ABSTRACT. I attempt to demonstrate that the concept of institutional fit and the closely related approach of institutional diagnosis can be improved with the process of formalization. In this context, the concept of fit is interpreted as a way of expressing certain theoretical propositions that relate a set of variables with each other and with an outcome. This perspective is demonstrated through the use of the Web Ontology Language to express several "theories of fit." Using a formal language to describe types of fit and their associated theories is argued to have much potential for advancing the scientific study of social-ecological systems.

Key Words: diagnostics; formalization; institutional fit; social-ecological systems

INTRODUCTION

The prescriptive concept of institutional fit, or that institutional arrangements should match “the defining features of the problems they address” (Young 2008:20) has intuitive appeal for scholars studying human-environment interactions. Institutional fit is closely related to the process of diagnostic analysis, whereby attributes of a problem are examined in order to identify the governance arrangements that might best address them (Young 2002, 2008, 2010). A key idea supporting both a diagnostic approach and the concept of institutional fit is that different environmental problems should be treated differently, while similar problems should be treated similarly. For example, environmental problems characterized by unpredictable dynamics with positive feedbacks that create high levels of irreversibility should be managed with an emphasis on early warning systems and high levels of institutional adaptive capacity (Young 2002).

In spite of its intuitive appeal, the concept of fit is limited by being a vague concept. Many questions are left unanswered and must be addressed by the analyst (Cox 2008). For example, what are the defining features to which institutions should be matched, and on what typology of institutional arrangements might we rely to conduct a systematic process of such matching? This grants scholars a fair amount of leeway in how they operationalize it, and limits its utility in meaningful academic communications because applications cannot easily be compared.

I believe that the existence of a fit between two objects can be usefully thought to mean that they relate in such a manner so as to produce a desirable outcome. Conversely, a misfit would relate the combination of such objects to an undesirable outcome. From here, we can introduce the concept of a “theory of fit” as a type of theory, where theories “place values on some of the variables identified as important in a framework, posit relationships among the variables, and make predictions about likely outcomes” (Schlager 2007:296). This interpretation of fit also means that there really is no such thing as an overall “fit” of a system, just as there could not be just one theory. A system could be fit with respect to one outcome and not to another.

A theory of fit, as a special type of theory, has several distinguishing characteristics. First, the outcome or dependent variable involved must be highly normative: it must be something that is explicitly desired. While this is a feature that is shared with many theories, it is a necessary component of a theory of fit. Secondly, there must be at least two independent variables, at least one of which describes a feature of a problem or of the natural environment, and at least one of which describes a social/technical property that may fit well or poorly with this feature.

Thirdly, these independent variables cannot independently contribute to the outcome. It is almost a mantra in some fields and ways of thinking that everything is linked to everything else. But this perspective differs from much of the work in social science that seeks to isolate the unique effects of each independent variable included in a model. Theories emphasizing causal interactions, then, represent a balance between the hyper-connected and hyper-isolated perspectives. They can be thought of as clusters or modules of causality (Young 2011), the variables of which interact more strongly with each other than they do with other variables to affect an outcome.

This association among the independent variables also implies that theories of fit are expressed at a sufficiently high level of specificity to warrant the description of such interactions. Theories that use highly aggregated general variables, such as one that stipulates that resource monitoring contributes to sustainable natural resource management, do not require such specificity. It is only when we get more specific that such interactions must be described. For example, certain types of monitoring, done by resource users or by governmental guards, may fit better with subtypes of resources and in concert with other social or ecological conditions. This observation follows a general theme in much of science: as we become...
more specific in the factors we examine, we must consider how these factors interact with each other and the relevant context more carefully.

Finally, we can make some general statements about the kinds of relationships that the variables in a theory of fit have with each other. First, each independent variable will tend not to have an unambiguously positive or negative relationship with the dependent variable. For example, just because a theory stipulates that variable X being high is beneficial in achieving an outcome does not mean that it being low is necessarily a bad thing, depending on what else changes in the system. If this relationship was unambiguously positive or negative, we could simply try to raise or lower the value of the independent variable as needed without worrying about how it fits with the other variables in the system. Secondly, this also affects the types of relationships we can expect (or not expect) between the independent variables. Basically, it seems unlikely that the relationship will be expressible arithmetically (e.g., additive or multiplicative). An additive relationship would imply independence, and a multiplicative relationship would imply that each independent variable is unambiguous in the direction (positive or negative) of its relationship to the dependent variable.

While we know that independent variables interact to produce outcomes, and many individual factors have been found to be important in environmental policy and management (Agrawal 2003), there has been little work done to organize the diversity of theories that relate several factors together. No standard method exists to label or organize them, other than in reference to specific analyses that confirm or disconfirm them. A useful response to this situation may be to develop a formalized way of enumerating and specifying theories of fit. Such formalization would disambiguate the idea of fit by disaggregating it into formally specified subtypes, thereby increasing its usefulness as a scientific concept. What follows in this paper, then, is an exploration of the following argument: formalization of theories of fit can improve the closely related concepts of fit and diagnosis.

**FRAMEWORKS AND THEORIES OF SOCIAL-ECOLOGICAL SYSTEMS**

The literature on the institutional fit of governance arrangements to environmental problems substantially overlaps with the literature on the sustainability of social-ecological systems (SESs) (see Folke et al. 2007). Anderies et al. (2004) define these as “social systems in which some of the interdependent relationships among humans are mediated through interactions with biophysical and nonhuman biological units.” Formally enumerating and specifying theories of fit is a large departure from the standard practice in both of these literatures.

Theories rely on concepts or variables. The study of SESs has not suffered from a lack of definitional offerings of concepts such as resilience and robustness (Holling 1973, Carpenter et al. 2001, Carlson and Doyle 2002, Walker et al. 2004, Levin and Lubchenko 2008). Unfortunately, there has not been much progress on achieving a broad consensus on both the most standard and accepted definitions of key terms, and less on how different terms relate to each other.

The lack of such common definitions in part reflects the fact that there is not a particularly well-established common framework for the study of SESs. Instead, there are many, including Gunderson and Holling’s (2002) Panarchy concept, the learning-based approach advocated by Pahl-Wostl (2009), Ostrom’s (2007) diagnostic framework, Anderies et al.’s (2004) approach, Janssen et al.’s (2006) network approach, and the robust control framework presented by Anderies et al. (2007). The role of a framework is to furnish the concepts that can be operationalized in order to construct theories, which posit causal relationships among these concepts (Ostrom 2005). Thus, without a common framework, conceptual clarification and cumulative theory generation is difficult.

Moreover, these concepts, and the theories that use them to relate various social and ecological variables to each other and to outcomes, tend to be stated textually and informally. The fact that theories of SESs are expressed mostly in natural language has the advantage that, because natural language is complex and fuzzy, it can describe a world that is equally complex and unclear. However, this comes at a cost. In the case of SES research, I would argue that this cost is the preponderance of natural language theories that can be and are interpreted somewhat differently by different scholars based on their own experiences with the terms used and relationships described. Such differences likely inhibit effective scientific communication. Moreover, because the expression of theories in natural language effectively glosses over the methods used to produce such a theory, the specificity of their actual meaning to these methods can be hidden from those who communicate with them.

All this implies that knowledge of SESs probably tends to be highly localized to particular scientists and particular systems. Thus, in addition to making the particular concept of fit more useful, formalizing theories of fit offers the possibility of improving communication among scientists regarding the many theories they employ by clarifying the meaning of theoretical statements.

**FORMALIZATION**

Formalization is the process of taking statements expressed in natural or nonformal language and expressing them in a formal language. A formal language is “a set of primitive expressions (e.g., symbols) and formation rules that specify how complex expressions can be constructed from the primitive ones” (Hinkel 2008:19). Two common examples of formal languages are the Unified Modeling Language and the Web Ontology Language (OWL).
I employ some basic features of OWL to formalize several theories of fit. OWL is built on a data model known as the Resource Description Framework (RDF). RDF represents all data as triples in the following form: subject-predicate-object. The subject is equivalent to a row ID in a traditional spreadsheet table, while the predicate is the column name, and the object is the value of a cell where the row and column intersect. Within the movement known as the Semantic Web, RDF has been put forward as an alternative to more dominant data models, namely relational databases. Being based on RDF, OWL is itself expressed in triples.

Employing RDF and OWL formalizes a domain of knowledge in a particular way. First, it constructs the world as classes of objects, the members of which are called individuals. Classes can have several relationships with each other. The most common is the “is a” relationship, where one class is a subclass of another, which is then called the superclass. If class A is a subclass of class B, and individual C is a member of class A, we can infer that C is a member of class B as well. This demonstrates the point that the formal meaning of any OWL expression amounts to the inferences that can be produced by that expression (Allemang and Hendler 2008). The superclass-to-subclass is a one-to-many relationship, with one superclass having many subclasses. Subclasses can also be given their own subclasses to create a taxonomy.

Each OWL class can have a number of properties assigned to it, and subclasses inherit the properties of their superclasses. For example, if we assign the property “biodiversity” to a class we call “ecosystems”, then every subclass of ecosystem (e.g., terrestrial or aquatic) would also have this property. OWL allows a user to enumerate the possible values that each property can take on as well.

Another very common relationship between classes, although not one formally included in OWL (meaning it does not automatically facilitate inference), is the “is a part of” relationship, where one class is a component of another class. Leaves are a part of trees, for example.

**APPLYING A FORMAL APPROACH TO THE CONCEPT OF FIT**

**The basic pattern**

There are several syntaxes that have been developed to express OWL. Rather than presenting the actual statements that would formalize theories of fit, I will illustrate this formalization through diagrams that are very similar to the graph notation for RDF. Rounded rectangles are classes, regular rectangles are individual members of classes, and arrows indicate relationships. It is important to know that these diagrams are to some extent caricatures of the OWL statements that would be used to implement the actual expressions. The full details of how they would be expressed in OWL can mostly be glossed over in the following discussion.

To begin the formalization, it helps to introduce and formalize one of the frameworks mentioned earlier that will serve as the basis for the expression of several theories of fit. This is the diagnostic framework introduced by Ostrom (2007). The SES framework breaks an SES into several primary components (resource systems, resource units, governance systems, and actors), as shown in Figure 1. In the future, there are plans to relabel the “users” component as “actors” to generalize the class. As a result, in the following discussion I will use this term instead. Each component is associated with a set of concepts, about which one could ask several questions.

**Fig. 1.** The main components of the social-ecological system (SES) framework (Source: Ostrom 2007).
Building on this, formalizing theories of fit with OWL presents a problem because fit is defined by how various SES components and their attributes interact to produce an outcome, which is something we cannot model with a single triple that describes only a binary relationship between the subject and the object. We need to express more than simply “institutions fit environment,” which would be using fit as a relationship between two classes and not as a class in itself. OWL, however, because it is built on the RDF data model, can describe only binary relationships between subjects and objects through predicates.

Fortunately, there is a standard solution to this problem which can be used to model any n-ary relationship (where n > 2). Figure 3 displays the method as applied to the concept of theories of fit. It shows “Theory of fit” as a class instead of a relationship between two classes. There are three types of relationships that we need to use to describe theories of fit. First, the theory of fit class has a “Uses” relationship with at least two subclasses of the “Variable” class (subclasses of which are variables). This relationship connects the theory to the variables that it uses.

Second, the theory of fit class has exactly one “Associates variables through” relationship to a class labeled “Variable association”, which we also introduce. A variable association describes how the variables involved in the theory supposedly interact to produce the outcome of interest. Finally, the theory of fit class has exactly one “Contributes to” relationship to an “Outcome” class. This class describes the outcomes that the independent variables, when associated with each other in a particular way, contribute to producing. The “Outcome” class can also be considered to be a subclass of the “Variable” class.

Examples of theories of fit
Figure 4 uses this formalization to illustrate a theory that has been expressed in much of the environmental governance literature (Hanna and Munasinghe 1995): that sustainable resource management can be encouraged if the extent of a governance system is spatially congruent to the extent of the resource system it governs. This theory is labeled “Extent fit,” which would be a subclass of the class “Theory of fit” from Figure 3.
“Extent fit” has a “Uses” relationship to two variables: “Governance system extent” and “Resource system extent.” The outcome it contributes to is labeled “Sustainable resource management.” This is a particular value of an outcome variable we might call “Resource management sustainability.” Because the variables in the theory contribute to a particular value of this variable, and values are modeled as individual members of variable classes, the rectangle in the figure is not rounded. The situation is similar with “High spatial congruence,” which is a value of a type of variable association that we could label “Spatial congruence.” The OWL statements used to model these two types of relationships (class to class vs. class to individual) are slightly different.\[5\]

The logic behind this theory is twofold. First, without high spatial congruence, spatial externalities could lead to resource degradation when actions taken outside the jurisdiction of the governance system adversely affect the resource system. Second, this congruence may help ensure that a consistent set of institutions are applied to an entire resource system, which presumably can be identified as such because there is a fair degree of biophysical homogeneity within its boundaries that would be amenable to the application of a consistent set of governance arrangements.

This reasoning is the motivation for such schemes as the Yellowstone to Yukon Conservation Initiative. It has been discovered that the ranges of many animals found in Yellowstone National Park extend well up into Canada. As such, in order to conserve these species, the thinking is that a governance system that covers these ranges must be implemented. Otherwise, externalities from outside the boundaries of the governance system can undermine the conservation efforts within it.

A second example of a theory of fit associates enumerated values of variables instead of congruence between variables. The theory here is labeled “Spatiotemporal fit.” The three variables involved are as follows: (1) “Resource unit variability,” a property of a resource unit; (2) “User mobility,” a property of user groups, which are a type of actor; and (3) “Social boundaries”, which again would be a property of a set of user groups.

The theory that combines particular values of each of these variables is expressed in Figure 5, and has been established in previous work (Niamir-Fuller 1998, Janssen et al. 2007). Under conditions of high resource unit variability, user groups need to be highly mobile in order to move to different areas where resources are more available at different times. This is facilitated if there are fuzzy social boundaries, meaning that ad hoc negotiations can be made between different user groups to allow access to resources under certain conditions. The variable association has changed from “Congruence” to “Combination,” meaning that it is the combination of the variable values that contributes to the outcome. If each of these values is present for these variables, it helps produce the outcome indicated.

Niamir-Fuller (1998) examines pastoral systems in the Sahel region of Africa that support this theory. In this area, the natural environment is “characterized by low net primary productivity and high variability in ecosystem structure and productivity, both spatial and temporal” (Niamir-Fuller (1998:257). The pastoral systems present have adopted several management strategies to respond to this, including active environmental monitoring and high levels of user mobility, as well as a system of flexible yet important boundaries establishing which groups have priority access to which areas. Access to a group’s priority area by another group is subject to “flexible negotiations” (Niamir-Fuller (1998:264).

This example also presents a question: Should environmental monitoring be included in this particular theory, since Niamir-Fuller (1998) thought this was an important factor in sustaining...
the systems she observed? A short, although not complete, answer to this question is that a useful criterion for the variables to be included in a theory is that they should each relate quite intimately with the rest, such that the effects of each on the dependent variable change with shifts in the values or presence or absence of the others. Environmental monitoring has been found to be important across a range of systems, not just those where spatiotemporal resource availability is high (Cox et al. 2010). Its efficacy has not been found to depend on the other variables of this theory. Thus, it may be more appropriate to stipulate its importance in its own theory, which may not need to be a theory of fit as I have defined it.

**Nesting and the Taos acequias**

I conclude with one final example that has spatial aspects similar to the two previous examples. Here we are concerned with spatial externalities and with the appropriateness of institutions to local biophysical circumstances, as we were with the first example. In addition, we are also concerned with resource nesting. As demonstrated by Olson et al. (2001) and many others, the distribution of biophysical properties in the environment can be grouped into spatially discrete units, which are in turn nested within each other at several levels. Within each spatially discrete unit, the biophysical properties are relatively homogeneous in comparison to those in other units. With less aggregated, smaller spatial units, this homogeneity becomes stronger. To conceptualize this pattern, Olson et al. (2001) conceive of a nested system of terrestrial ecological units, beginning with ecoregions and aggregating up to spatial realms.

Reflecting on this quality of many biophysical systems, Brunckhorst et al. (2006) specify a theory of environmental management, stating that a governance regime ought to be spatially nested itself in order to mirror the discrete and nested nature of spatial biophysical units. Under such a condition, the two objectives described earlier may be favored: institutions can be made appropriate to locally homogenous biophysical properties, and spatial externalities may be internalized by successive levels of governance. The theory of spatial nesting fit is shown in Figure 6. The variable association is spatial congruence, and the value of this association is high spatial congruence, duplicating those from the spatial extent theory of fit. This indicates that the manner in which governance is spatially nested should reflect, or be congruent with, the natural spatial nesting of the resource system.

An example of an SES that confirms this theory is the acequia irrigation system in the Taos valley of northern New Mexico (Cox 2010). An acequia is a community of farmers who share an irrigation canal. The acequia farmers are the descendants of the Spanish colonists who emigrated from what is now Mexico along the Rio Grande beginning in the 17th century. They carried with them in their heads the traditional Spanish irrigation institutions, and used them to survive in a high desert environment that covers much of northern New Mexico. In Taos valley there is little rain, leaving the farmers mostly dependent on the snowmelt-derived water that flows from the nearby Sangre de Cristo Mountains.

**Fig. 6. Spatial nesting fit.**

![Spatial nesting fit diagram](http://www.ecologyandsociety.org/vol17/iss4/art54/)

There are many important aspects of the Taos valley acequias that have enabled them to survive for several hundred years. One important feature of the acequia governance system is its nested quality, which mirrors the nested quality of the irrigation infrastructure the acequias maintain. Very small irrigation canals are nested below a larger canal from which they each derive their water through irrigation headgates. This larger canal feeds off of a canal that is larger still, or directly off of a river. At each spatial level of this resource system there tends to be a new management unit that governs a larger reach of a canal or river. This pattern has been noted in other traditional irrigation systems (Coward 1977).

Figure 7 illustrates this congruence in a section of Taos valley. Starting at the top of the figure, each successive box shades irrigated land based on the unit of governance that manages this land and the canals that feed it. At the lowest governance level, individual farmers manage their own irrigation turnouts and irrigated plots. Individual-level plots are shown with distinct shades, but the corresponding canals and headgates that feed them are not shown. At the next level, individual acequias manage the main canal from which these individual canals obtain their water. In some acequias there is in fact an intermediary level between these two to better reflect the degree of hydrological nesting. At the final level, the acequias in this section of the valley are united into a single governance community that manages the whole river through a formal repartimiento, or water-sharing agreement.

**FORMALIZATION AND DIAGNOSTICS**

The second part of the argument I introduced in the beginning of this paper is that formalization can help with the process of diagnosing which institutional arrangements and/or technologies might most effectively fit with important aspects
of environmental problems. The process of diagnosis has generally been presented to take the form of a set of questions that are asked of a particular system (Young 2002, Berkes 2007, Ostrom 2007). In conducting a diagnosis, the analyst needs help in determining which questions to ask about which variables because so many variables are potentially relevant and the analyst likely does not have the resources to ask about every particular variable. Simply put, a formalized approach could help organize these questions in a way that is useful to the diagnostician. Three ways it can do so are discussed here. Each would ultimately depend on a well-developed knowledge base containing OWL statements and accompanying individual-level RDF data. This in turn would depend on identifying a large number of SESs along with a fair number of theories of fit beyond the few examples I have discussed here. I believe it is a realistic expectation that such theories could be identified across a range of SESs.
To begin, the use of a formal language can help implement the diagnostic approach through the relationship of a superclass and its subclasses. This is an exercise in taxonomizing. We could, for example, create subclasses of SESs or their components. Each subsequent subclass of each component, for example a class A, represents a set of diagnostic questions. For example, if class A signifies that a particular resource system is a groundwater aquifer, then we need to ask questions that pertain exclusively to resource systems of this type, such as whether it is confined or unconfined. In this way, specifying whether a particular instance of an SES or component such as a resource system belongs to a particular subclass can help an analyst explore the most relevant questions for that type of system.

Secondly, the subcomponent relationship can also play an important role in the diagnostic process. If a class has a subcomponent (such as the SES and its components), this leads to an additional set of questions that need to be asked, based on the properties, subclasses, and subsequent subcomponents of that subcomponent class. For example, we could specify that two subcomponents of an environmental cap-and-trade mechanism (or for that matter a fishery quota system) are the initial limit on emissive or extractive activity, as well as a property-right arrangement that organizes the scarce rights whose value is in part created by the limit. Doing so provides a helpful structure whereby an analyst considering these types of institutional arrangements knows to ask questions about the subcomponents involved (e.g., what types of property rights are instituted).

Thirdly, an analyst could use the information embodied in theories of fit to guide a diagnostic analysis of a particular case. If an analyst observes a variable in a case and has a set of theories of fit in hand, he or she could then ask which theories this variable is involved in. If queried, a knowledge base containing formalized expressions of theories of fit could return all the theories that involve such a variable. From this, the analyst would know that these theories, and the variables they relate, represent potentially important questions to ask about his or her case.

Additionally, an analyst could weight the potential relevance of such theories by how similar the cases that support them are to the analyst’s case in question. This similarity would need to be evaluated based on whether or not two cases have the same values for critical variables (such as geographic size or the type of resource being managed). For example, if an analyst observes that there is spatial biophysical nesting in a watershed, and his or her case in question is very similar to other watershed management cases where spatial nesting of governance had successfully addressed this biophysical nesting, a good question to ask for his or her particular case would be whether or not there is corresponding nesting of the governance system.

**CONCLUSIONS**

I began this paper with two closely related questions: (1) can the process of formalization help make the concept of fit more useful to analysts, and (2) can such formalization help analysts conduct a procedure for diagnosis to find which institutions fit best within a particular situation? I believe the answer to both questions is yes. Unpacking the concept of fit has turned out to be a process of formalizing a particular type of theory: theories of fit. Each of these theories relates a set of variables that interact to contribute to an outcome, with at least some measure of independence from the presence of other factors not included in the theory. I believe that such a process of formalization helps improve the utility of the very general concept of fit by unpacking the different theories of fit.

Formalization can also help make the diagnostic process more explicit, while this process currently takes place mostly implicitly in the minds of scientists and analysts. As I indicated earlier, this potential would require a full-developed knowledge base of SESs and theories of fit, which has not so far been attempted. If this were conducted, however, this knowledge base would not only enable explicit social-ecological diagnosis. It would also represent a better organization of social-ecological knowledge. This organization could have additional taxonomic rigor, and could facilitate communication among social-ecological scientists beyond particular research project and team silos. If done well, it could also increase the rigor with which theoretical statements of SESs are made, and in so doing, move us towards cumulative knowledge, which is the hallmark of scientific progress.

---

**Responses to this article can be read online at:**


---

**Acknowledgments:**

Financial support from Formas through the program Human Cooperation to Manage Natural Resources (COMMONS) is acknowledged.

---

**LITERATURE CITED**


Allemang, D., and J. Hendler. 2008. *Semantic web for the working ontologist*. Elsevier, Burlington, Massachusetts, USA. [http://dx.doi.org/http://dx.doi.org/10.1016/B978-0-12-373556-0.00001-0](http://dx.doi.org/http://dx.doi.org/10.1016/B978-0-12-373556-0.00001-0)


[1] Formally, such relationships are called properties in OWL. OWL has two types of properties: object and data properties.

[2] For example, to assign a property to a class X, OWL actually uses several triples that formalize the existence of an anonymous superclass, of which the class X is specified as a subclass. This superclass defines a set of requirements for the members in it, and thus also for class X, as its subclass. One of these requirements can be having a particular property, which is then required for every member of class X. This structure is known as a restriction, in that it restricts (the membership of) class X to those individuals that satisfy the conditions specified for the new anonymous superclass (following the meaning of the subclass relationship).

[3] This follows one of two patterns established for expressing variables and their values using object properties. See http://www.w3.org/TR/swbp-specified-values/ for more details. For variables that take on many numerical values, data properties would be used.


[5] Technically, relationships to classes use the owl:someValuesFrom restriction type, while relationships to individuals use the owl:hasValue restriction type.