Journal of the Association for Information Systems
JAIS

Research Article

Evaluating and Developing Theories in the Information Systems Discipline

Ron Weber
Monash University
Ron.Weber@monash.edu

Abstract

This paper articulates a framework and criteria that can be used to evaluate the quality of theories. While the framework and criteria have general applicability, my focus is the evaluation of theories within the information systems discipline. To illustrate the usefulness of the framework and criteria, I show how they can be employed to pinpoint the strengths and weaknesses of a theory which, based upon citation evidence, has had a significant impact on other researchers within the information systems discipline. Because the evaluation of existing theories often provides the basis for refining existing theories or building new theories, I also show how the framework and criteria can be used to inform the development of high-quality theory.

Keywords: *Theory Evaluation, Theory Refinement, Theory Building, Constructs, Associations, States, Events, Boundary Conditions, Theoretical Importance, Theoretical Novelty, Parsimony, Meso-Level Theory, Falsifiability.*

* Cynthia Beath was the accepting senior editor. This article was submitted on 3rd March 2011 and went through two revisions.

Volume 13, Issue 1, pp. 1-30, January 2012

Nothing is as practical as a good theory (Lewin, 1945, p. 129).

1. Introduction

For many researchers, the development of theory within their disciplines is the central goal (the "jewel in the crown") of their research endeavors (e.g., Eisenhardt, 1989; Lewin, 1945; Parsons, 1950; Shapira, 2011). By articulating high-quality theory, they believe they are more likely to enhance their own knowledge of and other scholars' knowledge of the domain covered by their theory. Some also believe they can enhance practitioners' capabilities to operate effectively and efficiently in the domain covered by their theory. For these reasons, scholars in a number of disciplines have, from time to time, sought to articulate the nature of and characteristics of high-quality theory (e.g., Blalock, 1971; Dubin, 1978; Fetzer, 1993; Godfrey-Smith, 2003; Van de Ven, 1989; Weick, 1995, 1999; Whetten, 1989).

In spite of the importance that many researchers ascribe to theory, the development of new theory and the refinement of existing theories have been relatively neglected features of research within the information systems discipline. As a result, in the late 1990s and early 2000s, several editors of major information systems journals appealed for more theoretical contributions to be made to the discipline (e.g., Webster & Watson, 2002; Zmud, 1998).

To date, only a relatively small number of papers have been published within the information systems discipline that researchers might use to guide their development of new, *high-quality* theories and their refinement of existing theories (e.g., Gregor, 2006; Grover, Lyytinen, Srinivasan, & Tan, 2008; Markus & Robey, 1988). Moreover, even if information systems researchers were to seek guidance from other disciplines where attention to theory building has had a longer history, the processes they might use to develop and refine theories remain an arcane affair (e.g., Dubin, 1978; Freese, 1980; Sutton & Staw, 1995).

In this paper, I propose a framework and criteria that can be used to evaluate the quality of a theory. While my framework and criteria build on the work of many other scholars (e.g., see the references above), I believe my contribution is novel in its reliance on a theory of ontology to provide more formal and precise foundations for the evaluation of theory. I show how the framework and criteria I articulate can be employed to pinpoint the strengths and weaknesses of an existing theory within the information systems discipline – one that, according to citation evidence, has had a significant impact on many researchers within the discipline.

I focus on providing a framework and criteria for theory *evaluation* for four reasons. First, I seek to provide a means of identifying the likely usefulness of a theory as a basis for predicting and/or explaining real-world phenomena. Second, I wish to provide a way of pinpointing areas where empirical tests of the theory are likely to be problematic. Third, I seek to provide a method of identifying opportunities for refining a theory. Fourth, I wish to provide guidance in relation to the development of new, high-quality theory.

The structure of the paper is as follow. First, I briefly define some ontological constructs that enable me to define the meaning I ascribe to the term "theory" and to articulate the framework and criteria I propose for evaluating the quality of a theory more precisely. Second, I explain the meaning I ascribe to the term "theory". Third, I describe the framework and criteria that I propose for evaluating the quality of a theory. Fourth, I attempt to show the usefulness of the framework and criteria by applying them to the evaluation of an important, extant information systems theory. Fifth, I canvass how the framework and criteria can be used to inform the refinement of existing theories and the development of new, high-quality theories. Finally, I provide some reflections and conclusions.

2. Some Basic Ontology

Theories provide a *representation* of someone's perceptions of how a subset of real-world phenomena should be described. In this light, they can be conceived as *specialized* ontologies – *instances* of a *general* ontology (a theory about the nature of and makeup of the real world, in general). For this reason, I argue that any careful analysis of the notion of and components of a theory has to be rooted in a rigorously formulated *generalized* ontology. The elements of a *specific* theory can then be evaluated in terms of how well they map to or instantiate this generalized ontology.

In the sections below, I have used a formal, generalized ontology proposed by Bunge (1977, 1979) as the basis for my analyses. His ontology seeks to "stake out the main traits of the real world…in a clear and systematic way…to produce a unified picture of reality" (Bunge 1977, p. 5). I have chosen Bunge's ontology for two reasons: (a) it is the most rigorously formulated ontology that I have discovered; and (b) I have found it to be useful in elucidating ideas about theory that, in my view, have long remained vague and imprecise. Table 1 provides a succinct (and somewhat informal) explanation of some key constructs in Bunge's ontology.

Table 1. Some Fundamental Ontological Constructs

Construct	Explanation
Thing	The world is made of things. Things can be *substantial* or *concrete* (e.g., an information system user or a computer); alternatively, they can be *conceptual* (e.g., a mathematical set or a function). In this paper, my focus is primarily on *concrete* things.
Composite Thing	Some things are made up of other things (e.g., a system development team, which is a composite thing, is made of team members such as programmers or analysts, which are its components).
Property	All concrete things in the world possess properties (there are no formless things). Similarly, all properties in the world attach to some thing (properties do not exist in isolation from things). Properties are *not* things, however; they are separate ontological constructs that describe different elements (features) of the world. For example, a human (a concrete thing) may possess a property that he uses an information system, and a computer (a concrete thing) has the property of possessing a certain amount of internal memory.
Class	Things that possess at least one property in common constitute a class of things. For example, all humans who use an information system are members of the class of things called "information system users."
Attribute	We "know" about properties of things in the world through our *perceptions* of them. These perceptions may be more or less true. The way in which we perceive a property at a point in time (our representation of it) is called an *attribute*. Various types of attributes exist: • *Attributes in general* are attributes that belong to a *class* of things (e.g., all humans in the class "information system users" possess the attribute called "uses an information system"). • *Attributes in particular* are attributes that belong to *specific* things in a class of things (e.g., the specific person called "John" in the class of things called "users of information systems" possesses the particular attribute "uses an information system three times each day"). • *Intrinsic attributes* represent properties of individual things (intrinsic attribute in particular) or classes of individual things (intrinsic attribute in general). For example, a specific user of an information system called "Jane" has the intrinsic attribute in particular of "age=40 years," and the class of things called "information system users" has the intrinsic attribute in general called "age"). • *Mutual attributes* represent properties of two or more particular things (mutual attribute in particular) or two or more classes of things (mutual attribute in general). For example, system analysts and information system users have the mutual attribute in general called "level of shared understanding about the requirements for a new information system," which will take on a specific value (mutual attribute in particular) for a specific system analyst-information system user pair). • *Emergent attributes* (in particular or in general) are attributes of composite things that do not belong to their components. Nonetheless, they are related in some way to attributes of their components (e.g., the work productivity of a system development team has no meaning in terms of each team member, but it is related in some way to the productivity of each team member). • *Complex attributes* (in particular or in general) are attributes that are made up of the conjunction of simple attributes (e.g., the attribute "system quality" is composed of simpler attributes such as "response time," "data accuracy," ease of use," and so on).

Table 1. Some Fundamental Ontological Constructs (cont.)

Construct	Explanation
State	A vector of *attributes in particular* represents a state of a thing (its attributes in general along with their associated values). States can also be conceived as a complex attribute in particular. For instance, a particular user of an information system has two attributes in particular (measured on a 10-point scale) that relate to the information system: "perceived ease of use = 5" and "perceived usefulness = 8"). The vector (5,8) corresponding to the values of these two attributes represents the state of the user. The complex attribute *in general* might be called "perceived utility," and for the particular user it has the value (5,8).
Lawful State	Some states of a thing are deemed *lawful* (they obey natural or human-made laws); others are deemed unlawful. For instance, natural laws (the laws of physics) restrict the minimum response time that an online information system can achieve. Response times below this minimum amount are unlawful. A social law – legislation to protect the privacy of customers' data – might constrain an organization from outsourcing its information systems to providers in certain foreign countries.
Event	An event that a thing undergoes is represented by a change from one of its states to another of its states (at least one of its attributes changes values). For instance, in light of a user's ongoing use of an information system, her perceptions might change from one state (perceived ease of use=5, perceived usefulness=8) to another state (perceived ease of use=6, perceived usefulness=5). The event is represented by <(5,8),(6,5)>.
Lawful Event	Some events that a thing undergoes are deemed *lawful* (they obey natural or human-made laws); others are deemed unlawful. If an event has an unlawful beginning or end state, it will be unlawful. Some events are unlawful, however, even when their beginning and end states are *lawful*. For example, "has low experience with an information system X" and "has high experience with an information system X" are lawful states of a human thing. The event represented by the state change from "has low experience an information system X" to "has high experience with an information system X" is lawful; the event represented by the state change from "has high experience with an information system X" to "has low experience with an information system X" is unlawful.
History of a Thing	The history of a thing is a *sequence* (ordered set) of its states (e.g., the states that a thing traverses over time are ordered by time). For example, a user's perceived ease of use and perceived usefulness of a particular information system at three points in time might be represented by the following sequence of pairs: <(5,8),(6,5),(7,5)>. The three pairs show the history of the user.
Interaction between Things	Two things interact when the history of one thing is not independent of the history of the other thing. For example, consider two users, X and Y, of a particular information system. If they never meet during their use of the information system, assume X's history (at three points in time) of perceived ease of use and perceived usefulness is <(5,8),(6,5),(7,5)> and Y's history of perceived ease of use and perceived usefulness is <(6,7),(7,7),(7,8)>. If X and Y interact during their use of the information system, one or both might then have different perceptions about the ease of use and usefulness of the information system. For instance, Y might assist X to use the information system such that X's history is now <(5,8),(6,9),(8,9)>.

In the sections below, I show how Bunge's ontology can be used to make precise the meaning of terms frequently used in discussions about theory. I show, also, how this enhanced precision can be used to pinpoint the strengths and weaknesses of a theory.

3. Nature of Theory

Different researchers often ascribe different meanings to the term "theory." Indeed, the extant literature shows a considerable level of disagreement about what constitutes a theory, and what constitutes "strong" theory versus "weak" theory (see, e.g., Sutton & Staw, 1995, p. 371). In this section, therefore, I explain the meaning I ascribe to the term "theory" to provide a context for the conditions I lay out subsequently for a theory to be considered "strong."

3.1. A Particular View of What Constitutes a Theory

By *theory*, I mean a particular kind of *model* that is intended to *account* for some subset of *phenomena* in the real world. A theory is a *social construction* (Jaccard & Jacoby, 2010, pp. 7-10). It is an artifact built by humans to achieve some purpose. It is a *conceptual* thing rather than a concrete thing. Nonetheless, it has a *concrete* manifestation as a neuronal pattern in some person's brain.

By *phenomena*, I mean someone's *perceptions* of *facts* in the real world – the existence of things, the properties these things possess, the states these things experience, and the events these things undergo (see Table 1). The subset of phenomena in the world that the theory is intended to cover is called its *domain*. The phenomena in a theory's domain, in turn, can be partitioned into two subsets: (a) the *focal phenomena*, which are the primary focus of the theory; and (b) the *ancillary phenomena*, which are somehow associated either directly or indirectly with the focal phenomena.

The phenomena that are the focus of a theory usually apply to things in a *class* or things in several *classes*. Of course, it is possible to construct a theory about the properties, states, and events pertaining to an *individual thing* in the world (e.g., a *specific* person or a *specific* information system). For the most part, however, researchers are interested in constructing theories to account for phenomena that are common to more than one thing.

The phenomena that theories cover may be *static* phenomena (states of things at a point in time), *dynamic* phenomena (events that occur to things), or a *combination* of both static and dynamic phenomena. If the theory covers static phenomena, the researcher who proposes the theory should make clear whether the states of a thing (or things) that are covered are intended to be *stable* (in equilibrium) or *unstable* (in transition to equilibrium). Empiricists who wish to test a theory that covers static phenomena need to know whether they should measure the phenomena when they are in a stable state or an unstable state.

By *account*, I mean a theory assists its users to *explain* and *predict* its focal phenomena. Some researchers argue theories can have still another purpose – namely, to facilitate human *understanding* of the theory's focal phenomena (e.g., Hovorka & Lee, 2010). I do not see how a high-quality *explanation* of focal phenomena can occur, however, without first *understanding* the focal phenomena. For this reason, I intend the purpose of explanation to encompass the purpose of understanding.

By *model*, I mean an abstracted, simplified, concise *representation* of something else (phenomena) in the world. Models help us to comprehend the world by representing only those major features of the world that are important for our purposes. Often they provide only an approximate account of the complexity that exists in the real-world phenomena they cover. They compromise precision to achieve cognitive economy.

Theories are particular kinds of models, however (see Section 4 below). All theories are models, but not all models are theories. A model must satisfy certain conditions before I deem it to be a theory (see Section 4 below) – conditions that relate to rigorous specification of its "parts" and particular qualities of its "whole." Thus, the existence of a model is a *necessary* condition for the existence of a theory, but it is not a *sufficient* condition. The existence of a theory, however, is a *sufficient* condition for the existence of a model.

3.2. Some Prior Notions of Theory

To further clarify my notion of "theory," consider the taxonomy of theories proposed by Gregor (2006). Based on an extensive review of prior literature, she identifies *five* ways in which the term "theory" has been used: Type I – theories for analysis; Type II – theories for explanation; Type III – theories for prediction; Type IV – theories for explanation and prediction; and Type V – theories for design and action.

In my view, those she calls Type-I theories (theories for analysis) are *typologies* and not theories (see also Bacharach, 1989, p. 497). Typologies underpin precise definition of the constructs in a theory, but they lack some characteristics that I deem important to a theory (see my arguments in Section 4 below). Furthermore, in my view, those Gregor (2006) calls Type-V theories (theories for design and action) are *models* but not theories. As with typologies, models lack some characteristics that I deem important to a theory (again, see my arguments in Section 4 below).

Gregor's (2006) Type-II theories (theories for explaining) and Type-III theories (theories for predicting) may or may not, in my view, constitute theories (depending on how rigorously their "parts" have been articulated and the qualities possessed by their "whole"). For instance, Gregor argues (p. 625) some types of Type-II theories are used primarily as high-level "sensitizing devices." They sometimes lack clarity and precision in relation to their constructs, relationships among constructs, states they cover, and events they cover. Type-II theories often are used within the so-called interpretivist research paradigm (Klein & Myers 1999, p. 75). Where rigor is lacking, I see Type-II theories as constituting *models* but not theories. Similarly, Gregor argues (p. 626) Type-III theories do not always provide a clear account of the associations among the constructs they employ. If this is the case, again, in my view they do not constitute theories. They are models only, because they lack certain qualities needed to constitute a theory – namely, rigorous specification of all their "parts," which, in turn, undermines some qualities of their "whole."

In short, my notion of theory is best aligned with Gregor's (2006) Type IV theory – a theory for explanation and prediction. Nonetheless, my reluctance to ascribe the term "theory" to her other types of theory (Types I, II, III, and V) in no way is intended to denigrate the significance of these types of contributions in scholars' research endeavors. Indeed, Gregor (2006) provides compelling explanations for why they make important contributions to the development of knowledge. Rather, I am seeking to be clear about *my notion* of what constitutes a theory to provide a context for the arguments I develop below.

4. A Framework and Criteria for Theory Evaluation

In this section, I argue a theory should be evaluated from two perspectives. The first is the "parts" – the evaluation should focus on the quality of the individual components that make up the theory. I provide criteria for evaluating these components. The second is the "whole" – the evaluation should focus on the quality of the theory considered *in toto*. I also provide criteria for evaluating the whole.

Both forms of evaluation are important in assessing the quality of a theory. It is unlikely the quality of the whole will be high if the quality of the parts is not high. Nonetheless, high-quality parts are not a *sufficient* condition for a high-quality whole. To the extent a *model* satisfies the criteria for high-quality parts and a high-quality whole, it can be deemed a *theory*.

4.1. Parts

All theories have three parts: their constructs; their associations; and the states they cover. In addition, theories that cover dynamic phenomena have a fourth part – namely, the events they cover. When evaluating a theory, the focus initially should be on the quality of its parts.

The parts of a theory need to be described precisely because they circumscribe the *boundary* or *domain* of the theory – that is, the phenomena it is intended to cover. If researchers have a clear understanding of the theory's parts, they are better able to design tests that fall within the theory's domain rather than unwittingly testing the theory in an inappropriate context (a context outside the boundary of the theory). They are also better able to *filter* data they have collected so they undertake tests on only the subset representing phenomena in the domain the theory covers. Indeed, some scholars argue that a field's understanding of the boundary conditions associated with its theories is a good proxy for the quality of its theories and the state of the field more generally (e.g., Gray & Cooper, 2010, p. 627).

The following subsections explain the nature of each part. They also describe criteria that can be used to evaluate how well a researcher has articulated each of the parts. Figure 1 provides an overview of the analysis that follows.

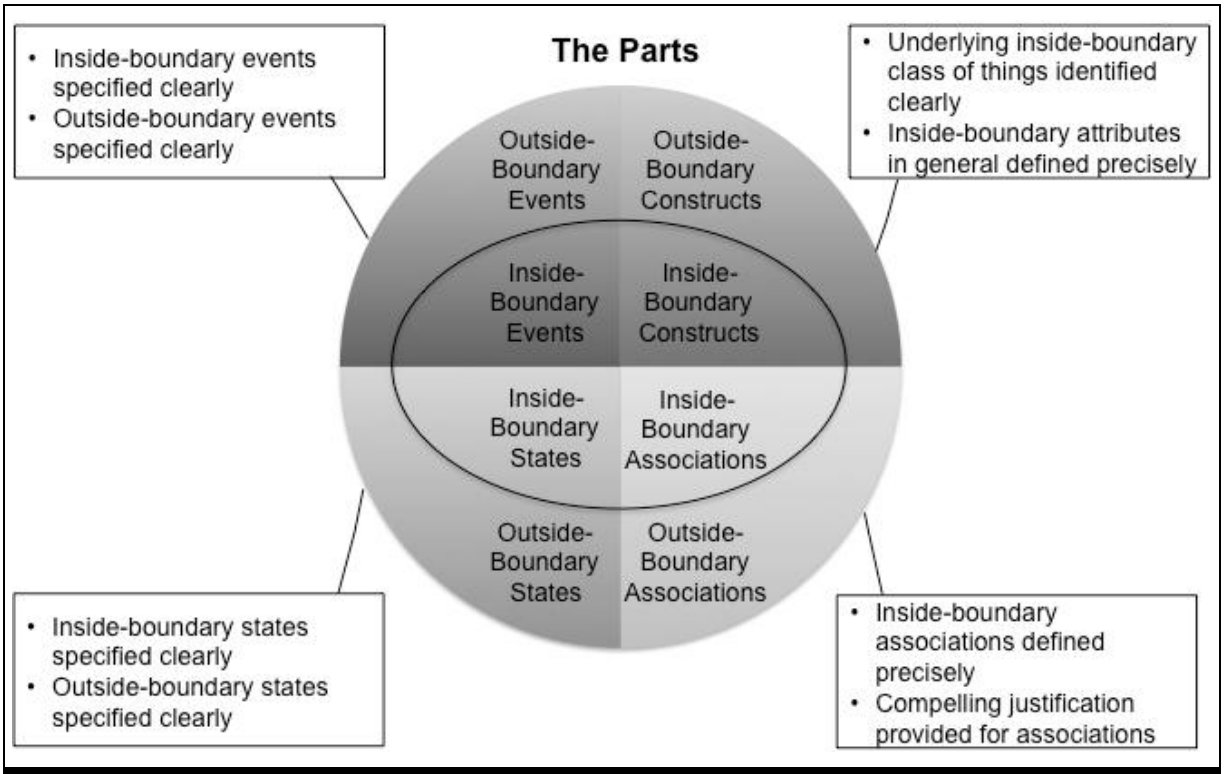

Figure 1. Framework and Criteria for Evaluating a Theory's "Parts"

4.1.1. Constructs

A *construct* in a theory represents an *attribute in general* of some *class of things* in its domain (as opposed to a *particular attribute* of a *specific* thing). The classes of things to which attributes in general pertain ought to be defined precisely to ensure that the *meanings* of each class and the things in each class are clear. Otherwise, the exact nature of the things that the theory covers will not be clear. Moreover, the meanings of the attributes in general that attach to the classes of things the theory covers are unlikely to be clear. Attributes do not float in the ether; they always attach to things. As a first step in clarifying the meaning of an attribute, therefore, the thing to which it attaches needs to be made clear.

For example, the well-known Technology Acceptance Model (TAM) (Davis, 1989) is a theory that had in its earliest versions (and many later versions) only one class of things in its domain – namely, individual users of some form of information technology. To be a member of this class, things had to possess only two attributes in general: (a) they had to be humans, and (b) they had to be users of some form of information technology. Similarly, Sambamurthy, Bharadwaj, and Grover (2003, p. 241) state that their theory covers only a particular class of things: "…our theory's boundary condition is firms operating in moderate to rapidly changing business environments, such as the high-tech, retailing, and financial services sectors."

Once the meanings of the classes of things that a theory covers are clear, the nature of each attribute in general that pertains to a particular class ought to be defined precisely. Unless the meanings of the attributes in general are clear, the meanings of any associations among them will not be clear. Moreover, developing credible (valid and reliable) empirical indicators of the attributes in general will be difficult (if not impossible). Interpreting the meaning of data collected about the attributes will also be difficult (if not impossible).

For example, Davis (1989, p. 320) defines precisely two attributes in general of TAM's single class of things (individual users of information systems): (a) *perceived usefulness*, and (b) *perceived ease of use*. Variations in the values of these two attributes in general for particular users of an information technology are the *ancillary* phenomena in TAM's domain. Variations in the values of a third attribute in general, *system usage*, are the *focal* phenomena in TAM (but, interestingly, Davis defines system usage only indirectly in his paper).

When two or more constructs in a theory represent attributes in general of the *same* class of things, care must be taken that the constructs are not different *proxies* for the *same* underlying property in general of the class of things. Otherwise, any association detected between variations in the values of the constructs simply manifests a variation in the values of a single underlying property in general of the class of things (in other words, it is a tautological association). Because we can only know the properties of things imperfectly (hence our use of attributes), the evaluation of whether attributes overlap in their representation of a property is important but is sometimes difficult to undertake.

For example, in TAM, perceived usefulness, perceived ease of use, and system usage are all attributes in general of a single class of things (the humans who use some form of information technology). The value of TAM as a theory depends in part upon these three attributes in general not being proxies for a single underlying property in general of the humans who use information technologies.

4.1.2. Associations

Associations in a theory can have multiple meanings. When evaluating the meaning to ascribe to an association, at the outset it is important to reflect upon whether a theory covers only static phenomena, dynamic phenomena, or a combination of both static and dynamic phenomena. It is also important to understand whether the constructs represent attributes in general of a *single* class of things or *multiple* classes of things.

If the theory covers *static phenomena*, an association shows that the values of one construct are somehow related to the values of another construct. The relationship is intended to reflect a *pattern* that is hypothesized to hold across *instances* of things in the class or classes of things that the theory covers. For instance, when "snapshots" of the phenomena that pertain to things in a class are taken at some point in time and the values of attributes of things in the class are examined, the theory might predict that high values for instances of one construct will tend to be associated with low values for instances of another construct.

Associations that cover static phenomena can be specified with varying levels of precision:

- Two constructs are simply shown as related to each other, but the sign is not shown.
- The sign of the association between two constructs is shown, which indicates that the values for instances of one of the constructs are positively or negatively related to the values for instances of the other construct.
- A functional relationship between two constructs is shown. For instance, the value of one construct is shown as twice the value of the other construct.

An association that covers static phenomena may show directionality if the values of one construct are believed to arise *prior to* the values of another construct. For example, at some point in time some researchers might seek to test TAM by measuring the values of perceived ease of use, perceived usefulness, and system usage. Even though they have captured the values of the three constructs at a single point of time, they might believe that the values for perceived ease of use and perceived usefulness would have arisen prior to the value for system usage. Thus, they would show directionality on the association between perceived ease of use and system usage and perceived usefulness and system usage.

An association between two constructs in a theory that covers *dynamic* phenomena (events) shows that a *history of values* for instances of one of the constructs is conditional on a *history of values* for instances of the other construct. In a diagrammatic representation of the association, often an arrow is placed on the association to show which construct's change in values precedes the other construct's change in values.

As with associations that cover static phenomena, associations that cover dynamic phenomena can be specified with varying levels of precision:

- Two constructs are simply shown as related to each other, but neither the sign nor the direction of the association is revealed. Perhaps uncertainty exists about the sign or direction of the association. Some indication is given, however, that changes in the value for an instance of one construct precede changes in the value for an instance of the other construct, even if the sign or direction is unclear (e.g., the uncertain nature of the dynamics is explained in a narrative).

- The sign of the association between two constructs is shown, which indicates that changes in the value for an instance of one of the constructs are positively or negatively correlated with subsequent changes in the value for an instance of the other construct. Uncertainty exists, however, about the direction of the association.

- The direction of the association between two constructs is shown, which implies the existence of causality or shows a time relationship among changes in the values for instances of the constructs – for instance, changes in the value for an instance of one construct *cause* a change in the value for an instance of the other construct, or a change in the value for an instance of one construct *precedes* a change in the value for an instance of the other construct.

- A functional association is shown between two constructs. In other words, the *amount* of change that occurs in the value for an instance of one construct is shown as a result of or subsequent to the *amount* of change that occurs in the value for an instance of another construct.

The constructs in an association may pertain to a single *class* of things or *multiple* classes of things. If two constructs represent different attributes in general of a *single* class of things, any association between the two attributes implies they are "lawfully" related in some way (association type (a) in Figure 2). In other words, for an instance of a thing in the class, the value or a change in the value of one of its attributes is related to a value or a change in the value of another of its attributes.

For example, in Sambamurthy et al.'s (2003) theory, the attributes in general named "digital options" and "agility" belong to a single class of things (firms of a certain type). Thus, the association that Sambamurthy et al. (2003) posit exists between these two attributes means they are lawfully related. They propose a two-way causal (and, therefore, time-sequenced) relationship in which "higher levels of agility will…enhance digital options" (p. 255) and "the impact of digital options…on agility" will be positively moderated by a third construct (namely, "entrepreneurial alertness," which is also an attribute in general of the type of firm their theory covers) (p. 253).

If two constructs represent different attributes in general of two *different* classes of things, any association between them means at least one instance of a thing in one class *interacts* with at least one instance of a thing in the other class (association type (b) in Figure 2). In other words, the history of a thing in one class of things is not independent of some thing in the other class of things. The nature of the interactions between the two things is manifested in the attributes that are related.

For example, an extension to the original TAM might posit that variations in the measured response time (as opposed to *perceived* response time) of online information systems might be associated with users' perceptions of these systems' usefulness. Whereas "perceived usefulness" is an attribute in general of a class of things called "online information system users," "measured response time" is an attribute in general of a class of things called "online information systems." At least one thing in one class interacts with at least one thing in the other class. For example, variations in a particular user's perceptions of the perceived usefulness of an online information system she is using are associated with variations in the measured response time of the system.

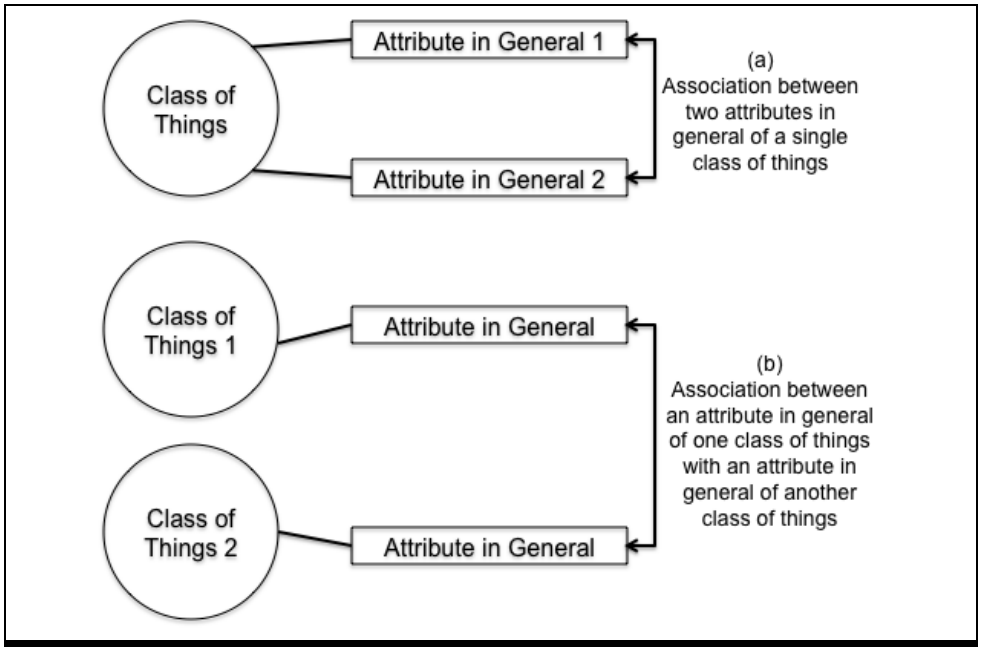

Figure 2. Two Types of Associations in a Theory

Theories that cover dynamic phenomena (events) sometimes *replicate* constructs in diagrammatic representations of the theory and show directional associations between the replicated constructs (e.g., Jaccard & Jacoby, 2010, pp. 159-161). For instance, a theory might be proposed to account for the evolution of system analysts' knowledge as they attempt to build an entity-relationship model of an application domain. Several instances of a construct called "system analyst's knowledge" might appear in a diagrammatic representation of the theory, all of which are linked by directional associations. The theory might show that one value of the construct (e.g., an "entities-known" state) precedes another value of the construct (e.g., a "relationships-known" state), which, in turn, precedes another value of the construct (e.g., an "attributes-known" state). In short, when system analysts build an entity-relationship model of an application domain, the theory indicates that they first acquire knowledge of the *entities* in the domain, then acquire knowledge of the *relationships* in the domain, and then acquire knowledge of the *attributes* in the domain.

Alternatively, diagrammatic representations of dynamic phenomena associated with a construct sometimes use a graph, where the possible values of an instance of the construct are shown on one axis and time is shown on the other axis (e.g., Monge, 1990, pp. 411-413, 415-419). The graph shows typical values of an instance of the construct as they unfold over time.

To the extent the nature of the associations among constructs in a theory is made more explicit, more-powerful empirical tests of the theory can be done (tests that are more likely to lead to the theory *not* being supported). For example, assume a theory shows only *associations* or only *directional associations* among its constructs. Perhaps counter-intuitively, *more-rigorous* research designs lead to *weaker* tests of such a theory (a paradox noted by Meehl, 1967). The reason is that even a small amount of covariation between the values of any two constructs in the theory is likely to be detected in the data obtained – that is, the null hypothesis is likely to be rejected. Whether the covariation reflects ambient noise in the data (the average variance that is common to unrelated constructs) or the *magnitude* of the association detected is practically significant, however, is another matter.

On the other hand, if the theory articulated associations with a *functional form* (e.g., concave, convex, stepped linear), posited the values of its associations' parameters (e.g., two constructs will be associated with a correlation coefficient of 0.70), or enunciated contingent (moderated) associations, more rigorous research designs are more likely to lead to obtaining data that shows *lack of support* for the theory (Edwards & Berry, 2010; Leavitt, Mitchell, & Peterson, 2010). In other words, stronger tests of the theory are possible.

Similarly, if a theory is intended to cover a time series of dynamic phenomena (events), stronger tests of the theory can be undertaken if change parameters are specified more precisely – for instance, whether the change is *continuous*, the *magnitude* of value or state changes at different points of time, the *rate of change* of values or states, the *trend* in values or states, the *periodicity* of changes in values or states (length of time between the same or similar values or states), and the *duration* for which a particular value of state remains constant (Abbott, 1990; Monge, 1990). The extent to which these sorts of parameters can be specified precisely, however, depends of the quality of the "narrative" used to explain the nature of the association (transition) between different values or states of a construct (e.g., Langley, 1999; Pentland, 1999; Poole, Van de Ven, Dooley, & Holmes, 2000).

Usually, a theory does not cover all possible associations among its constructs. Instead, researchers seek to make astute decisions about what associations to include in the theory and what associations to omit from the theory. The omission of an association among constructs in a theory does not necessarily mean none exists among the constructs. Rather, the researcher who proposes the theory has deemed the association to be outside the boundary of the theory.

Researchers ought to provide compelling justifications for whatever associations they decide to include within the boundary of their theory. On the one hand, a justificatory narrative may be compelling because it provides interesting, novel insights about how two or more constructs are related or a time series of events in a construct unfolds. It engages other researchers because it reveals aspects of the focal or ancillary phenomena that previously they had not contemplated. On the other hand, a justificatory narrative may be compelling because it is deemed uncontroversial or it is rooted in prior, widely accepted research–for instance, prior theoretical work or prior empirical work. In other words, it is compelling because it is relatively unsurprising–it resonates with knowledge that other researchers already possess. Whichever justificatory basis is used, the association should be material. It is included within the boundary of the theory because it is expected to account for a significant amount of shared variation between the constructs it covers, or it represents a strong causal link between the constructs it covers, or it shows an important time series of events in a construct it covers.

4.1.3. States

A theory should specify clearly and as precisely as possible the state space of things in the class or classes of things that it is intended to cover. In other words, it should stipulate those states that might arise for things in the class or classes of things that fall within its domain and for which it is intended to have explanatory and predictive power.

To determine the *state space* that falls within a theory's boundary, the range of values that each construct in the theory might cover first needs to be determined. All combinations of values (the Cartesian product) for the set of constructs in the theory then need to be considered. The set of combinations of values constitutes the *conceivable state space* for the set of constructs in the theory. Some combinations can be eliminated because they cannot occur *naturally* (they are unlawful, at least for the class or classes of things that the theory is intended to cover) and, thus, fall outside the boundary of the theory. Those that remain must then be evaluated to determine whether they are covered by the theory. In other words, states that can occur naturally must be partitioned into *inside-boundary states* and *outside-boundary states*.

To illustrate these concepts, assume we are developing a theory based on TAM about how system usage will change over time as users of some form of information technology experience changes in their perceptions of its ease of use and usefulness. At the outset, we might wish to be more specific than TAM about the class of things (individual users of some form of information technology) covered by our theory. For example, we might indicate our theory covers only individual users who are rational according to some criteria, already have some level of exposure to information technology, have volitional use of a form of information technology designed specifically to support individuals (rather than groups), work only in certain application domains, and are not subject to intense time pressures in the tasks they undertake. (Recall, I have argued a high-quality theory will be especially clear about the class or classes of things it covers.)

To be congruent with TAM, assume we have only three constructs in our theory (all are *attributes in general* of individual users of some form of information technology): *system usage*, *perceived ease of use*, and *perceived usefulness*. Assume, also, that the three constructs have been defined precisely. For simplicity, assume we have a valid and reliable measure of each construct, and each measure uses a 10-point scale (with 1 representing low values on the scale and 10 representing high values on the scale). In this light, our theory's *conceivable state space* contains 1,000 states (10×10×10).

We might then argue some states can be eliminated because they are unlawful. For example, assume we believe no user who falls within the class of things covered by our theory will have a value of system usage of 9 or 10, but will have values for perceived ease of use and perceived usefulness that are both equal to or less than 2. Thus, eight states are unlawful in our theory because we believe they will never occur "naturally": (9,1,1), (9,1,2), (9,2,1), (9,2,2), (10,1,1), (10,1,2), (10,2,1), (10,2,2). For instance, we might believe such states would occur only in relation to users who are irrational or forced to use the system. We might then conclude, however, that none of the remaining 992 states falls outside the boundary of our theory – in other words, the set of *outside-boundary states* is the null set. In short, we intend our theory to cover all remaining 992 states that might arise as a result of relationships among the three constructs.

Our TAM example illustrates that practically, for many theories, *each* state in the set of conceivable states of a theory cannot be considered to determine whether it falls within the boundary of the theory. The reason is that the number of conceivable states is too large. Nonetheless, our TAM example illustrates the kind of clarity about states that theoreticians must strive to achieve in specifying the boundaries of their theories. While they may find it difficult or impossible to list exhaustively all the states they need to consider, they might be able to describe the broad characteristics of states that fall inside or outside the boundary of their theory. For instance, they might seek to identify the characteristics of extreme-value states or unlikely states and to indicate whether such states fall inside or outside the boundary of their theory. Crafting a clear narrative to describe the subspace of the conceivable state space that the theory covers is often a difficult task. Perhaps for this reason, consideration of a theory's inside-boundary states is often missing or treated only cursorily in theoretical accounts.

4.1.4. Events

If a theory is intended to cover events, the event space that falls within the theory's boundary must also be articulated. At the outset, all conceptually possible *pairs* of inside-boundary states must be considered (recall, each event can be conceived as a before-state, after-state pair). These constitute the *conceivable event space* covered by the theory. Some combinations can be eliminated because they cannot occur naturally (they are unlawful). Those that remain must be evaluated to determine whether they are covered by the theory. In other words, they must be partitioned into *inside-boundary events* and *outside-boundary events*.

To illustrate these concepts, consider again our putative theory based on TAM about how system usage will change over time as users of some form of information technology experience changes in their perceptions of the technology's ease of use and usefulness. Recall, we have 992 inside-boundary states in our theory. In this light, the conceivable event space in our theory contains 984,064 events (992 before-states×992 after-states). We might conclude initially that some of these events are *unlawful* because they cannot occur naturally for users in the class of information technology users who fall within the boundary of our theory – specifically, those events that involve (a) an *increase* in the values of *both perceived ease of use* and *perceived usefulness* but a *decrease* in the value of *system usage*, or (b) a *decrease* in the values of *both perceived ease of use* and *perceived usefulness* but an *increase* in the value of *system usage*. We might believe, for instance, that such events would occur only for information system users who are irrational (those outside the boundary of our theory). In this light, the following two events are examples of those we deem to be unlawful: <(8,4,5)(6,5,7)> and <(3,4,3)(5,3,1)>.

Of the remaining lawful events, we might then deem any event involving a change in value of 6 or more in both *perceived ease of use* and *perceived usefulness* falls outside the boundary of our theory. In other words, our theory is not intended to cover events that involve large changes of values in these constructs. We might believe, for instance, that such events would occur only when some other infrequently occurring factor has had a major impact on information system users and, thus, our ability to explain and predict system usage via our theory is undermined. As a result, the following two events are examples of events that would fall outside the boundary of our theory: <(8,8,9)(7,2,1)> and <(3,3,1)(4,10,8)>.

As with states, our TAM example illustrates that practically, for many theories, *each* event in the set of conceivable events of a theory cannot be considered to determine whether it falls within the boundary of the theory because the number of conceivable events is too large. In this light, theorists might seek to identify the characteristics of extreme events or unlikely events and to indicate whether such events fall inside or outside the boundary of their theory. As with states, however, crafting a clear narrative that describes the inside-boundary events is often difficult. For this reason, it too is often missing or treated only cursorily in theoretical accounts.

4.2. Whole

A theory has *emergent* attributes – attributes of the theory as a *whole* rather than attributes of its parts. Many such attributes exist, and researchers often differ in their views on the significance they ascribe to each of them. Nonetheless, some emergent attributes have widespread acceptance among researchers as being significant when assessing the quality of a theory. The following subsections explain the nature of these attributes and describe criteria that can be used to evaluate the extent to which a theory possesses them. Figure 3 provides an overview of the analysis that follows.

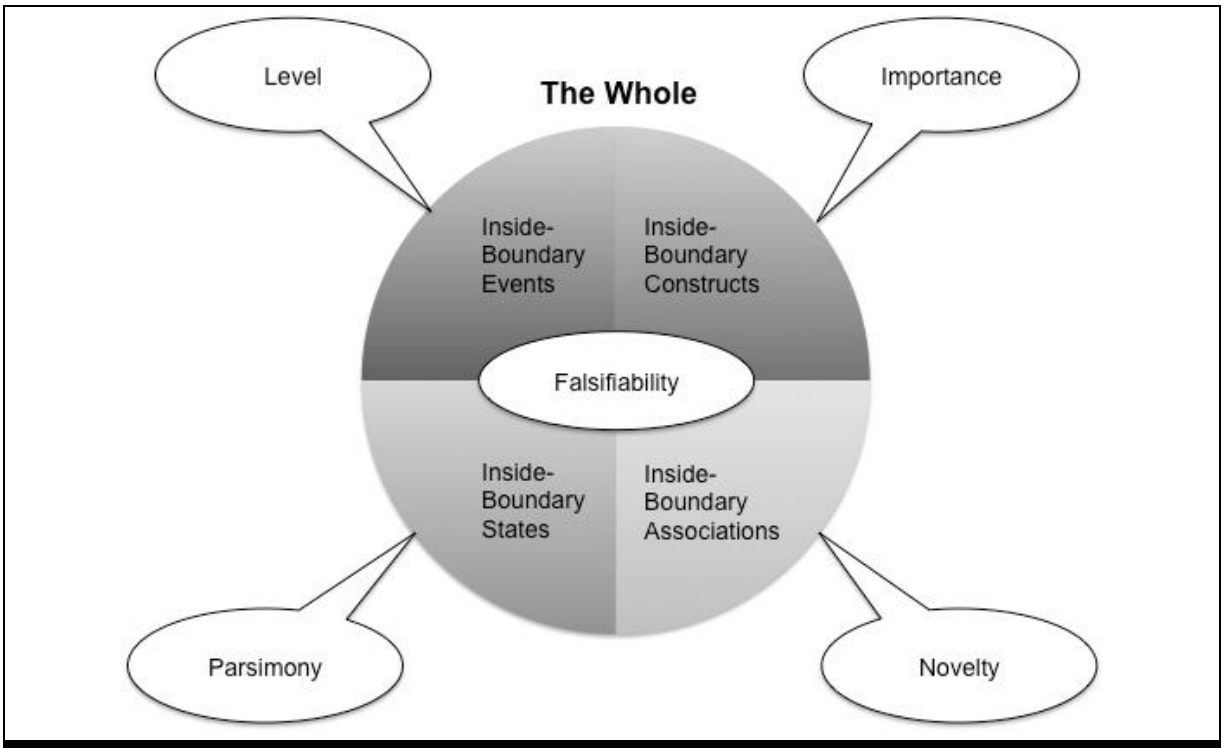

Figure 3. Framework and Criteria for Evaluating a Theory's "Whole"

4.2.1. Importance

The importance (or utility) of a theory is often assessed via judgments made about the importance of its focal phenomena (Corley & Gioia, 2011, pp. 17-19). Usually, there is little point to having a theory with rigorously specified constructs, associations, inside-boundary states, and inside-boundary events if it addresses uninteresting phenomena (Weber, 2003a). The focal phenomena might be deemed

important from the viewpoint of *practice* (improving the effectiveness and efficiency of some entity's activities). They might also be deemed important from the viewpoint of *research* (science). Potentially, enhanced understanding of the focal phenomena will provide key insights that enable theoretical or empirical progress to be made on some problem within a discipline.

Ex ante it might be difficult to judge the importance of a theory. At the outset, its potential impact on researchers and practitioners might be difficult to evaluate. Moreover, sometimes a theory provides insights not anticipated when it was first articulated. Such insights arise only when researchers engage with the theory and use it to inform their empirical work.

Ex post, however, various metrics can be used to assess the importance of a theory. For example, the extent to which researchers cite a theory provides an indicator of its impact on their work and, thus, its likely importance to them. Similarly, whether the theory underpins consulting work that practitioners undertake or provides the foundation for a successful patent or motivates the establishment of a new company is an indicator of its importance.

Citation evidence must be treated cautiously when it is used as a proxy for the importance of a theory. Some theories are appealing to researchers because they are relatively simple to test empirically. They are perceived as an easy route to journal publications. Whether they provide deep insights into the phenomena they cover, however, is another matter.

4.2.2. Novelty

The extent to which a theory is novel is an important factor determining (a) the value ascribed to it by researchers, and (b) the likelihood that papers describing the theory will be accepted for publication in major journals (e.g., Colquitt & Zapata-Phelan, 2007; Mone & McKinley, 1993; Corley & Gioia, 2011). Thus, judgments about a theory's novelty or originality and judgments about its contributions to knowledge appear to be closely related. Moreover, "revelatory" or "transformative" theoretical insights, in contrast to "incremental" theoretical insights, seem to be especially valued (Corley & Gioia, 2011, pp. 16-18).

Weber (2003b) describes several ways in which a theory might make novel contributions to a discipline. First, a theory's focal phenomena might not have been covered by prior theories. However, if the theory is simply a slightly modified version of an existing theory that has been applied to new phenomena, it is unlikely to be deemed novel. Second, a theory might be considered novel because it frames or conceives existing, well-known focal phenomena in new ways. For example, Lamb and Kling (2003) is a well-cited paper, perhaps because it argues phenomena associated with information system users need to be considered from a broader, social-actor perspective rather than a narrow, individualistic perspective (the latter perspective previously had been dominant in the literature). Third, a theory's novelty might arise because of important changes it makes to an existing theory – it might add and/or delete constructs and associations, define existing constructs and associations more precisely, or specify the boundary of the theory more precisely.

A theory will be deemed novel (perhaps after some time has elapsed) to the extent it changes the *paradigms* used by researchers to investigate phenomena within their discipline (Kuhn, 1996). It will command the attention of researchers if it provides a way of resolving "anomalies" within their discipline – that is, empirical observations of phenomena that existing theories are unable to explain or predict. It will also command the attention of researchers if it enables them to "see" or conceive new and interesting phenomena (phenomena that previously escaped their attention) or reconceptualize existing phenomena in new and interesting ways. Such theories break the cycle of "normal science" within a discipline and set new paths for the discipline to follow.

The *quality of the rhetoric* used by researchers to describe their theories also appears to be an important factor in determining the extent to which their theories are deemed novel (Locke & Golden-Biddle, 1997). Because science is a social phenomenon, researchers have to convince their colleagues that their work has value. In this light, the arguments researchers use to expound their theories' novelty must be crafted carefully; otherwise, their theories' contribution to knowledge might be overlooked.

After analyzing 82 papers published in the *Academy of Management Journal* and *Administrative Science Quarterly* (two high-quality, high-impact journals) between January 1976 and September 1996, Locke and Golden-Biddle (1997) concluded that researchers who had successfully demonstrated the novelty or contribution of their research use two rhetorical strategies. First, they "legitimize" their work by "constructing intertextual coherence." They "re-present and organize existing knowledge so as to configure a context for contribution" (p. 1029) (see also Webster & Watson, 2002). Second, they "subvert" or "problematize" the existing literature. They do so to show that opportunities exist for contributions to knowledge. One way in which the novelty of a theory can be assessed *ex ante*, therefore, is to evaluate how well its proponents enact Locke and Golden-Biddle's two strategies.

4.2.3. Parsimony

High-quality theories are parsimonious (e.g., Hempel, 1966, pp. 40-45; Popper, 2005, pp. 131, 272). They achieve good levels of predictive and explanatory power in relation to their focal phenomena using a small number of constructs and associations. By using a small number of constructs, they also limit the size of their conceivable state space and conceivable event space. As a result, it is often easier to articulate the nature of states and events that fall within the boundary of the theory. Thus, the boundary of parsimonious theories often can be stated more precisely because the class (or classes) of things, associations, states, and events covered by the theory can be defined precisely.

What constitutes a "small number" is in the eyes of the beholder. Nonetheless, Miller's (1956) classic paper on the "magical number seven, plus or minus two" suggests some guidelines. Humans appear able to manipulate about seven "chunks" of information in short-term memory. In this light, one might predict researchers would deem a theory to be parsimonious if it has no more than about seven constructs and seven associations (and perhaps the desired number in each case is less than seven) and the number of inside-boundary states and inside-boundary events is not large.

When building a theory, researchers are often tempted to include *more* constructs and *more* associations in an attempt to capture the "richness" of the phenomena they are seeking to predict or explain (and my experience is the inclusion of more constructs and associations is often a frequent request made by the reviewers of journal papers!). Parsimony dictates, however, that some constructs and associations must be *omitted* from a theory. In choosing constructs to omit, those whose instances have little variation in their values (states) are likely candidates. In choosing associations to omit, those where few instances of constructs are related to other instances of constructs are likely candidates.

Often, a trade-off must be made between parsimony and a theory's predictive and/or explanatory power. As the number of constructs and associations increases, the theory might be better able to predict and/or explain the focal phenomena. Nonetheless, at some point, users of the theory will deem it to be too complex. The goal is to achieve high levels of prediction and/or explanation with a small number of theoretical components (Ockham's Razor).

4.2.4. Level

Some theories cover a very narrow, constrained set of phenomena. They are often called "micro-level" theories. On the one hand, a micro-level theory's constructs and associations might be defined precisely. Moreover, its predictive and/or explanatory powers might be high in relation to the phenomena it covers. Because of the limited range of phenomena it covers, however, it runs the risk it will be deemed uninteresting and unimportant.

Some theories cover a broad range of phenomena. They are often called "macro-level" theories. In some ways, a macro-level theory might be compelling because it provides broad, overall insights into many phenomena. It has a high level of generality. Often, however, its constructs and associations are defined imprecisely. As a result, its predictive and/or explanatory powers in relation to the more-specific phenomena that are a researcher's focus are limited. In this regard, Weick (1995, pp. 389-390) points out that *generality* can be attained only by trading off a theory's accuracy and/or simplicity (parsimony). Moreover, if a theory is at too high a level and too general, it runs the risk that eventually it will be discredited because it ends up as a "theory of everything" in a discipline (e.g., see Davis, 2010, pp. 697-699) criticisms of New Institutional Theory (DiMaggio & Powell, 1983) in organizational research).

Merton (1957) argues the primary theories used in a discipline ought to be "middle-range" (or "meso-level") theories. On the one hand, such theories avoid "narrow empiricism." On the other hand, they avoid being so general in their coverage that it is difficult, if not impossible, to test them empirically. Moreover, meso-level theories often have value because they link the micro-level world and macro-level world in a discipline.

In spite of the wide acceptance of Merton's idea within many disciplines, the precise meaning of "middle-range theories" remains problematic (Boudon, 1991). Whether a theory is set at an appropriate level is a subjective matter. Also, a level that is too high or too low in one discipline might be an appropriate level in another discipline. Nonetheless, in the context of their discipline, in due course researchers make judgments about whether a theory is formulated at an appropriate level – whether it is too specific or too broad to be interesting and/or useful.

4.2.5. Falsifiability

Most, if not all, theories cannot be *proven* via empirical tests, because it is impossible to test the theory for all things, all instances of associations, all states, and all events that fall within its boundary. Instead, *support* for a theory grows when its powers of prediction and/or explanation remain robust across different tests of the theory (Godfrey-Smith, 2003, pp. 202-218; Hempel, 1966, pp. 33-46). If the theory has been articulated clearly, these tests can be designed strategically. They can be used to examine conditions researchers believe are most likely to lead to the theory being *falsified* (failing the empirical test) rather than supported (Doty & Glick, 1994; Popper, 2005, pp. 57-73). They facilitate "risk-taking" tests of the theory – tests that are likely to disconfirm the theory rather than to support it.

To be capable of falsifying a theory, researchers must be capable of generating precise predictions about the focal phenomena so they can undertake reasonably exact empirical tests of the theory. If the predictions they are able to generate are so vague that the status of empirical tests they undertake always remains problematic or, alternatively, the empirical outcomes can always be finessed (explained) using the theory, the value of the theory is undermined.

Similarly, if the classes of things, associations, states, and events that fall within the boundary of a theory are unclear, the meaning that can be ascribed to empirical tests that fail to support the theory is unclear. The results may simply mean that an invalid and unreliable empirical evaluation of the theory has been undertaken, or the evaluation applies to states and events that fall outside the theory's domain.

High-quality theories also suggest the kinds of empirical work that might lead them to be falsified (or conversely receive strong confirmation). For example, their putative applicability to a wide *variety* of phenomena will be clear (Hempel, 1966, pp. 33-37). Thus, researchers can infer how to test the theory under a large *number* and a great *diversity* of conditions. As discussed above, high-quality theories also enable researchers to "see" new phenomena – phenomena not conceived or considered at the time the theory was formulated (Hempel, 1966, pp. 37-38). Such phenomena provide strong tests of the theory because *a priori* the theory was not formulated to take these phenomena into account.

5. Using the Evaluation Framework and Criteria: An Example

To show how the evaluation framework and criteria I have proposed above can be used to pinpoint the strengths and weaknesses of a theory, assess the likely usefulness of a theory, highlight areas where empirical tests of the theory are likely to be problematic, and identify opportunities for refining and enhancing the theory, in the subsections below, I examine a paper by Griffith, Sawyer, and Neale (2003). This paper examines "the dynamics of knowledge development and transfer in more and less virtual teams" (p. 265). This paper is one of several published in a special issue of the *MIS Quarterly* on the topic of "Redefining the Organizational Roles of Information Technology in the Information Age." The stated purpose of the special issue was to "stimulate ***significant and innovative theoretical thought*** in response to the dramatic changes that had occurred in the 1990s regarding information technology and the transformational ways in which information technology was being applied to enable new forms of organizations and markets" (emphasis in original) (Zmud, 2003, p. 195).

In spite of the special issue's focus on "significant and innovative theoretical thought," it is unclear whether Griffith et al.'s (2003) paper articulates a *theory* or a *model*. Recall, contrary to some scholars (e.g., Dubin, 1978; Whetten, 1989), I have argued above that not all models are theories, whereas theories are models that possess specific attributes. The distinction is important, because I have argued *theories* must satisfy certain standards of rigor, whereas *models* can be used to lay broad foundations for understanding particular phenomena.

On the one hand, Griffith et al.'s (2003) paper states it "advances theory" (p. 265). Moreover, it articulates a number of propositions, which suggests its focus is theory building. On the other hand, it also states it is presenting a "stylized *model*" (my emphasis) of "individual and social knowledge…and how knowledge transfers among individuals and becomes available to the members of the team" (pp. 268-269). Moreover, throughout the paper, the term "model" is used frequently. Nonetheless, the legend for Table 1 of the paper (p. 281) is "Operationalization of Constructs to Test the *Theoretical Model*" (my emphasis), which suggests the model articulated in the paper is, indeed, meant to be a theory. In any event, for the sake of illustrating how the evaluation framework and criteria might be used, I have assumed Griffith et al.'s (2003) paper presents a *theory* of virtualness and knowledge in teams.

In evaluating Griffith et al.'s (2003) paper, I am acutely aware from my own work that theory building is often an extended *process* that involves many painstaking iterations (see, in particular, Weick's (1995) eloquent arguments about the *process* of theory building versus the *product* of theory building). In this regard, Griffith and her colleagues have continued to refine their "stylized model" to improve its rigor and, thereby, to better satisfy the conditions I propose for a model to be called a theory (Cadiz, Sawyer, & Griffith 2009; Griffith & Sawyer, 2010). Again, my purpose in evaluating Griffith et al.'s paper is to illustrate how the framework and criteria I propose can be usefully employed. Moreover, I am seeking to *develop* rather than to subvert or denigrate the substantial contribution that Griffith et al. have made. Like Gray and Cooper (2010, p. 622), I believe we invest too little effort in *developing* existing theories compared to *constructing* new theories. As a result, we have a "clutter" of partially articulated, partially tested theories in the information systems discipline that leads to "overload" and "disarray."

5.1. Parts

In this subsection, I evaluate how rigorously the constructs, associations, states, and events have been articulated in Griffith et al.'s (2003) theory. In this regard, while the evaluation framework and criteria I have proposed above pinpoint those parts of the theory that need to be assessed, readers of the theory still need to make their own judgments about how rigorously each part has been expressed (accordingly, my evaluation below reflects my own judgments). Even where judgments about rigor differ, however, the evaluation framework and criteria provide a way for researchers to structure their discourse about the quality of a theory's components.

5.1.1. Constructs

Griffith et al.'s (2003) paper presents constructs in four places. First, they are shown in Figure 2 (the "stylized model") of the paper (p. 269). Second, they can be gleaned from the 19 propositions stated in the paper (pp. 271-278). Third, Table 1 of their paper (p. 281) is intended to "catalogue the constructs and assessments necessary to test our model" (p. 280). Fourth, specific constructs are discussed at various places in the text of the paper.

A first problem with the paper's articulation of constructs is that inconsistencies exist among those shown in Figure 2 of the paper, those embedded within the propositions in the paper, and those listed in Table 1 of the paper. In this regard, Figure 2 of the paper appears to show 17 constructs employed in the theory. In my reading of the propositions, however, I can identify potentially 31 different constructs (see my Table 2 below). Yet Table 1 of Griffith et al.'s paper shows only 15 constructs that must be subject to "assessments."

At first glance, some inconsistencies appear to represent only *naming* inconsistencies. For example, Figure 2 of Griffiths et al. (2003) shows a construct called "Individualized Knowledge: Implicit," which

is cross-referenced to Proposition 2 (P2) in the paper. Based on the label given to this construct, one might expect it refers to the *level* or *amount* of implicit knowledge that a member of a virtual team possesses. The focus of Proposition 2, however, is on the *extent to which implicit knowledge can be transferred to explicit knowledge*. These are not the same constructs, even though the words "implicit knowledge" are used in both. Furthermore, the "assessment" (operationalization) of the "Individual Knowledge Types: Implicit" construct in Table 1 of Griffith et al.'s (2003) paper does not pertain to the extent to which implicit knowledge can be *transferred* to explicit knowledge (the construct used in P2). Rather, it refers to the "extent to which individuals rely on … knowledge which could be codified but has been made automatic by practice" (p. 281).

A similar problem exists with a number of other constructs – that is, the meaning that, at first glance, might be assigned to a construct shown in Figure 2 of Griffith et al.'s (2003) paper does not match the construct employed in the propositions (see Table 2 above). Moreover, in some cases, the construct used in the propositions (see Table 2 above) does not match the construct in Table 1 of their paper.

Table 2. Constructs in Griffith et al.'s (2003) Theory of Virtualness and Knowledge in Teams

No.	Class of Things	Attribute in General	Propositions in Griffith et al.	Brief Assessment of Construct Definition
1	Team	Level of Virtualness	P1a, P1b, P3a, P3b, P4a, P4b, P5a, P5b, P9a, P9b, P11a, P13	Defined clearly on pp. 267-268.
2	Team	Likelihood of Transforming Implicit Knowledge into Explicit Knowledge	P1a	Nature of implicit and explicit knowledge defined clearly on pp. 270-271. Likelihood of transformation defined somewhat indirectly.
3	Team	Likelihood of Having Access to Extant Explicit Knowledge	P1b	Nature of explicit knowledge defined clearly on pp. 270-271. Likelihood of having access defined somewhat indirectly.
4	Team	Extent to Which Implicit Knowledge Transferred to Explicit Knowledge	P2	Nature of implicit and explicit knowledge defined clearly on pp. 270-271. Amount transferred defined somewhat indirectly.
5	Team	Extent of Proactive Effort Made to Verbalize Rules, Terminology, and Descriptions	P2	Not defined clearly.
6	Team Member	Amount of Tacit Knowledge Acquired from Co-located Sources Transferred to Team	P3a	Nature of tacit knowledge defined clearly on pp. 270-271. Amount acquired from co-located sources and transferred defined somewhat indirectly (a complex attribute in general).
7	Team Member	Amount of Tacit Knowledge Acquired from Teammates	P3b	Nature of tacit knowledge defined clearly on pp. 270-271. Amount acquired from teammates defined somewhat indirectly.
8	Team	Level of Difficulty in Forming Collective Knowledge	P4a	Defined reasonably clearly on p. 273.
9	Team	Level of Experienced Richness of Communication	P4a	Not defined clearly. Brief comment about communication richness on p. 273.
10	Team	Amount of Collective Knowledge Accessible Via Technological Tools	P4b	Not defined clearly.
11	Team	Likelihood of Enacting an Independent Approach to Tasks	P5a, P5b	Defined reasonably clearly on p. 274.
12	Team	Amount of Shared Understanding of Tasks among Team Members	P5a	Nature of shared understanding defined clearly on p. 274. Amount of shared understanding defined somewhat indirectly.
13	Team	Level of Interdependence of Work.	P5b	Defined reasonably clearly on p. 274.
14	Team	Level of Access to Tools and Structures that Support Highly Interdependent Work	P5b	Not defined clearly.
15	Team	Level of Appropriation of Tools and Structures that Support Highly Interdependent Work	P5b	Not defined clearly.

Table 2. Constructs in Griffith et al.'s (2003) Theory of Virtualness and Knowledge in Teams (cont.)

No.	Class of Things	Attribute in General	Propositions in Griffith et al.	Brief Assessment of Construct Definition
16	Team	Amount of Shared Knowledge	P5b	Not clear whether shared *understanding* and shared *knowledge* are the same constructs. See Propositions 5a and 5b.
17	Team Member*	Level of Transition of Potential Team Knowledge to Usable Knowledge	P6	Nature of potential knowledge defined clearly on p. 275. Nature of usable knowledge defined somewhat indirectly on p. 275. Level of transition defined somewhat indirectly.
18	Team Member	Level of Individual Absorptive Capacity	P6, P7	Defined clearly on p. 275.
19	Team Member	Level of Virtual Work/Teamwork	P7	Nature of virtual work defined clearly on pp. 267-268. Not clear whether virtual *work* and virtual *teamwork* (both used in P7) are the same constructs.
20	Team Member	Level of Social Interaction of Team Members	P7	Nature of social interaction defined somewhat indirectly on p. 275. Level of social interaction defined somewhat indirectly.
21	Team	Level of Transition of Potential Knowledge to Usable Knowledge	P8, P12	Potential team knowledge and usable team knowledge defined on pp. 269-270. Level of transition between two types of knowledge defined somewhat indirectly.
22	Team Member	Level of Connections to Relevant Communities of Practice	P8	Defined somewhat indirectly on p. 276.
23	Team	Level of Access to Communities of Practice	P9a	Not clear whether construct 22 and this construct are the same. Do "connections" and "access" have the same meaning? Also, this construct applies to a team whereas construct 22 applies to a team member.
24	Team	Level of Tacit Knowledge from Members' Links to Communities of Practice Disseminated within Team	P9b	Defined somewhat indirectly on p. 276.
25	Team	Level of Transfer of Potential Knowledge to Usable Knowledge	P10	Potentially the same construct as construct 21. Do "transfer" and "transition" have the same meaning?
26	Team	Level of Transactive Memory	P10	Nature of transactive memory defined clearly on p. 277. Level of transactive memory defined somewhat indirectly.
27	Team	Level of Transactive Memory Development	P11a, P11b	Nature of transactive memory development defined clearly on p. 277. Level of transactive memory development defined somewhat indirectly.
28	Team	Level of Virtual Work	P11b	Nature of virtual work defined clearly on pp. 267-268. Level of virtual work defined somewhat indirectly. Note that the level of virtual work seems to be an attribute of both team (P11b) and team member (P7) (see also line 19 in this table).
29	Team	Extent to Which Technologies of Organizational Systems are Used to Support Transactive Memory Development	P11b	Not defined clearly.
30	Team	Level of Synergy	P12, P13	Nature of synergy defined clearly on p. 278. Level of synergy defined somewhat indirectly.
31	Team	Degree of Match Between Team Task and Technology Use	P13	Defined somewhat indirectly on p. 278.

Note: "*" means it is unclear whether attribute in general belongs to the class "team" or "team member."

A second problem with the paper's articulation of constructs (which is to some extent a corollary of the first problem) is that some are defined rigorously (e.g., the level of team virtualness and "individual knowledge types") but others are not. Moreover, the meaning of some constructs has to be elicited from the text used to articulate and support the propositions. Sometimes the meaning of these constructs is clear; sometimes it is not.

For example, a construct used in Proposition 4b of the paper is "Level of Collective Knowledge Accessible Via Technological Tools." Earlier in the paper (p. 273), collective knowledge is defined reasonably precisely as "explicit knowledge that has been internalized by the team members." What is meant by "technological tools," however, is discussed only somewhat obliquely. Nonetheless, it is not clear which of the following meanings should be ascribed to the construct used in Proposition 4b of the paper: (a) the *nature* of the collective knowledge formed by more-virtual teams means this knowledge is easier to access via "technological tools," or (b) more-virtual teams have more access to or greater facility with "technological tools" and, thus, find it easier to access collective knowledge, or (c) both meanings apply to the construct. If the theory is to be rigorously operationalized and tested, the meaning of the construct must be clarified.

Table 2 above contains my assessment of the extent to which each construct used in Griffith et al.'s (2003) paper has been defined rigorously. In part, some of the problems faced by Griffith et al. reflect more general problems faced by scholars who work in the knowledge-management area. For instance, whereas the meaning of constructs such as "amount" are straightforward in some domains (e.g., the amount of a good produced), they are problematic in the knowledge management domain. For example, what precisely is meant by the "*amount* of inflow of knowledge from both peer and supervising units" (Griffith et al., 2003, p. 274, my emphasis)?

One approach Griffith et al.'s (2003) paper might have used to clarify the meaning of *all* constructs is to employ a table similar to Table 2 above. Such a table could have shown the *class of things* that underlie each construct (team or team member) and the *attributes in general* associated with each class of things. To the extent possible, the table also could have provided a rigorous definition of the construct. Where a rigorous definition for a construct was difficult to provide, the table could have been used to indicate that further work was needed to refine the meaning of the construct. Other scholars could then focus their work on articulating these constructs more clearly.

The term "boundary" is not used within Griffith et al.'s (2003) paper. Nonetheless, at one point the paper indicates the theory is not applicable to all kinds of virtual teams: "[t]his model is presented from the perspective of virtual teams where membership is relatively stable, but with members having interaction both within the focal team, as well as with co-located others" (p. 269). Via this statement, therefore, Griffith et al. are seeking to be specific about the class of things that their theory covers.

Use of the evaluation framework and criteria motivates considerations of whether the class of things covered by Griffith et al.'s theory needs to be defined more precisely. For example: Does the theory apply to all kinds of tasks a virtual team with a relatively stable membership might undertake? Does it apply when the virtual team is made of members having substantial differences in culture? Does it hold throughout all phases of the virtual team's existence? Griffith et al.'s (2003) paper is silent on such questions.

In the absence of Griffith et al.'s (2003) paper having defined all constructs precisely, it is difficult to test their theory empirically. The reason is that valid and reliable measures cannot be devised for constructs that are not defined rigorously. Table 1 of their paper (p. 281) shows a number of constructs for which "[m]easures have to be developed," but valid and reliable measures cannot be developed unless the meaning of each construct is clear.

5.1.2. Associations

Griffith et al.'s (2003) paper states 19 propositions. Nine manifest a single directional association between two constructs (five positive associations and four negative associations). Two (P5a and P7) manifest two *mediated* associations involving three constructs (one construct is associated with another construct that, in turn, is associated with another construct). Eight manifest *moderated* associations – in other words, the "strength" of the directional association between two constructs is moderated by a third construct. From an ontological perspective, all types of associations in Griffith et al.'s (2003) theory manifest a hypothesized *interaction* between the constructs involved in the association. In other words, they are postulating that the history of at least one of the constructs in the association is *not* independent of the history of the other construct(s) in the association.

The paper's use of moderated associations strengthens the potential predictive and explanatory power of the theory that has been articulated. Moreover, while the paper does not use the terms "cause" and "causal" when discussing the propositions, nonetheless, causality is implied in the arguments provided to support many of the propositions. For example, it seems clear the authors believe the existence of virtuality in a team *causes* certain outcomes to occur in relation to how different types of knowledge are transferred among team members (but perhaps they have cautiously avoided use of causality terminology). To the extent the propositions imply causality either implicitly or explicitly, the predictive and explanatory power of their theory is enhanced further.

Some of the arguments provided in Griffith et al.'s (2003) paper to support a number of their associations are rigorous and compelling. For instance, their Proposition 1a states (p. 271): "More virtual teams are more likely to transform implicit knowledge into explicit knowledge than less virtual teams." The constructs of "level of team virtualness," "implicit knowledge," and "explicit knowledge" are first defined carefully in the paper (pp. 267-271). Griffith et al. then provide compelling arguments (p. 271) to support an association among these constructs – for example, "[t]eams who spend less time together on task, are located further apart, and who make greater use of technological tools…will be more likely to transfer knowledge in explicit…forms because the technology supports the declarative nature of explicit knowledge." They also cite earlier research to support their argument.

Two factors undermine the rigor of arguments provided by Griffith et al. (2003), however, to support other associations. First, as discussed above, some constructs are not defined clearly. As a result, the meaning of any associations that employ these constructs will lack clarity in some respects. Second, because of the large number of constructs and associations employed in the theory, it is difficult to provide rigorous argumentation to support them all. Inevitably, some associations will be better argued than others.

For instance, Proposition 5b in Griffith et al.'s (2003) paper states: "Access to and appropriation of tools and structures that support highly interdependent work will moderate this result on shared knowledge." By "this result," Griffith et al. mean their previous proposition (Proposition 5a), which states: "More virtual teams have a greater likelihood of enacting an independent approach to their tasks and, therefore, are expected to have less shared understanding of their tasks than less virtual teams."

In my view, the constructs in Proposition 5b called "Level of Access to Tools and Structures that Support Highly Interdependent Work" and "Level of Appropriation of Tools and Structures that Support Highly Interdependent Work" are not defined rigorously (see Table 2 above). What exactly is meant by "access"? What exactly is meant by "appropriation"? Are there trade-offs between "access" and "appropriation"? For example, what happens when teams have *high levels of access* to tools that support interdependent work but *low levels of appropriation* of these tools? Similarly, what happens when teams have *low levels of access* to tools that support interdependent work but *high levels of appropriation* of the tools that they can access? How is the extent to which tools support interdependent work to be assessed? Is this an attribute in particular of a tool that exists independently of humans? Or is this a "socially constructed" attribute in particular of a tool? Are shared understanding and shared knowledge the same constructs (see Table 2 above)? As the nuances in possible meaning of these constructs are teased out, the need for more careful argumentation to support both Propositions 5a and 5b in Griffith et al. (2003) becomes apparent.

Because the paper does not specify all associations in the theory rigorously, a researcher's ability to test the theory empirically is undermined. Researchers will lack the understanding they need to be able to evaluate whether the theory's 19 propositions hold empirically when they observe the outcomes of a test of the theory.

Moreover, the associations that underpin the nine directional and two mediated propositions in the theory are subject to the paradox articulated by Meehl (1967) – namely, *stronger* research designs will yield *weaker* tests of these associations (recall, the null hypothesis might be rejected because any covariation detected by stronger empirical tests might reflect ambient noise in the data or an association that is statistically significant but not *practically* significant). Ideally, Griffith et al. (2003) would have stated these associations in functional form or provided values that they deem to be *practically* significant in terms of

each association's parameters (Edwards & Berry, 2010). However, the eight moderated associations in Griffith et al.'s (2003) theory provide more theoretical precision than their directional associations and, thus, allow stronger tests of their theory to be undertaken (Edwards & Berry, 2010).

5.1.3. States

Griffith et al.'s (2003) paper contains only a limited and somewhat indirect description of those states that fall inside the boundary of their theory and those states that fall outside the boundary of their theory. For instance, in their discussion of the "virtualness" construct, they indicate that pure face-to-face teams are unlikely to occur in today's technological environment and that pure virtual teams (those that never meet face-to-face) differ in a "non-linear way" from teams where their members do meet (if only occasionally) (p. 268). Thus, their theory covers states that apply to "hybrid teams" only – those teams that do not have extreme values for the virtualness construct.

If, as I have argued, Griffith et al.'s theory has 31 constructs (see Table 2), the state space that arises from the Cartesian product of the different sets of values that each of these constructs might assume will be very large (even with a restricted range of values for the "virtualness" construct). It is understandable, therefore, that Griffith et al. would have had difficulty crafting a narrative to describe precisely the subset of the conceivable state space covered by their theory.

5.1.4. Events

As with states, Griffith et al.'s (2003) paper contains only a limited and somewhat indirect description of those events that are inside the boundary of their theory and those events that are outside the boundary of their theory. For instance, because their theory covers only a restricted range of values for the "virtualness" construct, the events that their theory covers will be limited to those associated with this restricted range of values for this construct (manifested in the before-state and after-state pairs that are used to describe an event).

Nonetheless, the large conceivable state space associated with Griffith et al.'s theory leads to a large conceivable event space. As a result, it is understandable that Griffith et al. will have had difficulty crafting a narrative to describe the subset of the conceivable event space covered by their theory.

5.2. Whole

In this subsection, I evaluate the theory proposed in Griffith et al.'s (2003) paper as a whole. The evaluation of the theory's emergent attributes is more of an exercise in judgment than the evaluation of its parts.

5.2.1. Importance

The introduction of Griffith et al.'s (2003) paper provides some clear and compelling reasons why the theory's domain phenomena are important for practice. The paper points out that the management of teams and knowledge are important ways of creating "synergies in … resources" and "increased value" for organizations (p. 266). Moreover, with the emergence and ongoing refinement and development of collaboration technologies and the increasing globalization of workforces, virtual teams are becoming more prevalent. Thus, the successful operation of virtual teams is now critical to the success of many organizations (e.g., Lowry, Zhang, Zhou, & Fu, 2010).

From a research perspective, the paper argues the theory proposed potentially provides a foundation for other researchers who wish "to identify the limiting conditions for effective learning and knowledge transfer across the range of traditional, hybrid, and virtual teams" (p. 280). This outcome clearly has been achieved, because Google Scholar™ shows Griffith et al.'s (2003) paper has been cited more than 300 times. Many researchers have, thus, found the paper useful in their work.

5.2.2. Novelty

Prima facie, it does not appear Griffith et al.'s (2003) paper has been paradigm changing in the sense it has fundamentally altered the ways researchers view phenomena associated with virtual teams and knowledge transfer. Thus, the paper follows a normal-science approach (Kuhn 1996). Nonetheless, the theory described in the paper can be deemed novel for several other reasons.

First, at the time Griffith et al.'s (2003) paper was published, the theory proposed included a number of constructs that, if not completely new, had received only cursory attention in the extant research literature. For example, Table 1 of the paper (p. 281) shows several constructs that require "measures to be developed." Table 1 of the paper also contains a number of constructs that, to the best of my knowledge, had not been canvassed extensively by researchers (e.g., "level of social interaction limited by virtual work undertaken").

Second, the paper includes a number of associations that had received either no or only cursory attention in the research literature that existed at the time the paper was prepared. For example, based on the paper's analysis of existing literature, the eight moderated associations proposed in the theory appear to be new.

Third, the "package" of constructs and associations included in the theory was novel. While at the time the paper was published other researchers might have canvassed *subsets* of the constructs and associations covered by Griffith et al.'s (2003) theory, the "whole" was new. The theory covered team-virtualness and knowledge-transfer phenomena in novel, interesting, and important ways.

In the context of Locke and Golden-Biddle's (1997) two strategies for demonstrating the contribution to knowledge of a piece of research, Griffith et al. (2003) first *construct intertextual coherence* using Locke and Golden-Biddle's tactic of "synthesized coherence" – making connections between literatures that historically have been somewhat disjointed (Locke & Golden-Biddle, 1997, pp. 1030-1035). Their paper enacts Locke and Golden-Biddle's second strategy, *problematizing the existing literature*, by using the tactic of "incompleteness" – that is, showing the existing literature can be characterized by knowledge gaps or lacunae (Locke & Golden-Biddle, 1997, pp. 1030-1035).

These tactics are manifested in the way the paper frames its contribution: "The model is largely drawn from the extant literature….Our contribution is in combining the results from the prior literature in a way that is amenable to an assessment of the opportunities and challenges presented by considering more and less virtual teams from the perspective of knowledge" (p. 270). In this light, Griffith et al.'s (2003) paper has tacitly followed Locke and Golden-Biddle's recommendations for demonstrating novelty and contribution via the rhetoric used to contextualize a piece of research. As a result, overall, the paper's rhetoric is engaging and compelling.

5.2.3. Parsimony

As I have indicated above, I believe the theory proposed in Griffith et al.'s (2003) paper contains:

- 31 constructs (rather than 17 constructs, as shown in Figure 2 of the paper, or 15 constructs, as shown in Table 1 of the paper);
- 21 associations (these are manifested in 19 propositions in the paper).

Based on a simple count of the number of constructs and associations in the theory, it is difficult to conclude it is parsimonious. As a result, one might predict this lack of parsimony would undermine the theory's impact on other researchers. Interestingly, as indicated above, citation data suggests otherwise. Given the large number of citations to Griffith et al.'s (2003) paper, the theory proposed in the paper clearly has engaged the interest of other researchers. Thus, contrary to expectations, lack of parsimony has not weakened its impact.

Nonetheless, the large number of constructs and associations in Griffith et al.'s (2003) theory is likely to undermine any attempts made to "prune" it – that is, to reduce the number of constructs and associations and to articulate its inside-boundary states and inside-boundary events more precisely (Leavitt et al., 2010). For instance, what are the implications for the theory if an empirical test shows lack of support for one (or some subset) of its associations? Does "one null result or even a few null results…really justify moving backwards in the logical chain to argue the theory is disconfirmed"? (Leavitt et al., 2010, p. 646). Moreover, with so many constructs and associations, it will be difficult to find a comparable theory against which the predictive and explanatory power of the theory can be evaluated (Leavitt et al., 2010, pp. 649-654).

5.2.4. Level

In my view, Griffith et al.'s (2003) paper has articulated a middle-range (meso) theory. The range of phenomena the theory covers is reasonably broad. Thus, the authors cannot be accused of narrow empiricism. Moreover, while a number of the constructs have yet to be defined rigorously and to be operationalized, it is possible to conceive how ultimately these outcomes might be achieved. In short, the theory is framed at a level that enables it to be employed to generate useful predictions about, insights about, and understanding of the theory's focal phenomena.

5.2.5. Falsifiability

I have argued above that some parts of the theory proposed in Griffith et al.'s (2003) paper have been articulated clearly and other parts have *not* been articulated clearly. Where clarity exists, rigorous empirical tests can be undertaken to test the theory. Potentially, the outcomes of such tests will lead researchers to conclude the theory is not supported (i.e., the theory can be falsified). For those parts of the theory that are not articulated clearly, however, attempts to falsify it are problematic. Empirical tests that produce "unfavorable" outcomes may simply mean researchers have used invalid or unreliable measures of constructs. Alternatively, they have failed to understand the nature of an association between constructs. They also might have tested the theory in a context that falls outside its boundary.

5.3. Summary Evaluation

Table 3 provides a summary of the more-detailed evaluation I have carried out above of Griffith et al.'s (2003) paper.

Table 3. Summary Evaluation of Griffith et al.'s (2003) Paper

Criterion	Summary Evaluation
Parts	
Constructs	Some constructs are defined precisely; others are not. Inconsistencies exist among the definitions of some constructs at different places in the paper. Little discussion occurs on the boundary of the theory in terms of the class of classes of things it covers. A brief comment is made about the theory applying to virtual teams whose membership is relatively stable and whose members have both remote and local interactions.
Associations	Use of moderated associations strengthens the potential predictive and explanatory power of the paper. Some arguments used to support associations are rigorous and compelling. Others are not clear and compelling, because they include constructs that are not defined precisely. Moreover, the large number of constructs and associations employed makes rigorous articulation of all associations difficult to accomplish.
States	Little discussion occurs on those states that are inside the boundary of the theory and those that are outside the boundary of the theory.
Events	Little discussion occurs on those events that are inside the boundary of the theory and those that are outside the boundary of the theory.
Whole	
Importance	The paper provides clear and compelling reasons why the theory is important for practice. The paper is highly cited, which manifests its substantial impact on other researchers.
Novelty	The paper introduces new constructs and associations. It also covers team-virtualness and knowledge-transfer phenomena in novel, interesting, and important ways. The theory covers gaps or lacunae in the existing literature.
Parsimony	The theory is not parsimonious, because it has a large number of constructs and associations and also covers many possible states and events.
Level	The theory is framed at an appropriate level as a middle-range (meso) theory.
Falsifiability	Those parts of the theory that have been articulated precisely can be subjected to rigorous empirical tests. Thus, these parts potentially can be falsified. Those parts of the theory that have not been articulated rigorously, however, are difficult to test empirically. Thus, the outcomes of tests of these parts of the theory are problematic as a means of supporting or falsifying the theory.

6. Using the Framework and Criteria to Inform Theory Refinement and Construction

The framework and criteria I have proposed can also be used to inform researchers seeking to refine an existing theory or to build a new theory. As they construct their modified or new theory, researchers should be mindful of matters they need to address from the perspective of achieving high-quality outcomes in relation to the parts and whole of their theory. In essence, the framework and criteria can be used to test the quality of the work they are undertaking as their refined or new theory unfolds.

If researchers are seeking to *refine* an existing theory, the framework and criteria can first be used to pinpoint areas where the existing theory can be improved. Researchers' analyses might indicate some constructs are not well defined, some associations are not articulated clearly, the inside-boundary states and events have not been specified clearly, the theory lacks parsimony, and the importance of the theory is not well argued. Researchers can then seek to rectify the problems they identify.

For instance, my analysis of Griffith et al. (2003) using the framework highlights some areas where their theory might be improved. A number of their constructs and associations need to be defined more rigorously, and the inside-boundary states and events of their theory need to be articulated more clearly. Once these outcomes are achieved, it might be possible to see more clearly which constructs and associations should be discarded from the theory because they are deemed not to be material in explaining and predicting variations in the theory's focal phenomena (usable team knowledge). As a result, a more parsimonious version of the theory might be achieved. Moreover, a more rigorous specification of constructs, associations, and inside-boundary states and events would facilitate more rigorous empirical tests of the theory and, thus, an enhanced ability to falsify the theory.

If certain types of defects become apparent, however, researchers need to reflect carefully on the merits of seeking to enhance the theory. For example, if the theory's importance or novelty are unclear, or it has been framed at an inappropriate level, researchers should be circumspect about whether actions -- such as defining constructs more rigorously, providing a clearer rationale for associations, and articulating inside-boundary states and events more completely -- will produce substantial improvements in the quality of the theory.

For example, my analysis of Griffith et al.'s (2003) theory using the framework I have proposed shows that it rates highly in terms of importance, novelty, and level. In this light, *prima facie* good reasons exist to try to refine or enhance the theory rather than build yet another theory of virtual teams and knowledge-transfer phenomena (Gray & Cooper, 2010; Hambrick, 2007). My analysis using the framework pinpoints some ways in which an enhanced theory might be achieved.

If researchers are seeking to articulate a new theory, their first concern should be the choice of the focal phenomena. They must select focal phenomena that their colleagues ultimately will deem to be important, either because the focal phenomena's importance is readily apparent, or the rhetoric the researchers provide to support the importance of the focal phenomena is compelling. Sometimes a key issue is to reframe well-known phenomena in new, interesting ways or to point out important phenomena that previously had gone unseen (Weber, 2003a). The focal phenomena must also be conceived at a level that allows a meso-level theory to be formulated.

Once the focal phenomena are defined clearly, researchers can then build the parts of the theory – constructs, associations, inside-boundary states, and inside-boundary events. They can seek to ensure the theory is falsifiable through clear, precise definitions or specification of the theory's parts. Through selective choice of constructs, associations, and inside-boundary states and events, they can also seek to ensure the theory is parsimonious. They must then realistically evaluate their theory's "utility" and "originality" (Corley & Gioia, 2011). If they deem it to be useful and novel, they must carefully craft their rhetoric with the objective of convincing their colleagues that their theory has these characteristics (Locke & Golden-Biddle, 1997).

7. Conclusions

The framework I have proposed above facilitates an evaluation of the quality of an existing theory. It also informs researchers who are seeking to build a new theory or refine an existing theory. As they construct their new or modified theory, they should be mindful of matters they need to address from the perspective of achieving high-quality outcomes in relation to their theory's parts and whole. In essence, the framework and criteria can be used as a set of checkpoints to test the quality of the work they are undertaking.

The framework does not assist, however, in choosing the focal phenomena and the ways these phenomena might be conceived, nor does it assist in choosing a theory's constructs, associations, and inside-boundary states and events. To a large extent, these choices remain creative acts that affect, in particular, the quality of the whole – a theory's importance, its novelty, its parsimony, and so on (Feyerabend, 1975; Jaccard & Jacoby, 2010, pp. 39-73; Weick, 1989). In the information systems discipline (and in a number of other disciplines), I believe a rich vein of research lies in seeking to better understand the characteristics of those choices that lead to the articulation of high-quality, high-impact theories.

Acknowledgements

An early version of this paper was presented at the *Information Systems Foundations Workshop* at The Australian National University, Canberra, 30 September - 1 October 2010. I am indebted to participants in the workshop for their comments on the paper. I am also indebted to (a) Terri Griffith for her helpful feedback on my analysis of her paper, (b) Suzanne Rivard, Bob Zmud, and another reviewer for their constructive review comments, and (c) Cynthia Beath for her counsel, feedback, and support during multiple iterations of the paper. Of course, these colleagues bear no responsibility for any errors, omissions, or views expressed in the paper.

References

Abbott, A. (1990). A primer on sequence methods. *Organization Science*, *1*(4), 375-392.

Bacharach, S. B. (1989). Organizational theories: Some criteria for evaluation. *Academy of Management Review*, *14*(4), 496-515.

Blalock, H. M. (1971). *Causal models in the social sciences*. Chicago, IL: Aldine-Atherton.

Boudon, R. (1991). Review: What middle-range theories are. *Contemporary Sociology*, *20*(4), 519-522.

Bunge, M. (1977). *Treatise on basic philosophy: Volume 3: Ontology I: The furniture of the world*. Dordrecht, Holland: D. Reidel Publishing Company.

Bunge, M. (1979). *Treatise on basic philosophy: Volume 4: Ontology II: A world of systems*. Dordrecht, Holland: D. Reidel Publishing Company.

Cadiz, D., Sawyer, J. E., & Griffith, T. L. (2009). Developing and validating field measurement scales for absorptive capacity and experienced community of practice. *Educational and Psychological Measurement, 69*(6), 1035-1058.

Colquitt, J. A., & Zapata-Phelan, C. P. (2007). Trends in theory building and theory testing: A five-decade study of the Academy of Management Journal. *Academy of Management Journal*, *50*(6), 1281-1303.

Corley, K. G., & Gioia, D. A. (2011). Building theory about theory building: What constitutes a theoretical contribution? *Academy of Management Review*, *36*(1), 12-32.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *1*3(3), 319-340.

Davis, G. F. (2010). Do theories of organizations progress? *Organizational Research Methods*, *13*(4), 690-709.

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, *48*(2), 147-160.

Doty, D. H., & Glick, W. H. (1994). Typologies as a unique form of theory building: Toward improved understanding and modeling. *Academy of Management Review*, *19*(2), 230-251.

Dubin, R. (1978). *Theory building* (Rev. ed). London: Free Press.

Edwards, J. R., & Berry, J. W. (2010). The presence of something or the absence of nothing: Increasing theoretical precision in management research. *Organizational Research Methods*, *13*(4), 668-689.

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Journal*, *14*(4), 532-550.

Fetzer, J. H. (1993). *Philosophy of Science*. New York: Paragon House.

Feyerabend, P. K. (1975). *Against method: Outline of an anarchistic theory of knowledge*. Atlantic Highlands, N.J.: Humanities Press.

Freese, L. (1980). Formal theorizing. *Annual Review of Sociology*, *6*, 187-212.

Godfrey-Smith, P. (2003). *Theory and reality: An introduction to the philosophy of science*. Chicago: University of Chicago Press.

Gray, P. H., & Cooper, W. H. (2010). Pursuing failure. *Organizational Research Methods*, *13*(4), 620-643.

Gregor, S. (2006). The nature of theory in information systems. *MIS Quarterly*, *30*(3), 611-642.

Griffith, T. L., & Sawyer, J. E. (2010). Multilevel knowledge and team performance. *Journal of Organizational Behavior, 31*(7), 1003-1031.

Griffith, T. L., Sawyer, J. E., & Neale, M. A. (2003). Virtualness and knowledge in teams: Managing the love triangle of organizations, individuals, and information technology. *MIS Quarterly*, *27*(2), 265-287.

Grover, V., Lyytinen, K., Srinivasan, A., & Tan, B. C. Y. (2008). Contributing to rigorous and forward thinking explanatory theory. *Journal of the Association for Information Systems*, *9*(2), 40-47.

Hambrick, D. C. (2007). The field of management's devotion to theory: Too much of a good thing? *Academy of Management Journal*, *50*(6), 1346-1352.

Hempel, C. G. (1966). *Philosophy of natural science*. Englewood Cliffs, N.J.: Prentice-Hall.

Hovorka, D. S., & Lee, A. S. (2010). Reframing interpretivism and positivism as understanding and explanation: Consequences for information systems research. *Proceedings of the 2010 International Conference on Information Systems*, St. Louis, MO.

Jaccard, J., & Jacoby, J. (2010). *Theory construction and model-building skills: A practical guide for social scientists*. New York, NY: Guilford Press.
Klein, H. K., & Myers, M. D. (1999). A set of principles for conducting and evaluating interpretive field studies in information systems. *MIS Quarterly*, *23*(1), 67-93.
Kuhn, T. S. (1996). *The structure of scientific revolutions* (3rd ed.). Chicago, IL: University of Chicago Press.
Lamb, R., & Kling, R. (2003). Reconceptualizing users as social actors in information systems research. *MIS Quarterly*, *27*(2), 197-235.
Langley, A. (1999). Strategies for theorizing from process data. *Academy of Management Review*, *24*(4), 691-710.
Leavitt, K., Mitchell, T. R., & Peterson, J. (2010). Theory pruning: Strategies to reduce our dense theoretical landscape. *Organizational Research Methods*, *13*(4), 644-667.
Lewin, K. (1945). The research center for group dynamics at Massachusetts Institute of Technology. *Sociometry*, *8*(2), 126-136.
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.
Lowry, P. B., Zhang, D., Zhou, L, & Fu, X. (2010). Effects of culture, social presence, and group composition on trust in technology-supported decision-making groups. *Information Systems Journal*, *20*(3), 297-315.
Markus, M. L., & Robey, D. (1988). Information technology and organizational change: Causal structure in theory and research. *Management Science*, *34*(5), 583-598.
Meehl, P. E. (1967). Theory-testing in psychology and physics: A methodological paradox. *Philosophy of Science*, *34*(2), 103-115.
Merton, R. K. (1957). *Social theory and social structure* (Rev. ed.). Glencoe, IL.: Free Press.
Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81-97.
Mone, M. A., & McKinley, W. (1993). The uniqueness value and its consequences for organization studies. *Journal of Management Enquiry*, *2*(3), 284-296.
Monge, P. R. (1990). Theoretical and analytical issues in studying organizational processes. *Organization Science*, *1*(4), 406-430.
Parsons, T. (1950). The prospects of sociological theory. *American Sociological Review*, *15*(1), 3-16.
Pentland, B. T. (1999). Building process theory with narrative: From description to explanation. *Academy of Management Review*, *24*(4), 711-724.
Poole, M. S., Van de Ven, A. H., Dooley, K., & Holmes, M. E. (2000). Process theories and narrative explanation. In M. S. Poole, A. H. Van de Ven, K. Dooley, & M. E. Holmes (Eds.), *Organizational change and innovation processes: Theory and methods for research* (pp. 29-55). Oxford, England: Oxford University Press.
Popper, K. (2005). *The logic of scientific discovery.* London: Taylor & Francis e-Library.
Sambamurthy, V., Bharadwaj, A, & Grover, V. (2003). Shaping agility through digital options: Reconceptualizing the role of information technology in contemporary firms. *MIS Quarterly*, *27*(2), 237-263.
Shapira, Z. (2011). "I've got a theory paper–do you?": Conceptual, empirical, and theoretical contributions to knowledge in the organizational sciences. *Organization Science*, *22*(5), 1312-1321.
Sutton, R. I. & Staw, B. M. (1995). What theory is not. *Administrative Science Quarterly*, *40*(3), 371-384.
Van de Ven, A. (1989). Nothing is quite so practical as a good theory. *Academy of Management Review*, *14*(4), 486-489.
Weber, R. (2003a). The problem of the problem. *MIS Quarterly*, *27*(1), iii-ix.
Weber, R. (2003b). Theoretically speaking. *MIS Quarterly*, *27*(3), iii-xi.
Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly, 26*(2), xiii-xxiii.
Weick, K. E. (1989). Theory construction as disciplined imagination. *Academy of Management Review*, *14*(4), 516-531.

Weick, K. E. (1995). What theory is not, theorizing is. *Administrative Science Quarterly*, *40*(3), 385-390.

Weick, K. E. (1999). Theory construction as disciplined reflexivity: Tradeoffs in the 90s. *Academy of Management Review*, *24*(4), 797-806.

Whetten, D. A. (1989). What constitutes a theoretical contribution? *Academy of Management Review*, *14*(4), 490-495.

Zmud, R. (1998). Editor's comments. *MIS Quarterly*, *22*(2), xxix-xxxi.

Zmud, R. (2003). Special issue on redefining the organizational roles of information technology in the information age. *MIS Quarterly*, *27*(2), 195.

About the Author

Ron WEBER is Dean, Faculty of Information Technology, Monash University. He obtained his PhD from the University of Minnesota. His research interests focus primarily on conceptual modeling. Ron is a past president of the Association for Information Systems, a past co-chair of the International Conference on Information Systems, and a past editor-in-chief of the *MIS Quarterly*. He is a Fellow of the Association for Information Systems, the Australian Computer Society, the Institute of Chartered Accountants in Australia, CPA Australia, and the Academy of the Social Sciences of Australia.