradigms

Arguments based on empirical research involve theoretical interpretation of the grounds (the data). The researcher chooses theoretical warrants based on their ability to provide explanations within a particular research program. For example, in the economic discourse on organizational boundaries, a researcher invokes Transaction Cost Economics as a warrant to interpret a make-or-buy decision through the lens of efficient contracting. Of course, the same make-or-buy decision could just as well be approached by adopting warrants from theories of power or competence (Ketokivi & Mahoney, 2020; Santos & Eisenhardt, 2005). Adopting different warrants leads to different interpretations of the data. Indeed, a useful way to understand the role of theoretical warrants is to think of them as interpretational *lenses* through which we make sense of the grounds.

The most obvious choice situation in constructing an argument is the choice of warrants: Theoretical warrants in particular guide the researcher's attention to certain grounds, deeming others irrelevant. Research on organization boundaries offers an illustration.

Consider two researchers who both seek to understand firm scope: Which activities does the firm conduct

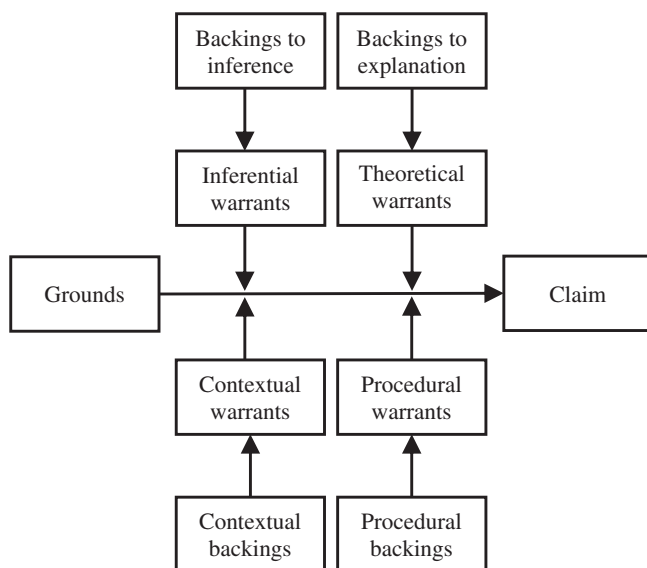


FIGURE 3 Elaborating the warrant structure

TABLE 1 Four types of warrants

Warrant	Definition	Example	Backing	Notes
Theoretical	Using theory as a sensemaking device in the interpretation of the grounds (data)	Using transaction cost economics to interpret the make-or-buy decision through the lens of efficient contracting	A theoretical body of knowledge (e.g., organization economics) and paradigmatic assumptions (e.g., efficiency)	Theories invoked must be established, they must have “money in the bank” (Meehl, 1990, p. 115)
Inferential	Using methodological rules to draw inferences from the grounds	Inferring causality from correlation in a non-experimental research design	Methodological body of knowledge (e.g., statistical theory)	Inference is not explanation (e.g., causal inference is <i>not</i> causal explanation)
Procedural	Applying an <i>established</i> set of <i>formalized</i> procedures	Inductive multiple-case research, the theory-testing method, and other formalized research designs	Methodological formalism (rule-following); unambiguous communication of results	Applying a <i>template</i> may mask idiosyncrasies in the author’s actual reasoning processes
	Applying an <i>idiosyncratic</i> set of <i>explicit</i> procedures	Gioia Methodology	Transparency; tractability	Making procedures explicit does not imply formalism
Contextual	Incorporating <i>common knowledge</i> about the context	Arguments about Southwest Airlines ascribe, without empirical evidence, a low-cost strategy	The empirical body of knowledge (e.g., industry studies, case studies of particular firms)	Not to be confused with <i>direct</i> empirical evidence, which is in the grounds of the argument

in-house and which ones does it outsource? Suppose further that one researcher adopts the power view and the other the efficiency view. The former is typically couched in theories of resource dependence (Pfeffer, 1987), and the latter, in organization economics (Williamson, 1971). While the choice of one’s theoretical foundation may be a matter of preference and policy, the implications of these choices extend to how the argument is ultimately structured and how the grounds are bridged to the claims.

What grounds are considered relevant in the two views? In the power view, the key question to be examined is who *controls* the assets, whereas the efficiency view focuses on who *owns* them. The choice of focusing on control versus ownership has fundamental implications not only for how the grounds are bridged with the claims, but also for which grounds will be considered relevant in the first place. Indeed, what counts as a relevant fact depends on the warrant. For example, the efficiency researcher will find corporate balance sheets useful and informative, but the power researcher will find them less informative, perhaps even irrelevant. Indeed, in the power view one of the central questions of interest is whether one can effectively exercise control over assets one does not own: Will a powerful final assembler be able to control what kinds of technologies its suppliers use, which customers they serve, and how they allocate their capacity? The power researcher will find the dynamics of

the exchange relationship, most notably its power asymmetry, more informative than balance sheets.

Different interpretations of the same grounds are sometimes thought of as competing explanations. However, while alternative interpretations may compete *for our attention*, they often cannot be compared in a *methodologically* sound manner: Asking whether the antitrust argument is *better* than the efficiency argument in the Schwinn case misses the point that the two arguments are based on different paradigms. More generally, the question of whether one set of unobservable terms and mechanisms explains an outcome better than another set of distinct, unobservable terms and mechanisms, is methodologically meaningless. When explanations involve theoretical concepts, “there is no best explanation” (Boylan & O’Gorman, 1995, p. 73).

4.2 | Inferential warrants: Appealing to methodology

Inferential warrants are used to justify reasoning choices that involve empirical (as opposed to theoretical) generalization. For example, a researcher would want to invoke an inferential warrant to present the empirical generalization that *X* correlates significantly with *Y*. In contrast with theoretical warrants, inferential warrants are used

in the non-explanatory parts of the argument: That X and Y correlate is not a theoretical claim but simply an inferential step that rests on an empirical generalization. Ultimately, inferential warrants complement theory by enabling the transitions from raw data to empirical generalizations.

Inferential warrants can effectively be described by examining the use of statistical models and statistical evidence. Rules about statistical significance are perhaps the most commonly used inferential warrants in statistical research: Upon observing that the p -value of an estimate is .013, the author invokes the warrant “ $p < .05$ means the estimate is statistically significant” to infer that the parameter in question differs from zero.

Inferential warrants have inferential backings. For example, the calculation of p -values is based on theory of statistical inference developed by statisticians Ronald Fisher, Jerzy Neyman, and Egon S. Pearson. Therefore, the general use of p -values has a methodological backing in probability theory and theory of estimation. But why is a probability of less than 1 out of 20 considered *statistically significant*, that is, what is the backing of the “ $p < .05$ means the estimate is statistically significant” warrant? It is obvious that while the rule operates as an inferential warrant, it receives no backing from statistical theory. This should be obvious: Labeling a probability of less than 5% *significant* simply means that we will have set the level of Type I error at 5%. This decision is made not on methodological but on practical grounds. To be sure, a Type I error rate of 5% may in some contexts have serious consequences. In medical research for instance, it translates into a *false positive*, which may be unacceptable at 5% (Colquhoun, 2019). In management research, we seldom think in terms of false positives, and most of the time, Type I error rate is set to 5% by default, without any explicit contextual consideration. This leaves the backing of the warrant unspecified.²

The problematic aspects of p -value cutoffs are well known, but it is instructive to make the connection to warrants: The problem with cutoffs as warrants is that they lack a methodological backing. This is obviously not limited to p -values, instead, it applies to all instances of simple rule following, which are abundant in statistical research (Lance & Vandenberg, 2008, 2012). Interestingly, some academic journals are taking an explicit position on the use of cutoffs: *Strategic Management Journal*, for example, “no longer accepts papers for publication that report or refer to cutoff levels of statistical significance (p -values)” (publisher's web site, accessed September 25, 2019). Rules of thumb are becoming less effective as inferential warrants.

Perhaps one of the reasons appeals to the statistical-significance warrant persist is that we have not developed

credible, alternative inferential warrants to help determine when an estimate is worth our attention. The tell-tale sign that this work seems to still lie ahead of us is the fact that as late as 2017, the American Statistical Association was compelled to hold a two-day symposium on statistical inference, with p -values as the central topic. The symposium resulted in a special issue of the *American Statistician* (Volume 73, Supplement 1) with a total of 43 articles published. We find ourselves in a strange impasse that has lingered on for decades: Everybody knows the warrant has no backing, but we keep pretending as if it did.

We do not claim to have an answer to how to get out of the impasse, but we think we can delineate the kind of an approach it would take. We would have to start with the objective of developing an explicit understanding of inferential warrants and their role in an argument that involves statistical analysis. Most importantly, an inferential warrant must be understood as a *local* principle applied in the context of the specific argument. McShane et al. (2019, p. 238, emphasis added) argued that universal inferential rules fail to incorporate “related prior evidence, plausibility of mechanisms, study design and data quality, real world costs and benefits, novelty of finding, and other factors *that vary by research domain*.” Echoing the same sentiment, Betensky (2019, p. 115) suggested that “the p -value requires context, not a threshold.” It is refreshing to see statisticians incorporate context into their arguments instead of simply trying to formulate more sophisticated, contextually detached rules. From the point of being a plausible warrant in the context of an argument, universal rules of inference, no matter how sophisticated, will eventually suffer from essentially the same shortcomings as the $p < .05$ rule (McShane et al., 2019, p. 238). We need better contextualization, not more complexity.

Understanding the local nature of inferential warrants is a good starting point for a conversation, because it forces us to address the issue locally instead of relying on statistics experts outside the local scientific community to provide a solution. But there is also another implication that can be expressed in terms of warrants: Inferential warrants are always used and evaluated in conjunction with other types of warrants, most notably, theoretical ones. Indeed, this is how we interpret McShane et al.'s (2019) call for incorporating “plausibility of mechanisms” not only as part of theoretical inquiry but also in evaluating inference.

4.3 | Procedural warrants: Being explicit about the research process

Procedural warrants seek to justify an argument by focusing on the general research design. Here, we distinguish

between two kinds: Warrants that invoke an established, formalized research design, and those that rely on explicit but idiosyncratic research designs. We discuss the two in turn.

There are several established, formalized research designs for linking empirical grounds with theoretical claims. Two widely used research designs in management research are the *theory-testing method* and Eisenhardt's (1989) *multiple-case research*. The former has its intellectual roots in the normative philosophy of science literature on the confirmation of scientific theories (Hempel & Oppenheim, 1948; Popper, 1959; Salmon, 1989; Whewell, 1840). In this literature, the methodological focus is on how empirical predictions are derived from theory and subsequently confirmed (or disconfirmed) using data. This literature addresses mainly the methodology of the natural sciences, making its specific tools inapplicable to the social sciences. Specifically, a central difference between the natural and the social sciences is that the latter do not contain the kinds of *universal laws* that are required in the application of tools such as Hempel's (1965) *deductive-nomological model* (Gorski, 2004). However, both management researchers in particular and social scientists more generally have adopted the general principle of *methodological formalism* from the methodology of the natural sciences: The theory-testing method seeks to formalize the way empirical predictions are derived from theory. Similarly, multiple-case research subscribes to formalism by emphasizing the inductive process by which claims emerge from the grounds. Both the theory-testing method and the multiple-case method endorse the principle that reasoning in scientific research should adhere to a specific form.

Formalized procedures, as warrants, receive their backing from the principle of methodological formalism. Formalism is endorsed, because it makes the procedures in a specific argument recognizable, thus leading to unambiguous communication of results. However, we must understand that this backing is subject to critique. An obvious critique is that the relationship between the warrant and its backing can be argued to be tautological: Formalized procedures are used because formalization is beneficial. Indeed, critics (e.g., Chick, 1998) point out that formalism is not so much a methodological principle as it is a *virtue*. Therefore, formalism belongs to the same category as parsimony: Some research communities tend to be partial toward parsimonious explanations simply *for the sake of parsimony*. Virtues are something the community simply takes as granted, and consequently, often grants the author the warrant to invoke them in arguments without further justification.

Some researchers seem to think that in order to make one's reasoning explicit, one must apply methodological

formalism by adopting a general research design "template." But making one's reasoning explicit does not mean one must adhere to methodological formalism, which brings us to the second variant of procedural warrants: invoking explicit and systematic yet idiosyncratic research designs. The idiosyncratic approach to procedural warrants involves justifying one's research design by elaborating and justifying the practices of data production and analysis in the context of one's research.

Gioia et al. (2013) have developed an idiosyncratic approach to the analysis of data in qualitative research. *Gioia Methodology* draws on multiple intellectual foundations such as Grounded Theory and ethnography, but is in many ways unique in its general research design. Further, in stark contrast with Eisenhardt's (1989) research design, Gioia et al. (2013, pp. 25–26) explicitly reject methodological formalism: "[O]rganizational researchers seem to be applying [Gioia Methodology] as a template, [...] as a 'formula,' essentially reproducing the exact format of the data structure from recently published studies. Even a number of methodology sections now seem to be adopting formats and procedural descriptions that are almost identical to those in the published works. This trend is something of a concern, because we envision the approach as a 'methodology,' rather than a 'method'—that is, we see it as a flexible orientation toward qualitative, inductive research that is open to innovation, rather than a 'cookbook'." This explicit rejection by the architects of Gioia Methodology merits attention precisely because some researchers mistakenly understand it to be a formalized general method. Indeed, Langley and Abdallah (2011, p. 107) explicitly discussed what they called "the Gioia method" as a template for qualitative research. This seems to confuse method and methodology: Methods are *tools* for knowledge creation, methodologies are broader epistemological *principles* applied in research.

Unlike Eisenhardt's (1989) research design, Gioia Methodology seeks systematicity and transparency *without rule following*. One of the manifestations of this is the fact that whereas Eisenhardt (1989) emphasizes inductive reasoning, Gioia et al. (2013, p. 21) acknowledge the use of creativity, innovation, and the use of abductive reasoning. The shift from induction to abduction ascribes the researcher an active role as a reasoner and interrogator of data. In stark contrast, Eisenhardt emphasized the primacy of not the researcher but the data, arguing explicitly for the benefits of formalism (Eisenhardt & Graebner, 2007, p. 25). Whereas both Eisenhardt's and Gioia's approaches pertain to theory development and invoke Grounded Theory, the crucial difference in emphasis makes the two fundamentally different approaches to research design: Eisenhardt's approach subscribes to formalism and offers a template, Gioia's does neither.

4.4 | Contextual warrants: Appealing to common knowledge

Contextual warrants are founded on commonly held empirical knowledge, and are used when researchers draw on this knowledge. Contextual warrants are commonly encountered when the researcher studies a well-known case organization. Research-based arguments regarding Southwest Airlines, Toyota, or Zara, often incorporate general knowledge of these specific firms as warrants in arguments. For example, researchers of the airline industry are seldom called to present direct empirical evidence that Southwest's competitive strategy is based on low cost (Lapr   & Scudder, 2004) or that JetBlue is a focused airline (Mellat-Parast et al., 2015). Further, binary categorizations—an airline is *either* focused *or* unfocused—are also commonly accepted without explicit justification.

While in some cases such empirical knowledge is truly general in that it can be considered common knowledge, there are also empirical research programs within scientific communities that can serve as a source of contextual warrants. A good example of an empirical research program relevant to operations management is the *Program on Vehicle and Mobility Innovation*, currently hosted by Wharton Business School, but originally founded at MIT in 1979 as the *International Motor Vehicle Program*. The participants of this academic consortium coalesce around topics related to mobility and automotive innovation. Obviously, those who formulate arguments specifically regarding the automotive industry can turn what we already know about the context into warrants in their arguments. Indeed, this is what researchers do more generally as well: They make use of the existing empirical evidence and knowledge of the context as warrants as they interpret their own evidence (the grounds).

5 | TWO EXAMPLES: PRODUCTION GEOGRAPHY AND HOME CARE

To illustrate the use of warrants in empirical research articles published in *Journal of Operations Management*, we discuss two examples from our own research: (a) the production-location decision (Ketokivi, Turkulainen, et al., 2017) and the organization of home care (Groop et al., 2017). We use examples from our own research not to highlight our own work, but precisely because the warrants in arguments are often left implicit; the authors' own reflection is required to make them explicit.

5.1 | The production location decision

Ketokivi, Turkulainen, et al. (2017) analyzed 35 decisions to locate the final stage of production (such as final assembly) in a specific geographic location. Of particular interest were situations in which the production site was located in a high-cost (high-GDP) country. The study focused on supply-chain structure and inter-functional interdependencies.

The central concepts the authors used to make sense of the location decisions were *formalization*, *specificity*, and *coupling* (Ketokivi, Turkulainen, et al., 2017, pp. 22–24). Importantly, these concepts did not emerge from the data or the empirical analysis, they were adopted *a priori* as warrants to make sense of the massive amounts of contextual, detailed data. In the following, we examine these warrants and their backings in more detail. In retrospect, the first author of the Ketokivi, Turkulainen, et al. (2017) article thinks the reader would have benefited from the following examination, because it would have (a) made the argument more transparent, and (b) more effectively explicated a crucial boundary condition. Without the explication of the warrants, the structure of the argument remains opaque and the boundary conditions implicit.

Formalization and *coupling* are warrants that are based on the Structural Contingency Theory of organizations, in particular, the work on organizational differentiation and integration (Lawrence & Lorsch, 1967). Choosing Structural Contingency Theory as the warrant directs attention to the structural features of supply chains. But to be sure, examining whether buyers and suppliers are loosely or tightly coupled or whether their relationships are based on formalization is only one out of many ways to make sense of complex supply-chain relationships. Why not examine, say, supply-chain power instead? Further, conventional supply-chain researchers might suggest taking a time-based approach and focus on lead times. These are both valid approaches to supply chains, and the only intellectually honest answer why these were not considered is because the authors *chose* the structural approach, just like the power researcher would choose the power approach and the researcher interested in lead times the time-based approach. In research on operations management, the importance of structural aspects has been argued in the theoretical and empirical literature on *organizational integration* in particular (Swink et al., 2007).

The central point for the purposes of this article is that while the structural features were ultimately examined empirically (Ketokivi, Turkulainen, et al., 2017, p. 25), they did not emerge from the data; instead, they constitute the theoretical warrants used to bridge the

grounds to the claims. Equally important is to understand that there is no methodological way to argue that the structural approach is superior to other approaches; adoption of warrants is a matter of choice.

Specificity operates in the argument in a similar way; it provides an *a priori* concept. However, specificity is different from the other two, because it is based on another theoretical warrant, namely Transaction Cost Economics. Importantly, invoking Transaction Cost Economics introduces an essential boundary condition: It introduces the assumption that minimizing the cost of transacting is essential. This is a critical assumption, because it implicitly demotes the revenue aspect of supply chains to a concern of secondary importance. In retrospect, the first author of the article has come to the conclusion that this point in particular should receive explicit attention in the article: The key arguments presented in the article are essentially arguments for *cost efficiency*, and must not be confused with arguments for power or competence (Ketokivi & Mahoney, 2020). Again, researchers interested in the power aspects of supply chains might find the choice of focusing on efficiency arbitrary, and they would be correct: It is not hyperbole to describe the choice of the warrant as arbitrary, in the sense that there are always multiple options from which to choose. Of course, *arbitrary* does not mean that *anything goes*; researchers must obviously always justify their choices to their audience. Even though those reading the published version of the article cannot see it, the authors did indeed have to justify their choice as part of the peer-review process.

5.2 | Organizing the delivery of home care

Using an *intervention-based research design*, Groop et al. (2017) sought to improve the productivity of a home care delivery system in the greater Helsinki region in Finland. Here, home care delivery refers to the provision of medical and social services to promote assisted living to senior citizens who are still able to live at home if they receive assistance with certain activities, such as taking their daily medication or bathing. These services are provided by a team of *traveling caregivers*, that is, nurses and social workers who each visit the homes of 5–10 customers during a work shift. The central operational question is the following: How should caregiver travel routes be planned, and how should overall caregiver capacity be planned?

In this case, the choice of warrants has a more practical objective: How should the practical problem be *framed* (Simon, 1997) in order to design a solution that improves system productivity? One option would be to frame the

problem as the conventional *traveling-salesperson problem*, and accordingly, seek to minimize caregiver travel time. Following this premise, one would “assign caregiver visits to all nearby customers consecutively” (Groop et al., 2017, p. 14). Another option would be to seek efficient use of overall caregiver capacity throughout the work shift. This would, among other things, involve allowing caregivers to perform only time-critical visits (e.g., administering Mrs. Johnson's insulin shots) during peak hours, and leveling demand throughout the shift by allocating non-time-critical visits (e.g., giving Mr. Williams a bath) to non-peak hours. These two options—minimize travel time or level capacity—can be thought of as alternative warrants to be applied to the design of the system. Here, one would adopt the warrant that leads to a comparatively more efficient solution.

Unlike in the production location example where the objective was to *explain* why production was located in a specific geographic location, the objective in the home care case was to *help design* a system that delivers efficient care. Here, the two warrants are best thought of not as theoretical but as inferential, because they provide inferential rules for reasoning: The choice of, say, minimizing travel time provides an unambiguous inferential rule for determining how the system should be designed. The authors found that in the specific case of home care and the objective of increasing overall system productivity, the inferential warrant of leveling caregiver capacity offered a comparatively more effective solution. Of course, in other contexts or in the case of other outcome variables, the traveling-salesperson formulation might be superior.

Note that the two warrants in the home care case are mutually exclusive, and therefore, cannot be reconciled. Specifically, under the traveling salesperson formulation, a caregiver would consecutively visit both Mrs. Johnson and Mr. Williams if the two happened to live in the same building. Under the demand-leveling formulation, the caregiver would be *prohibited* from visiting Mr. Williams (a non-time-critical visit) in conjunction with the 8 a.m. peak-time visit to Mrs. Johnson's home, even if the two lived in the same building.

Groop et al. (2017) found that given the objective of productivity improvement in the context of home care, the capacity-leveling warrant seemed to provide a comparatively better solution. This led the researchers to a number of conclusions that took the form of *design propositions* (Denyer et al., 2008). One of the design propositions was to “[m]aintain excess capacity in a common resource pool consisting of skill group buffers” (Groop et al., 2017, p. 14). Using the terminology of the Toulmin framework, this design proposition is a claim that rests on two pillars: (a) data (grounds) that show that applying

this design proposition improves productivity; and (b) the warrant that the system should seek to level demand.

In this example, being mindful of the warrant is important, because it explicates the principle applied in seeking a solution to the problem. Note that the principle only considers the nature of the demand and ignores issues such as customer satisfaction and employee well-being, which are obviously relevant system objectives as well. Indeed, the researchers found that different stakeholders had different expectations and preferences. Specifically, the only thing Mrs. Johnson expects is that a registered nurse show up at her door at exactly 8:00 a.m. to administer her insulin shot; she does not care about leveling system demand. Mr. Williams, in turn, does not necessarily need to take his bath exactly at the same hour every day, but we venture to guess that he prefers a daily visit from the same, male caregiver who knows how Mr. Williams prefers to take his bath. Neither of these preferences is explicitly incorporated into the demand-leveling warrant, in fact, Mr. Williams's preference in particular might be in direct conflict with it: His preference would introduce a system constraint—"Mr. Williams must be visited by the same, male caregiver"—that makes leveling demand more difficult.

In retrospect, the second author of the Groop et al. (2017) article thinks that even though the warrants were actually made explicit (Groop et al., 2017, p. 14), they could have been discussed in more detail, and their contextual essence could have been highlighted more. The four design propositions presented in the article (Groop et al., 2017, p. 14) are fundamentally context dependent and can be prescribed only in light of the demand-leveling warrant. In fact, the first design proposition is the warrant: "Level demand by off-loading non-time-critical visits to off-peak hours" (Groop et al., 2017, p. 14).

6 | REFLECTIONS AND IMPLICATIONS

What can we gain from an explicit examination of argument structure in general and the typology of warrants in particular? What should scholars do differently? What should we teach to our doctoral students?

In offering prescription, we must tread carefully so as not to replace one orthodoxy with another. If reasoning is a local, context-dependent activity, it would be inconsistent to offer general prescription. Indeed, the only general prescription we feel comfortable offering is this: Researchers should always reason themselves instead of appealing to someone else's reasoning. Or more accurately, if one chooses to rely on someone else's reasoning (e.g., by adopting a formalized research design), one must

understand what the central warrants are and what kinds of backings support them.

The typology of warrants we offer is not a checklist, nor is it a set of formalized, readily actionable guidelines; it merely provides a possible starting point for scholars to systematically examine the justification of their arguments. Without further details of the specific research situation and the knowledge interest, we have nothing else to prescribe. Simply put, there is nothing in a scholar's reasoning that is either self-validating, or, that can be validated by an appeal to someone else's reasoning.

Suspension of explicit, transparent reasoning ultimately invites scholarship based on dogma, convention, and mimicry. How could such scholarship ever produce knowledge that is relevant to anyone except the academic community? How is it possible that dogmatic approaches designed by academic could result in managerially relevant knowledge claims? We propose that transparency in reasoning can provide a foundation for methodological rigor that ultimately feeds directly into practical relevance as well. Indeed, in many practically oriented fields such as law, the explication of warrants is an integral part of the argument; the Schwinn case is a good example. How could we even begin to understand the essence of the Schwinn debate if we did not explicate the competing warrants and their backings?

In addition to the general prescription of encouraging explicit reasoning, we offer three more detailed guidelines: (a) Evaluate the argument, not the claim; (b) Mind the qualifier; and (c) Incorporate an examination of backings into researcher training. To supplement these guidelines, we have also prepared a checklist of questions that reviewers and editors can use in evaluating manuscripts. The checklist can be found in the Appendix to this article.

6.1 | Evaluate the argument, not the claim

In his experimental studies on the psychology of science, Mahoney (1977, p. 161) found that "[i]n addition to showing poor interrater agreement, reviewers were strongly biased against manuscripts which reported results contradictory to their theoretical perspective." We suggest that the failure to explicate warrants offers a plausible interpretation: Absent the explication of warrants, instead of evaluating whether the claim is supported by the grounds, reviewers simply examine whether they find the claim itself reasonable.

Conservation of belief is a basic human condition to which no scientist is immune (Nickerson, 1998, p. 176). All of us must pay attention to two central mechanisms

of confirmation bias that are likely present in the evaluation of arguments:

1. Conservatism: We may think that scientific discoveries—claims that really challenge our predispositions—tend to be resisted mainly by non-scientific actors and institutions, but in reality they are in fact most strongly resisted by *scientists themselves* (Nickerson, 1998, p. 194).
2. Theory persistence: Scientists tend to develop a vested interest in the specific theories they develop (Nickerson, 1998, p. 195).

Here, it is important to distinguish confirmation bias from the more familiar notion of publication bias (Kepes et al., 2012). Publication bias stems from selective reporting: Studies that report a significant relationship between *X* and *Y* are more likely to be published than those reporting no significant relationship. This may lead to a biased conclusion about the strength of statistical associations. Specifically, if nine out of 10 studies show no relationship between *X* and *Y* and the tenth study shows a positive relationship, only the one showing a positive relationship will be published, and we end up with a positively biased understanding of the relationship between *X* and *Y*. Accordingly, analyses of confirmation bias often focus on inferential (statistical) issues, such as suppression of findings due to small magnitudes and statistical insignificance (Harrison et al., 2017).

In contrast, confirmation bias is not about biased inference, it is a form of *prejudice* that stems from one's paradigmatic predisposition. Specifically, if the presenter of an argument leaves the warrants unspecified, those evaluating the argument have no option but to “fill in the blanks” and introduce the warrants themselves. Different evaluators may invoke different warrants, and consequently, ascribe different meanings to the findings. If the author's claim coincides with this meaning, the claim is accepted, but if it does not, the claim is rejected, or at least, called into question. To describe this as bias is in our view misguided; it is descriptively more accurate to describe it as a form of prejudice.

Eliminating confirmation bias altogether is wishful thinking, but we posit that acknowledging the evaluator's task can be useful. To this end, Henle (1962) noted that this task is *logical*, meaning that the evaluator must examine the logical structure of the argument in its entirety. This is not a trivial point, because as Henle (1962, p. 370) noted, we often evaluate merely “the content of the conclusion, not the logical form of the argument.”

Both authors of this article have on several occasions been in a situation where a reviewer in a journal peer-

review process has, in one form or another, expressed the sentiment that he or she “does not buy our argument.” However, a closer inspection reveals that the reviewer does not have a problem with the argument, but rather, primarily with the claim that is made. In his or her critique, the reviewer does not demonstrate the structure of the argument to be invalid or implausible, only that the conclusion is at odds with his or her understanding.

To be sure, we do not blame the reviewer for making a mistake, being unreasonable, or rejecting a plausible argument: We fully acknowledge that the problem may well have stemmed simply from the fact that we, as authors, failed to make our warrants explicit. Indeed, only over the years and with accumulating experience have we realized the importance of making warrants explicit. Our first guideline, “evaluate the argument, not the claim,” is therefore aimed at both the evaluator *and* the author of the argument.

Understanding the structure of the argument in its entirety also unveils something essential about the nature of the claim: The conclusions of our research activities are not *truth claims* but *knowledge claims*. Again, this is not a semantic point: A truth claim is a statement about the world (an ontological notion), whereas a knowledge claim is a statement of what we assert to know (an epistemological notion). In our arguments, we seek to convince our audiences that our assertions are plausible. Whether our claims are true is never an actionable criterion in the evaluation of an argument. The reason is obvious: Neither the author nor the evaluator have access to what is true. The only part of the argument where truth is even remotely relevant is when we examine whether we have the facts correct and whether the data are valid; but these are questions about the grounds, not the claim. Finally, understanding the distinction between a truth claim and a knowledge claim also helps us understand the crucial role of qualifiers, which leads us to our second guideline.

6.2 | Mind the qualifier

As the Schwinn example showed, the central point of disagreement was specifically on how the claim should be qualified. Further, Williamson's rebuttal—like rebuttals more generally (Toulmin et al., 1979, chapter 10)—was aimed not at the claim but its unconditional qualifier. That rebuttals are not aimed at the claim but its qualifier has important implications.

The use of qualifiers such as *likely*, *probably*, and *plausibly* is often casual—we must make it more rigorous. In the context of scientific research, we find it useful

to link qualifiers to the use of scientific reasoning, because they can provide insight on what kinds of qualifiers we should introduce to a claim. In the following, we discuss deductive, inductive, and abductive reasoning in turn.

Deductive reasoning is logically coherent in the sense that the premises unambiguously imply the conclusion. Accordingly, the use of unconditional qualifiers such as *certainly* and *evidently* should be used only when the claims can be shown to follow from the grounds in a deductive way. Here, we must also keep in mind that even though the conclusion follows from the premises in an unambiguous manner, its credibility hinges on the credibility of the premises from which it is deductively derived. This is important, because as Mill ([1843] 1882, p. 981) noted, it is always possible to make one's reasoning deductively valid by introducing *ad hoc* premises. This is a logical fallacy that is sometimes labeled an *ad hoc rescue*.

Inductive reasoning is probabilistic in the sense that the claim involves a generalization of the grounds. If indeed the claim is an empirical (as opposed to theoretical) generalization, one can use qualifiers such as *likely* or *probably*. For example, we could make the empirical prediction that increasing the frequency of transacting makes the in-house production of a component *more likely*.

Abductive reasoning is neither logically coherent nor probabilistic, because it involves inference not to an empirical generalization but to *an explanation* (Lipton, 2004). Abductive reasoning leads to explanations that are *plausible*. At the same time, underdetermination of theories implies that there are always other plausible explanations in addition to the one offered. In the Schwinn case, both the Supreme Court's claim that Schwinn broke the law and Williamson's claim that Schwinn sought economic efficiency are to be treated as alternative explanations, and therefore, as results of an abductive reasoning process. There are no grounds (i.e., empirical data) available to us that would lead to either of the two conclusions being *evidently* correct. In the Schwinn case, only those who made the decisions and know whether it was for the sake of monopoly power or transacting efficiency know the answer, and all these executives have long since passed away.

With plausible claims, it would be similarly incorrect to ascribe probabilities: It is misguided to ask whether one explanation or interpretation is *more likely* than the other. We may of course casually think of probabilities, but there is no data or analysis that produces a probability estimate for either case. When we say an explanation is plausible, we are merely suggesting that we have arrived at an interpretation that is reasonable given the grounds and the warrants. Whether the interpretation is

logically justified (in the deductive sense) or likely to be true (in the inductive sense) is not relevant, because the conclusion is a product of not merely the grounds but also the warrants and their backings. The plausibility of a claim is always contingent on the warrant.

6.3 | Incorporate an examination of backings into researcher training

We hope to have demonstrated that researchers must understand not only the warrants they use but also the foundation on which they ultimately stand—their backings. This has implications for doctoral education in particular. Specifically, rather than training generations of rule followers, doctoral programs should aim at equipping future researchers with a sense of autonomy and responsibility. This calls for developing an understanding of various methodological traditions instead of merely learning to apply them. More generally, rather than teaching compliance, focus should be on the use of discretion and judgment.

When formalized research designs are taught in doctoral seminars, instructors should ensure that students understand both that formalization is effectively invoked as a warrant, and that the backing of the formalization warrant is becoming increasingly contested in the management research community (Köhler et al., 2019). A case in point, in covering the Gioia Methodology, instructors should clarify what Gioia and colleagues meant when they proposed that their approach constitutes a methodology, not a method. Equally important is to understand why they strongly resisted the idea that their approach constitutes a template for qualitative empirical research (Gioia et al., 2013). As we have witnessed in several symposia and professional development workshops at the Academy of Management over the past 5 years, the management research community is clearly becoming increasingly critical of the idea of thinking of methodology in terms of “templates.”

The same applies to inferential warrants often used in statistical research. The ubiquitous use of demonstrably pseudo-methodological rules of thumb—VIF < 10, Cronbach's alpha > .70, et cetera—is a good example. If those teaching statistical methods to doctoral students incorporated an explicit evaluation of the backings of inferential warrants into their seminars, we predict that the uninformed use of these rules of thumb would rapidly decline. Further, this would be easy to achieve: All we would have to do is look at what the authors of various cutoff criteria actually said, and we would realize that these criteria are simply “urban legends” (Lance et al., 2006, p. 202). For example, once we acknowledge

that what Nunnally (Nunnally & Bernstein, 1994, p. 249) actually wrote about acceptable levels of reliability is that they depend on how the measure is used, we can explore the question of how sensitive parameter estimates are to increasing measurement error in the context of the specific model and dataset. This effectively transforms the classroom discussion of measurement reliability from a yes/no -issue into a discussion of degrees and context. We predict that if students, early in their doctoral education, developed an understanding of how important it is to examine how increasing measurement unreliability attenuates model estimates, they would be less likely to invoke arbitrary cutoffs as warrants in their arguments; they would know the cutoff warrant does not have a methodological backing.

6.4 | In conclusion

A claim made based on empirical analysis rests on two pillars: evidence and warrants. The former is self evident and requires no elaboration other than specifying its precise role in an argument: Evidence is found at the grounds of an argument. But in order to use evidence as the basis of a claim, the argument must invoke a warrant, or warrants. Both in scientific inquiry and arguments in other contexts (e.g., politics, law, or management), warrants are often implicit in the argument. Further, one sometimes gets the impression that some researchers think that warrants are not needed, that “the facts speak for themselves.” Even a casual glance at claims made in scientific texts reveals that this is misguided: Facts do not speak, the researcher does. Further, an explicit analysis of the structure of an argument uncovers a complex set of warrants, which in scientific inquiry can be theoretical, inferential, procedural, and contextual. Only by understanding the use of warrants can we rigorously examine an argument and the credibility and the boundary conditions of the claims that are made.

ACKNOWLEDGMENTS

We thank Editors Suzanne de Treville and Tyson Browning, as well as Department Editor Mikko Rönkkö and his review team, for detailed, constructive critique throughout the review process. We thank Dr. Rönkkö in particular for pushing us to make the message maximally relevant to research practice. We also thank collectively all those who have given us feedback on earlier versions of the manuscript. In particular, we are indebted to the participants of the JOM Mini-Conference on “What Warrants our Claims?” organized by JOM Editors on January 30, 2021. We also thank the Executive Ph.D. students at the Center for Values-Driven Leadership at Benedictine

University, who gave valuable feedback to us from the point of view of doctoral students working on their dissertations. Finally, this manuscript has benefited from scores of conversations with colleagues too numerous to list. These discussions have been essential, because argument evaluation is ultimately all about research practice, which means it must be informed by those who practice research. All errors and omissions are our sole responsibility.

CONFLICT OF INTEREST

The authors declare no potential conflict of interest.

ORCID

Mikko Ketokivi  <https://orcid.org/0000-0003-4510-4949>

ENDNOTES

- ¹ Here, we adopt the definitions by Bagozzi and Phillips (1982, p. 465). *Theoretical concepts* are “abstract, unobservable properties or attributes of a social unit or entity [that] achieve their meaning through formal connections to empirical concepts as well as through their definition.” *Empirical concepts* are “properties or relations whose presence or absence in a given case can be inter-subjectively ascertained, under suitable circumstances, by direct observation.”
- ² Kennedy-Shaffer (2019) provides an enlightening historical interpretation of the $p < .05$ rule. In the time before computers, p -values had to be calculated by hand. But instead of hand-calculating the p -value in every instance (which was time-consuming), researchers used the chi-square tables that Pearson and Fisher had provided to expedite the process. These tables included a number of arbitrarily chosen cutoffs based on the number of standard deviations: “The value for which $p = .05$, or 1 in 20, is 1.96 or nearly 2; it is convenient to take this point as a limit in judging whether a deviation is to be considered significant or not. Deviations exceeding twice the standard deviation are thus formally regarded as significant” (Fisher, 1925, p. 47, italics added). There is obviously no methodological backing for such convenience.

REFERENCES

- Bagozzi, R. P., & Phillips, L. W. (1982). Representing and testing organizational theories: A holistic construal. *Administrative Science Quarterly*, 27(3), 459–490.
- Bain, J. S. (1968). *Industrial organization* (2nd ed.). New York: Wiley.
- Bedeian, A. G. (2003). The manuscript review process: The proper roles of authors, referees, and editors. *Journal of Management Inquiry*, 12, 331–338.
- Betensky, R. A. (2019). The p -value requires context, not a threshold. *American Statistician*, 73(sup1), 115–117.
- Bird, A. (1998). *Philosophy of science*. London: Routledge.
- Bork, R. H. (1965). The rule of reason and the per se concept: Price fixing and market division. *Yale Law Journal*, 74(5), 775–847.
- Boylan, T. A., & O’Gorman, P. F. (1995). *Beyond rhetoric and realism in economics: Towards a reformulation of economic methodology*. London: Routledge.

- Bronfenbrenner, M. (1944). Production functions: Cobb-Douglas, interfirm, intrafirm. *Econometrica*, 12(1), 35–44.
- Cachon, G. P. (2003). Supply chain coordination with contracts. In S. Graves & T. de Kok (Eds.), *Handbooks in operations research and management science, volume 11: Supply chain management* (pp. 227–339). North-Holland: Elsevier.
- Chick, V. (1998). On knowing one's place: The role of formalism in economics. *Economic Journal*, 108(451), 1859–1869.
- Coase, R. H. (1988). *The firm, the market, and the law*. Chicago: The University of Chicago Press.
- Colquhoun, D. (2019). The false positive risk: A proposal concerning what to do about p-values. *American Statistician*, 73(sup1), 192–201.
- Denyer, D., Tranfield, D., & van Aken, J. E. (2008). Developing design propositions through research synthesis. *Organization Studies*, 29(3), 393–413.
- Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review*, 14(4), 532–550.
- Eisenhardt, K. M., & Graebner, M. E. (2007). Theory building from cases: Opportunities and challenges. *Academy of Management Journal*, 50(1), 25–32.
- Feyerabend, P. (1978). *Science in a free society*. London: NLB.
- Fisher, M. L. (1997). What is the right supply chain for your product? *Harvard Business Review*, 75(2), 105–116.
- Fisher, R. A. (1925). *Statistical methods for research workers*. Edinburgh: Oliver and Boyd.
- Gehman, J., Glaser, V. L., Eisenhardt, K. M., Gioia, D., Langley, A., & Corley, K. G. (2017). Finding theory-method fit: A comparison of three qualitative approaches to theory building. *Journal of Management Inquiry*, 27(3), 1–18.
- Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. *Organizational Research Methods*, 16(1), 15–31.
- Gorski, P. S. (2004). The poverty of deductivism: A constructive realist model of sociological explanation. *Sociological Methodology*, 34(1), 1–33.
- Groop, J., Ketokivi, M., Gupta, M., & Holmström, J. (2017). Improving home care: Knowledge creation through engagement and design. *Journal of Operations Management*, 53–56, 9–22.
- Grover, V., & Malhotra, M. K. (2003). Transaction cost framework in operations and supply chain management research: Theory and measurement. *Journal of Operations Management*, 21(4), 457–473.
- Harman, G. H. (1965). The inference to the best explanation. *Philosophical Review*, 74(1), 88–95.
- Harrison, J. S., Banks, G. C., Pollack, J. M., O'Boyle, E. H., & Short, J. (2017). Publication bias in strategic management research. *Journal of Management*, 43(2), 400–425.
- Hempel, C. G. (1965). *Aspects of scientific explanation and other essays in the philosophy of science*. New York: Free Press.
- Hempel, C. G., & Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science*, 15(2), 135–175.
- Henle, M. (1962). On the relation between logic and thinking. *Psychological Review*, 69(4), 366–378.
- Kennedy-Shaffer, L. (2019). Before $p < 0.05$ to beyond $p < 0.05$: Using history to contextualize p-values and significance testing. *The American Statistician*, 73(sup1), 82–90.
- Kepes, S., Banks, G. C., McDaniel, M., & Whetzel, D. L. (2012). Publication bias in the organizational sciences. *Organizational Research Methods*, 15(4), 624–662.
- Ketokivi, M., & Jokinen, M. (2006). Strategy, uncertainty and the focused factory in international process manufacturing. *Journal of Operations Management*, 24(3), 250–270.
- Ketokivi, M., & Mahoney, J. T. (2020). Transaction cost economics as a theory of supply chain efficiency. *Production and Operations Management*, 29(4), 1011–1031.
- Ketokivi, M., & Mantere, S. (2010). Two strategies for inductive reasoning in organizational research. *Academy of Management Review*, 35(2), 315–333.
- Ketokivi, M., Mantere, S., & Cornelissen, J. P. (2017). Reasoning by analogy and the progress of theory. *Academy of Management Review*, 42(4), 637–658.
- Ketokivi, M., Turkulainen, V., Seppälä, T., Rouvinen, P., & Ali-Yrkkö, J. (2017). Why locate manufacturing in a high-cost country? A case study of 35 production location decisions. *Journal of Operations Management*, 49–51, 20–30.
- Köhler, T., Smith, A., & Bhakoo, V. (2019). Feature topic for ORM: “Templates in qualitative research methods”. *Organizational Research Methods*, 22(1), 3–5.
- Kouvelis, P., Chambers, C., & Wang, H. (2009). Supply chain management research and production and operations management: Review, trends, and opportunities. *Production and Operations Management*, 15(3), 449–469.
- Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago, IL: The University of Chicago Press.
- Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 91–196). Cambridge: Cambridge University Press.
- Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria: What did they really say? *Organizational Research Methods*, 9(2), 202–220.
- Lance, C. E., & Vandenberg, R. J. (Eds.). (2008). *Statistical and methodological myths and urban legends: Received doctrine, verity, and fable in the organizational and social sciences*. New York: Routledge.
- Lance, C. E., & Vandenberg, R. J. (Eds.). (2012). *More statistical and methodological myths and urban legends*. New York: Routledge.
- Langley, A., & Abdallah, C. (2011). Templates and turns in qualitative studies of strategy and management. *Research Methodology in Strategy and Management*, 6, 105–140.
- Lapr , M. A., & Scudder, G. D. (2004). Performance improvement paths in the U.S. airline industry: Linking trade-offs to asset frontiers. *Production and Operations Management*, 13(2), 123–134.
- Laudan, L. (1990). Demystifying underdetermination. In C. W. Savage (Ed.), *Scientific theories (Minnesota studies in the philosophy of science)* (Vol. 14, pp. 449–472). Minneapolis, MN: University of Minnesota Press.
- Lawrence, P. R., & Lorsch, J. W. (1967). *Organization and environment: Managing differentiation and integration*. Boston: Harvard University Press.
- Lipton, P. (2004). *Inference to the best explanation*. London: Routledge.
- Locke, K., & Golden-Biddle, K. (1997). Constructing opportunities for contribution: Structuring intertextual coherence and “problematizing” in organizational studies. *Academy of Management Journal*, 40(5), 1023–1062.
- Longino, H. E. (1990). *Science as social knowledge: Values and objectivity in scientific inquiry*. Princeton, NJ: Princeton University Press.

- Lycan, W. G. (1998). Theoretical (epistemic) value. In E. Craig (Ed.), *Routledge encyclopedia of philosophy* (pp. 340–343). London: Routledge.
- Mahoney, M. J. (1977). Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy & Research*, 1, 161–175.
- Mantere, S., & Ketokivi, M. (2013). Reasoning in organization science. *Academy of Management Review*, 38(1), 70–89.
- McShane, B. B., Gal, D., Gelman, A., Robert, C., & Tackett, J. L. (2019). Abandon statistical significance. *American Statistician*, 73(sup1), 235–245.
- Meehl, P. E. (1990). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry*, 1(2), 108–141.
- Mellat-Parast, M., Golmohammadi, D., McFadden, K. L., & Miller, J. W. (2015). Linking business strategy to service failures and financial performance: Empirical evidence from the U.S. domestic airline industry. *Journal of Operations Management*, 38, 14–24.
- Mill, J. S. ([1843] 1882). *A system of logic, ratiocinative and inductive, being a connected view of the principles of evidence, and the methods of scientific investigation* (8th ed.). New York: Harper & Brothers.
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175–220.
- Niiniluoto, I. (1999). Defending abduction. *Philosophy of Science*, 66 (Proceedings), S436–S451.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- Peirce, C. S. (1955). The scientific attitude and fallibilism. In J. Buchler (Ed.), *Philosophical writings of Peirce* (pp. 42–59). New York: Dover.
- Pesch, M. J., & Schroeder, R. G. (1996). Measuring factory focus: An empirical study. *Production and Operations Management*, 5(3), 234–254.
- Pfeffer, J. (1987). A resource dependence perspective on intercorporate relations. In M. Mizuchi & M. Schwartz (Eds.), *Intercorporate relations: The structural analysis of business* (pp. 25–55). New York: Cambridge University Press.
- Popper, K. R. (1959). *The logic of scientific discovery*. London: Hutchinson & Co.
- Porter, M. E. (1981). The contributions of industrial organization to strategic management. *Academy of Management Review*, 6(4), 609–620.
- Quine, W. V. O. (1951). Main trends in recent philosophy: Two dogmas of empiricism. *Philosophical Review*, 60(1), 20–43.
- Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: MIT Press.
- Ragins, B. R. (2015). Editor's comments: Developing our authors. *Academy of Management Review*, 40(1), 1–8.
- Salmon, W. C. (1989). Four decades of scientific explanation. In P. Kitcher & W. C. Salmon (Eds.), *Scientific explanation (Minnesota studies in the philosophy of science)* (pp. 3–219). Minneapolis, MN: University of Minnesota Press.
- Santos, F. M., & Eisenhardt, K. M. (2005). Organizational boundaries and theories of organization. *Organization Science*, 16(5), 491–508.
- Simon, H. A. (1997). *Administrative behavior* (4th ed.). New York: Macmillan.
- Skinner, W. (1969). Manufacturing - missing link in corporate strategy. *Harvard Business Review*, 50(3), 136–145.
- Skinner, W. (1974). The focused factory. *Harvard Business Review*, 52(3), 113–121.
- Skinner, W. (1996). Manufacturing strategy on the “S” curve. *Production and Operations Management*, 5(1), 3–14.
- Stanford, K. (2017). Underdetermination of scientific theory, In Zalta, E.N. (Ed.), *The Stanford encyclopedia of philosophy*. Stanford, CA: Center for the Study of Language and Information. Available online at <https://plato.stanford.edu/entries/scientific-underdetermination/>.
- Starbuck, W. H. (2003). Turning lemons into lemonade: Where is the value in peer reviews? *Journal of Management Inquiry*, 12(4), 344–351.
- Swink, M., Narasimhan, R., & Wang, C. (2007). Managing beyond the factory walls: Effects of four types of strategic integration on manufacturing plant performance. *Journal of Operations Management*, 25(1), 148–164.
- Toulmin, S. E. ([1958] 2003). *The uses of argument* (Updated ed.). Cambridge: Cambridge University Press.
- Toulmin, S. E., Rieke, R., & Janik, A. (1979). *An introduction to reasoning*. New York: Macmillan.
- Vokurka, R. J., & Davis, R. A. (2000). Focused factories: Empirical study of structural and performance differences. *Production and Inventory Management Journal*, 41(1), 44–55.
- Whewell, W. (1840). *The philosophy of the inductive sciences, founded upon their history*. London: John W. Parker and Son.
- Williamson, O. E. (1968). Economies as an antitrust defense: The welfare tradeoffs. *American Economic Review*, 58(1), 18–36.
- Williamson, O. E. (1971). The vertical integration of production: Market failure considerations. *American Economic Review*, 61(2), 112–123.
- Williamson, O. E. (1985). *The economic institutions of capitalism*. New York: Free Press.
- Williamson, O. E. (2002). The merger guidelines of the U.S. Department of Justice: In perspective. Talk given at the 20th anniversary of the adoption of the 1982 Merger Guidelines, June 20, 2002, The Antitrust Division of the U.S. Department of Justice. Full text of the talk is available at <http://www.justice.gov/atr/hmerger/11257.htm>.

How to cite this article: Ketokivi M, Mantere S. What warrants our claims? A methodological evaluation of argument structure. *J Oper Manag*. 2021;67:755–776. <https://doi.org/10.1002/joom.1137>

APPENDIX: A CHECKLIST FOR EDITORS AND REVIEWERS

In this appendix, we offer four sets of questions to guide the peer-review process. We believe that if reviewers start paying closer attention to argument structure, it will ultimately encourage authors to follow suit and pay attention to the issues as well. We further pose these questions first and foremost to peer-reviewers (including editors and associate editors who make recommendations and decisions to publish manuscripts), because we believe that promoting the explication of argument structure requires systematic, collective attention. We also suggest that journals should explicitly start considering *argumentation rigor* as a key component of *methodological rigor* in their editorial policies. However, even though we think everything must start with the gate-keepers of the academic community, we encourage authors to address these questions as well, whether journal editors and reviewers explicitly require it or not. In conjunction with each set of questions, we offer a brief reflection why the questions are important.

What are the central claims made? Have the authors made the claims explicit, or do I have to “assemble” them from several sections of the manuscript?

During our careers as reviewers spanning over 20 years, we have found that problems often start at our inability to identify the key claims the author is making in a manuscript (we openly acknowledge that some of our own published articles may have suffered from this very problem). Instead, we have to infer them by simultaneously considering different parts of the manuscript: Some parts of the claim may appear in the introduction, others in the discussion or conclusion sections. Further, sometimes the claims authors make are confounded with research motivation; these two must be kept separate. Our recommendation is that to the extent possible, the key claims be explicitly incorporated into the introduction section of the manuscript (Locke & Golden-Biddle, 1997).

What have the authors done to ensure that I am not only able to identify what the key claims are, but that I also *understand* them the way the authors intended?

Even when something is explicitly stated, it does not follow that it will be understood the way the author intended. Authors must ensure that the key concepts of

the claim are well defined and the logic of the claim is transparent. Here, we encourage paying special attention to qualifiers, which should never be used casually (see Section 6.2). We also discourage the use of general terms such as *strategy*, *structure*, *culture*, *performance*, or *institution* in claims, because these terms are used in different meanings in different contexts. Most importantly, one should use concepts in the same meaning they are used in the theoretical conversation or research program in which the manuscript participates.

What have the authors done to *justify* their claim, that is, what is the structure of the argument? Do I understand what the central grounds and warrants are?

It is crucial to understand that having addressed the first two sets of questions, one has *not even begun* to address argument structure. In addressing this third question, the reviewer must set aside all considerations of whether the claim is understandable, novel, or appealing (points covered in addressing the first three questions), and focus squarely on its justification. This calls for an understanding of the grounds and the warrants on which the claim stands.

Should I grant the authors the warrants they seek to invoke? If I feel compelled not to grant a specific warrant, how do I justify the rejection?

Some warrants are granted, others are not. If in particular a warrant is rejected, the reviewer must be able to explicate why this is the case. Here, we want to distinguish between challenging the warrant based on lack of coherent backing versus rejecting it based on paradigmatic preference.

Some warrants simply have no credible backing; various statistical rules of thumb are a good example. If the warrant has no credible backing, or if the warrant is not coherent with its backing, the reviewer is justified in rejecting the warrant on exclusively methodological grounds. Assuming the reviewer is correct, this rejection is not contestable. Symmetrically, the author is justified in rejecting the reviewer's warrant if it does not have a credible backing. Unfortunately, both editors and reviewers sometimes assume that reviewers are more competent than authors (Starbuck, 2003, p. 345).

Rejection of a warrant becomes more complicated when the author has chosen to invoke a warrant that has a credible backing, but when there are simultaneously other warrants that could have been chosen. A reviewer who adopts, say, a power perspective to supply chains

may reject the organization economist's cost minimization as the warrant (see Section 5.1). This rejection may be subject to debate, because it can be read simply as the promotion of the reviewer's paradigmatic preference over the author's.

Importantly, here the situation is *not* symmetric: In the case of conflicting paradigmatic preferences, we argue that we should privilege the author's choice over the reviewer's. A reviewer is in a position of power over the authors, and in our field also enjoys the benefit of anonymity; giving the weaker party the benefit of the doubt in any power dynamic would seem like a prudent governance principle. The collective opinion in the academic community bears this out: In his survey of authors of academic management journals, Bedeian (2003, p. 334) found that only 6% of authors (who obviously also

serve as peer reviewers) agreed with the statement "If there is a disagreement between author and referee about a matter of opinion, the author should be required to conform to the referee's position." In a similar vein, in discussing the principle of developmental review, Ragins (2015, p. 2) noted that developmental reviews do not involve "telling the authors what to do, or taking over the authors' voice," instead, a developmental reviewer should "empower authors rather than write the paper for them." It is unfortunately the case that editors and reviewers sometimes think that they are not only more competent than authors, but also that reviewers' and editors' *opinions* have more validity than the authors' (Starbuck, 2003, p. 345). We wonder if this is the position these same editors and reviewers would take when they author articles themselves.