CHAPTER X– Defining the Ill-Defined: From Abstract Principles to Applied Pedagogy

Benjamin D. Nye,[1] Michael W. Boyce,[2] Robert Sottilare[2]
[1] USC Institute for Creative Technologies (ICT), [2]Army Research Laboratory / Oak Ridge Associated Universities

INTRODUCTION

Attempts to define ill-defined domains in intelligent tutoring system (ITS) research has been approached a number of times (Fournier-Viger, Nkambou, & Nguifo, 2010; Lynch, Ashley, Pinkwart, & Aleven, 2009; Mitrovic & Weerasinghe, 2009; Jacovina, Snow, Dai, & McNamara, 2015; Woods, Stensrud, Wray, Haley, & Jones, 2015). Related research has tried to determine levels of ill-definedness for a domain (Le, Loll, & Pinkwart, 2013). Despite such attempts, the field has not yet converged on common guidelines to distinguish between well-defined versus ill-defined domains. We argue that such guidelines struggle to converge because a domain is too large to meaningfully categorize: every domain contains a mixture of well-defined and ill-defined tasks. While the co-existence of well-defined and ill-defined tasks in a single domain is nearly universally-agreed upon by researchers; this key point is often quickly buried by an extensive discussion about what makes certain domain tasks ill-defined (e.g., disagreement about ideal solutions, multiple solution paths).

In this chapter, we first take a step back to consider what is meant by a domain in the context of learning. Next, based on this definition for a domain, we map out the components that are in a learning domain, since each component may have ill-defined parts. This leads into a discussion about the strategies that have been used to make ill-defined domains tractable for certain types of pedagogy. Examples of ITS research that applies these strategies are noted. Finally, we conclude with practical how-to considerations and open research questions for approaching ill-defined domains.

This chapter should be considered a companion piece to our chapter in the prior volume of this series (Nye, Goldberg, & Hu, 2015). This chapter focuses on how to understand and transform ill-defined parts of domains, while the prior chapter discusses commonly-used learning tasks and authoring approaches for both well-defined and ill-defined tasks. As such, this chapter is intended to help the learner understand if and how different parts of the domain are ill-defined (and what to do about them). The companion piece in the authoring tools volume discusses different categories of well and ill-defined tasks, from the standpoint of attempting to author and maintain an ITS.

WHAT IS A LEARNING DOMAIN?

It is easy to think of examples of domains (e.g., math, writing, physics, accounting), but harder to find examples of precise definitions of why certain topics are considered learning domains while other topics are typically not (e.g., fire, hearing, the universe). This implies that domains tend to be defined bottom-up, as clusters of knowledge, skills, and tasks that co-occur, either in educational contexts (e.g., curricula) and/or vocational contexts (e.g., expert skillsets). Treating domains as clusters of knowledge, the boundaries of such clusters would be well-defined when their connections and dependencies to knowledge within that domain are much denser than to knowledge considered outside of that domain (i.e., distinct boundaries). So then, a domain can be considered a set of things that:

A) Define an expert in some field (e.g., expertise). This view is useful from a training standpoint, where a domain is modeled after an archetypal expert. This connects strongly with assessment literature, in that most assessments attempt to discriminate between novices and experts (Alexander, 1992); or

B) Are commonly taught together (e.g., pedagogy for an academic discipline). This view is useful for communicating about learning and accommodates knowledge and skills that might underlie a wide range of specializations (e.g., such as reading).

Subject matter domains appear to be descriptive in nature: learning objectives, tasks, knowledge, and skills are grouped into domains due to either the history of academic disciplines (e.g., how we know to teach it) or the broader societal context (i.e., how we know to use it). While some of these differences appear to be due to relationships between knowledge (e.g., despite being a strong prerequisite for physics, the domain of math is considered separate from physics, since too many other domains use math but not physics), others are probably path-dependent historical artifacts (i.e. just how we have always done it).

From a beginning learner's standpoint, any domain may be perceived as poorly defined because its features, rules, and guiding principles are unknown. This is also the first stage of forming a domain from a societal standpoint, since all domains initially begin without experts. In this very beginning phase, the domain is not ill-defined (e.g., lack of agreement) but undefined (e.g., not even specified at a level to ground a discussion). From this stage, an understanding of the domain evolves and causes it to become distinct from other domains (useful for categorization) and also enables identifying patterns that are useful to achieve outcomes that are valued within that domain. In this way, society's defining and understanding of a domain must follow phases similar to those of a beginning learner. Initially, some learners "ad-lib" since there are either no formal rules or guiding principles for a domain, or the learner is unaware of them. For example, dancing may appear to be an ill-defined domain to a teenager on a dance floor for the first time, but is well-defined for a professional dancer in a Broadway play or a ballerina in a ballet production. The differences in perception of the definition of a domain may rest in the learner's familiarity with successful and unsuccessful models of performance, their capabilities to implement a successful model (e.g., talent), their degree of preparation invested to perform well on a consistent basis (e.g., practice), and also a societal agreement about what success means in that domain (e.g., goals).

Starting from an initial lack of structure, domains gain definition and value from their ability to transfer learning, either from one person to another (e.g., pedagogy, communication) or from one task to another (e.g., leveraging similarities between domains). Starting from an initial state where exploration is arbitrary or random, several theories of how to be successful in that domain will tend to emerge. In other words, learners may have several paths to successful performance but lack consensus on a single best methodology. Over time, the more successful methods should win out and continue (i.e., memetic competition with survival of the fittest). Eventually, a limited set of methods may become popular both because they are effective *and* because they are sufficiently simple to be shared with the masses. This leads to a convergence and ultimately to well-defined domains (e.g., ones with high levels of agreement). However, complex solutions with higher levels of performance may not have this same level of consensus.

For example, in the 1860s when the game of baseball was evolving, nearly every batter had a different approach to hitting the ball. Some were more successful than others, such as choking up on the bat (shortening it) to increase their bat speed when facing a pitcher who threw very fast. Over time, the task of hitting a baseball went from "undefined" (lacking any rules or principles) to "ill-defined" (e.g., lacking agreement, the domain changed faster than consensus was reached, or the domain changed due to learning about it, such as in some game-theoretic domains). In baseball, part of the problem was that some people defined paths to success and could perform well using these models, but most people could not: players' strategies interacted with their physical capabilities, as well as game-theoretic considerations (e.g., changing your hitting would change how opponents reacted). This is still true today. While there is consensus on what some elements of what makes a successful baseball hitter, professional hitters still have a variety of strategies and conceptualizations for this process. Hitting instructors model some critical behaviors which are necessary to successfully hit a baseball, but training is limited by the fact that reaction times must be fast (e.g., strategies must be trained until nearly automatic, making them mostly mutually-exclusive), that different players' athletic talents may be better-suited for certain approaches, and that the performance domain adapts to them (e.g., pitching opponents react to changes in approach). These factors combine to form a domain where the goals are well-defined, but the key knowledge to teach varies based on both individual and contextual factors that vary over time.

The distinction between expertise and pedagogy is important when conceptualizing ill-defined domains. A domain is typically considered to be ill-defined because the expertise required to do the tasks is unclear. However, from an instructional standpoint, the pedagogical practices and tools to teach the domain might also be unclear (i.e., ill-defined domain pedagogy). This distinction is the difference between domain knowledge (how to do tasks) and pedagogical domain knowledge (how to teach doing those tasks). The purpose of an ITS (and instruction in general) is to reach learning objectives by applying pedagogical domain knowledge to support a learner in developing domain knowledge.

Well-defined expertise implies the ability to design well-defined pedagogy, but ill-defined aspects of expertise might not necessarily imply ill-defined pedagogy. When domain expertise is well-defined, one or more instructional frameworks are typically also well-defined. Whenever at least one framework is known to be effective, we would typically consider the domain pedagogy to be reasonably well-defined. On the other hand, while cultural competency is poorly defined as expertise, cultural immersion through living in a culture is considered to be highly effective for learning. While pedagogy typically builds on understanding the expertise for a domain, there are other interventions which are general because they either based on human cognition (e.g., affective support) or basic principles of learning (e.g., time on task increases learning). With that said, it is possible to imagine or build "trick domains" where near-universal pedagogy fails (e.g., ones where the more time you spend training, the less you know, by rewarding habits that are counterproductive in the long term). However, these are not domains that are actually taught in practice. This raises the final point about domains: while there are many components that can be ill-defined, there are fewer that are ill-defined for domains that we care about tutoring. As such, it is important to consider that while some ill-defined components of domains may be challenges to overcome through better ITS design, others may be red flags that a domain may not yet be mature or that the domain is framed in terms that become outdated faster than they can be trained.

In summary, we posit the following points that underpin our discussion of ill-defined domains:

1) Nearly all domains have a mixture of well-defined and ill-defined components.
2) ITS design for domains can be ill-defined due to components of either domain expertise or its pedagogy.
3) The range of ill-defined domains is restricted due to the nature of the domains that are desirable to train.

With those premises stated, we can then attempt to identify components that make up a learning domain and how they can be ill-defined.

What Are the (Ill-Defined) Components of a Domain?

Intelligent tutoring systems can be designed based on either the behavior of domain experts or based on the practices of teachers and other instructional experts. In both cases, a domain is primarily specified in terms of the set of *tasks* that a person or group performs (Nye, Goldberg, & Hu, 2015). Multiple domains may share the same tasks, though those tasks may have different *relationships* with each other within one domain versus another. Such relationships might be temporal (e.g., prerequisites), part-whole (e.g., subtasks), or include other types of interactions.

The tasks in a domain can be framed using Markov Decision Processes (MDP; Bellman, 1957), in that it can be assumed to have *states*, *actions*, and transition probabilities between states based on the actions performed. Both states and actions are constructs, in that states and actions represent the space of features and interventions that are useful for decision-making or communicating decisions. In an MDP, these would be the state vector features, state space, and the action space. Tasks, both individually and in aggregate, also imply *goals*: certain states or trajectories of states that are preferred to other states (e.g., the utility function for an MDP). Finally, *assessments* are an important component, because they represent tasks or measures that can monitor goals based on the states and actions (e.g., a function over observable nodes in a partially-observable MDP).

Defining Components for Expertise versus Pedagogy

These components are shown in Figure 1, both in terms of how they apply to the domain expertise itself and to the domain pedagogy (i.e., the tasks of teaching someone to learn that domain expertise). So then, an expert is someone who can do all the domain tasks that a typical expert can do, as inferred from an existing pool of assumed experts (Hoffman, 1988). Likewise, during instruction, a learner is considered to have mastered the content when they can meaningfully complete all the learning tasks for the domain. Tasks that are useful for learning are often not the same ones seen in real-life, but instead tend to emphasize key concepts or are bounded due to contextual constraints such as cost or safety. Theories of assessment bridge expertise with instruction, by aligning learning tasks to domain tasks (Alexander, 1992; Fitzpatrick, Hawboldt, Doyle, & Genge, 2015; Gipps, 1994).

For a completely well-defined domain, every component of Figure 1 would be straightforward to specify and there would be no disagreement between different experts. In practice, particularly for domain pedagogy, there is almost always some disagreement or uncertainty about optimal specifications. This is because instructional design is indeed a design problem, which are characterized as one type of ill-defined task. An ITS requires a well-defined instructional design (Domain Pedagogy, right side Figure 1) that can be implemented computationally. To do this, the instructional design must accommodate the various well-defined and ill-defined aspects components of the domain expertise.

Domain Expertise	Domain Pedagogy
Tasks: What tasks do experts do? - **States/Features:** What features do experts monitor? - **Actions:** What do experts do to influence transitions between states?	**Tasks:** What are the training tasks? - **States/Features:** What actions should a teacher/ITS monitor? - **Actions**: What interventions should be used to improve learning?
Relationships: How do tasks interact?	**Relationships:** How should tasks interact?
Goals: What task states or pathways are considered better?	**Goals:** What are the learning objectives? How do these relate to task expertise?
Assessment: What tasks and outcomes measure performance?	**Assessment:** What tasks and outcomes measure changes due to learning?

Figure 1: Components of Domain Expertise (Doing) vs. Domain Pedagogy (Teaching)

Assuming these components for a domain and its pedagogy, what does it mean to be ill-defined? In practice, any one of these components is ill-defined when there is no convergence to a countable number of agreed-upon elements. For example, domain experts may agree about the key tasks for a domain (well-defined tasks), but may disagree about the set of features that are relevant to those tasks (ill-defined features for task states). Since components can be independently ill-defined, the degree that a domain is ill-defined depends on not just the level of disagreement but also on which components are ill-defined.

As an example, Figure 2 contrasts a well-defined domain task (basic Algebra) against a highly ill-defined domain task (country stability). In a fully-defined task, it is possible to determine at least the order of value for states and also assign value to the actions that the user took. In a predictable task, the value of actions is directly determined by the value of the states, though in a chaotic or probabilistic domain, action value might be only loosely connected to state utility (i.e., optimal actions still sometimes have bad outcomes, and vice-versa). For a highly ill-defined task, it may be difficult to determine the actions and features of the state that are relevant (e.g., a “mess”; Ackoff, 1981). Figure 3 gives examples of domains where either the importance of states and/or actions are better or worse defined. A second major difference is that experts agree about solutions and goals for well-defined tasks (e.g., simplifying an Algebra equation), but ill-defined domains can have fundamental disagreements over optimal solutions. For example, the United States and Taliban both have vastly different goals for a stable government in Afghanistan.

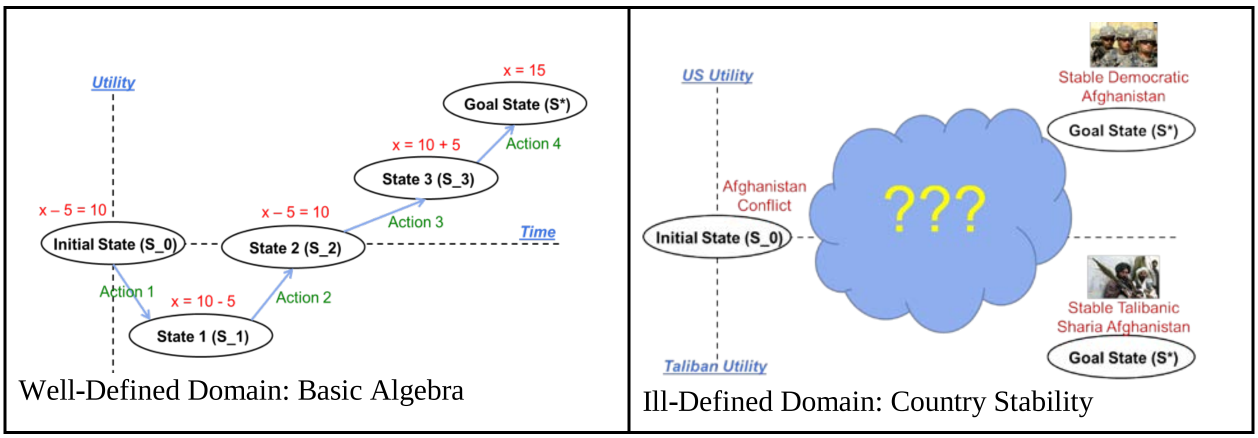

Figure 2: Contrasting the Components of a Well-Defined vs. a Highly Ill-Defined Domain

	Action Values Known	Action Value Unknown
State Utilities Known	Can evaluate all actions and suggest good solution paths *Ex. Economic decision tree*	Know the value of states, but can’t surely deduce what actions do *Ex. Stock market simulation*
State Transition Gradients Known	Can rank states relatively and can suggest next step or bug fix. *Ex. Solving an Algebra problem*	Know better/worse changes, but can’t suggest concrete actions *Ex. Automated essay assessment*
State Values Categorical	Can detect good/bad actions, but can’t rank improvement *Ex. Going out of bounds in a race*	Ill-defined task, due to emergent dynamics or subjective values *Ex. Building a better world*

Figure 3: Levels of Definition for the Value/Utility of Task States and Actions

FUNDAMENTAL AMBIGUITIES (WHAT/WHY) VS. UNDERLYING AMBIGUITIES (HOW)

Fundamental ambiguities about the boundaries of a domain are present when experts cannot agree on either the tasks or on the goals (i.e., what and why). These ambiguities mean that experts disagree about what important activities occur and what would be considered a good outcome. These indicate uncertainty about what should be taught within a given domain, in terms of what a domain is. Selecting tasks determines that you won't teach math in an English curriculum, for example. Ambiguities over good/bad states are when you agree about how to frame the domain, but don't necessarily agree on the importance of outcomes or actions. For example, history is sometimes criticized for emphasizing events such as wars as compared to other important events such as trade or peaceful regime change (Nash, Crabtree, & Dunn, 2000). History is also an example with ambiguities about goals: historians have served many roles within cultures, ranging from neutral archivists who preserve the past, to interpreters who try to communicate it to the present, to revisionists who re-interpret history to try to change our future view of the past (Grele, 1981). The relative importance of different tasks may also be debated. These are fundamental ambiguities about what constitutes expertise and what goals one should be able to accomplish by studying history. Other core disciplines, such as philosophy, ethics, legal argument, and social policy also hit such issues. ITSs for domains with fundamental ambiguities are relatively rare, and tend to focus on the process of studying the domain, rather than arriving at specific outcomes (Pinkwart, Ashley, Lynch, & Aleven, 2009; Easterday, Aleven, Scheines, & Carver, 2011).

Underlying ambiguities include task states, task actions, and relationships between tasks. These more granular ambiguities are present when experts can agree on tasks enough to make meaningful comparisons between them, but disagree about how to represent what the feature space, the state space, or what actions are available during a task. Many domains commonly identified as ill-defined are due to these issues, such as programming (Mitrovic, Koedinger, & Martin, 2003) and design (Chesler, Arastoopour, D'Angelo, Begley, & Shaffer, 2013). Finally, assessment ambiguities can occur when it is unclear which benchmarks should can be usefully applied to measure progress toward goals. This type of ill-defined component can occur because the ideal goals are not directly observable (e.g., long term economic growth or learning), which requires measures that attempt to estimate when goal states and paths are reached.

The above issues that affect domain knowledge and expertise apply similarly to domain pedagogy, where the task is building an instructional design (e.g., a curriculum, either adaptive or linearly sequenced). Much like for the domain itself, instruction has tasks (learning tasks), goals (learning objectives), relationships between tasks (e.g. knowledge prerequisites), and assessments to monitor progress toward the learning objectives. Likewise, for each learning activity, a pedagogical feature state and possible interventions exist. While all of these elements can be ill-defined, fundamental disagreements at this level occur at the goal level: being unable to agree on the learning objectives. Unlike the domain expertise, disagreements about learning tasks are not necessarily fundamental, since these are the means-to-an-end. Otherwise, the avenues for ill-defined domain pedagogy are similar to those for domain expertise.

So then, ill-defined domain expertise does not necessarily imply ill-defined domain pedagogy, provided that the learning objectives and goals remain well-defined. However, ill-defined components of a domain must inherently constrain the available pedagogical options. In short, ill-defined domains can cause certain pedagogical options to become incoherent or impossible. For example, if the desirability of task states cannot be compared, then it is impossible to give error-correction feedback. This has significant implications for the types of instruction and pedagogical strategies that can be applied.

Given that we conceptualize components as ill-defined due to ambiguous components and a lack of agreement between experts, it is important to identify root causes that may underlie such disagreements. Components may be ill-defined for a number of reasons (not exhaustive):

1) **Subjectivity** occurs when different experts disagree about either the relative importance of tasks or about the value of states within tasks. If the goals are subjective, two experts might see the same events, with one declaring it a ringing success while the other expert declares it an utter failure.
2) **Under-specification** occurs when domain tasks or goals tend to be context-specific, but it is infeasible or impractical to represent the space of contexts that impact the tasks. Under-specification can underlie subjectivity, such as comparing computer code by an author who prefers brief code versus one who prefers extensive code documentation. These preferences are typically due to different design goals, and both experts might agree if one goal or the other was stated or explicitly weighted.
3) **Representation/mental-model mismatches** are a related problem, where experts may have different levels of specification for different tasks (e.g., unique mental models; Ososky, 2013) or when experts have similar mental models but communicate them differently (e.g., different names, same concepts; Norman, 1983). This can be observed as disagreement about the specific features or actions involved, despite agreement about goals and the value of final outcomes.
4) **Assessment/measurement confounds** occur when the states and values may have relatively high levels of agreement, but are unobservable and assessment methodologies struggle measure those key outcomes.
5) **Rapidly evolving domains** occur when domain knowledge becomes outdated faster than it is possible to train the knowledge or skills. For example, sommeliers specialize in the process of tasting wines and typically a specific regional specialty, rather than attempting to be an expert in all wines. This is because the variants of wine released in a year is far more than any one person could taste.
6) **Wicked problems** or “messes” occur when multiple tasks are highly interconnected, evolve over time, and also prone to subjective goals (Ackoff, 1981; Rittel & Webber, 1973). Confounds include imperfect information, unintended consequences, probabilistic results, emergent outcomes, and many stakeholders with different value systems (Silverman et al., 2010). These are commonly found in complex social problems, such as public policy. Wicked problems are complex because not only do different experts disagree at a given time, the same expert might have different opinions depending on the time horizon of interest.
7) **Uncontrollable problems** occur when the domain is inherently ill-defined, in that the tasks are not actually human-controllable because they are inherently chaotic or the information to meaningfully direct them is unobservable (Kalman, 1959). Such tasks are not teachable, since there is only the illusion of control. Related to this are problems where control is only temporary (e.g., cyclical or otherwise non-asymptotic), because entropy rapidly washes out human inputs and their traces from the system (e.g., building sandcastles). Another related category is systems that with high inertia or tendencies toward entropy, where change due to human inputs are only observable after long periods of effort despite no detectable changes (e.g., breaking a boulder by kicking it for years). The latter type might be considered energy-responsive, in that a certain critical mass of inputs (either all at once, or over time) are needed to cross a tipping point where change can be observed.

With the exception of truly uncontrollable problems, pedagogical strategies have been developed to accommodate domains that suffer from these root causes for different aspects of the domain goals, tasks, relationships between tasks, and assessments. Examples of strategies to address ill-defined components are noted in the next section.

STRATEGIES AND SOLUTIONS: SIMPLIFYING, SPLITTING, TRANSFORMING, & ASSESSING ILL-DEFINED COMPONENTS

Four main strategies have been proposed to simply ill-defined aspects of a domain so to make it pedagogically tractable for an ITS: authority-based simplifications that limit their training to a subset of a complex domain (Roberts, 2000), hybrid ITS that split the domain into well-defined and ill-defined tasks that are tutored differently (Fournier-Viger et al., 2010), transforming ill-defined domain tasks into pedagogically well-defined tasks (Mitrovic et al., 2003), and expanded assessments beyond traditional objective, criterion-based assessments (e.g., explicit handling of subjectivity; Nye, Bharathy, Silverman, & Eksin, 2012).

Simplifying the Domain: Authority-Based Simplifications

The first strategy for simplifying ill-defined aspects of a domain is entirely direct: simply choose a point-of-view and stick with it. This approach originates in literature on the study of complex social systems, often-called messes or wicked problems (Ackoff, 1981; Rittel & Webber, 1973). Choosing beneficial courses of action in a socio-cultural system is ill defined because the framing of the problem determines the value of the solution (Rittel & Webber, 1973). Confounds include imperfect information, unintended consequences, probabilistic results, emergent outcomes, and many stakeholders with different value systems (Silverman et al., 2010). This makes the goals for actions and outcomes in social systems subjective, due to differences in value systems or framing. This ambiguity is intractable because it is based in serious moral and philosophical questions, such as “How much money is a life worth?" and “Is health or security more important?" Even for an individual, quantitatively answering such questions involving life and death tends to be difficult and inconsistent (Waldmann & Dieterich, 2007). Due to their complexity, issues such as policy problems, economic interventions, and military strategies are often analyzed using forecasting simulations to better understand potential outcomes (Ichikawa et al., 2010; Silverman et al., 2010). However, when such approaches are used, actions often have implications beyond immediate outcomes and contribute to later emergent outcomes.

Roberts (2000) notes that wicked problems can be reduced using authoritative simplifications. This approach transforms the problem by using a subset of expert perspectives to reduce the number of stakeholders. The training objectives selected by funding sponsors and developers intrinsically introduce authoritative simplifications. For example, a training environment for the US military will assess performance in terms of allied interests and US. This can reduce certain globally subjective issues into objective tasks. Domain experts can also provide authoritative structure. However, experts often differ on useful courses of action to improve a social problem so subjectivity cannot be entirely eliminated. The Complex Environment Assessment and Tutoring (CEATS) prototype for simulation-based training of counter-insurgency operations applied this methodology to convert a faction-based predictive model of district stability into a learning environment (Nye et al., 2012). CEATS employed “objective” measures of performance (e.g., increasing alliances with local groups) that were actually just special cases of subjective measures of performance where the subject (US and International Forces) was held constant.

Authority-based simplifications are particularly useful in two cases. The first case is when all other methods to transform or simplify the domain have been exhausted, and it becomes necessary to simply settle on a (possibly even arbitrary) representation. In some cases, this is not necessarily bad. For example, agreeing on expert knowledge components to adaptively select learning tasks can be quite difficult, while methods to infer useful knowledge components to perform this task are quite mature (Ritter, 2015). As such, if data can be readily collected, it is a reasonable strategy to make best-guess Knowledge Components (KCs) and prerequisites and then apply sufficient sampling to explore the space of alternatives to find a more-optimal set. The second case is when stakeholders that are critical to adoption would prefer to focus on certain tasks or features, as compared to others. In this case, good standards in user-centric design dovetail with the need to simplify ill-defined components. By consulting with the stakeholders involved in adopting, using, and maintaining the system, it may be discovered that certain domain tasks, features, or assessments are more useful and hence more likely to be used to support learning. This is particularly important since the best designed ITS is useless unless people use it.

Splitting Up the Domain: Handling Well-Defined and Ill-Defined Tasks Separately

One strategy applied to address ill-defined domains has been hybrid tutoring which splits the domain into different tasks and features based on their well-defined and ill-defined characteristics (i.e., divide-and-conquer). Hybrid tutoring approaches apply multiple qualitatively different pedagogical interventions to different tasks (Fournier-Viger, et al., 2010). Since many domains have a have a mixture of well-defined and ill-defined problems, the first strategy is to identify and split the domain into the tasks, which are well define versus those that are ill-defined. Ideally, a hybrid design would guide tutoring based on the nature of the problem. However, hybrid-tutoring designs are an emerging topic and lack established design principles.

Fournier-Viger et al. (2011) describes one application of a hybrid ITS to assessing and tutor the ill-defined domain of controlling a three-dimensional robot arm with seven dimensions of freedom. This ITS applied model-tracing tutoring to a set of well-defined spatial cognition procedures, partial task models (sequential patterns) that were extracted from annotated actions sequences, and constraint-based modeling to a path-planning simulator to catch critical failures (e.g., damaging the arm). This approach represented the domain from multiple levels, with each level able to support different types of feedback. Model tracing provided feedback, hints, and worked examples for the high-level process and basic controls for the robot arm (e.g., checking cameras for potential obstructions before moving the arm). Partial task models were used to establish similarity of the user's performance to expertise archetypes (e.g., expert, intermediate, novice) and could be used to suggest next-step hints. Finally, the constraint-based model was used to detect and remedy invalid actions.

Guiterrez-Santos, Marvikis, and Magolus (2010) proposed subsumption architectures to break up a complex problem into a series of layers. In this approach, the ITS assumes higher-level tasks with subtasks that determine the state and performance at the higher level as well (i.e., learners do not directly act on the high-level task). A subsumption architecture can provide one mechanism to implement a hybrid ITS, where different layers monitor and attempt to tutor subtasks, and high layers monitor overall task and learning goals to help determine what pedagogical interventions should be applied. Guiterrez-Santos et al. (2010) tutored micro-worlds using this approach, by tutoring a higher-level abstraction of the micro-world. Dividing the system into layers did this: the expresser (the exploratory environment), computation (attempts to solve well-defined tasks), and aggregation (integrates computational solutions into the context of the larger exploration goals). It does this through a rule-based approach where information from different types of data are mapped to satisficing criteria (i.e., conditions for "good enough" solutions as opposed to empirical optima). There is also a check to see if that portion of the solution maps closely to any of the already existing solutions.

Transforming the Domain: Translating Ill-Defined Components into Well-Defined Pedagogy

Transforming the domain imposes additional structure onto ill-defined tasks to support training. Transforming the domain can also partially address the issue of emergent outcomes. Strategies for mid-game chess and Go rely on strategic features of board structure such as territory as a proxy for direct game-state policy evaluation, which is otherwise an intractable problem (Richards, Moriarty, & Miikkulainen, 1998). Domain experts in complex socio-cultural problems such as counter-insurgency and economics use similar feature-based approaches to assess the situation (U.S. Army, 2010). Such proxy measures of future value complement the immediate outcomes of actions. However, for complex domains or simulations of such domains emergent outcomes can occur that conflict with expert expectations. This means that formative and summative assessments may be inconsistent (e.g. good actions, bad outcomes).

Transforming Fundamental Ambiguities

A number of techniques have been applied to try to transform a domain to eliminate fundamental ambiguities, such as what constitutes a good legal case. The most common high-level approach to such domains is to focus on the process (i.e., actions) that are performed, with less emphasis on the specific content. For example, the PolicyWorld game breaks policy thinking (highly ill-defined) into structured inquiry learning steps (Easterday et al., 2011). As such, it partially sidesteps the issue of which specific policies are good or bad. Instead, it focuses on helping the learner understand the actions and steps that are necessary to make an informed decision about a policy. Systems such as the Legal Argument Graph Observer (LARGO) apply similar transformations where legal arguments (usually expressed as text) are instead represented using graphs, to scaffolded structured argumentation (Pinkwart , Ashley, Lynch, & Aleven, 2007).

A strongly related approach involves having learners perform traditionally unstructured tasks (e.g., argumentation) using highly-structured tools that limit them to a better-defined analog. Graphical visualizations (e.g., concept graphs) and form-based entry are common solutions. Lynch and Ashley (2014) proposed a graphical structure that allows for logical relationships, to discover areas where the student arguments contained conflicting information. Similar approaches were used by both PolicyWorld and LARGO (Easterday et al., 2011; Pinkwart et al, 2007). While imposing an artificial task or structure reduces the ecological validity of the learning task as compared to the domain tasks, this sacrifice can give better analysis and traction for other pedagogical goals (e.g., structuring logical arguments) by disentangling them from more complex problems (e.g., writing up a rhetorically-persuasive legal argument).

Metacognitive and affective tutors can also be applied to tasks that are not well-represented by the ITS, since they attempt to train reusable skills that should improve performance on more than just the current domain. As an example of non-cognitive modeling, Baker and Corbett (2014) identified a series of features which can help to detect whether a student is undergoing robust learning through metacognitive activities, such as the amount of time spent after receiving a feedback message related to lesson content, pauses between answering questions that the student should have knowledge of, attempts to game the system or participating in off-task behavior. These detectors can serve to support the instructor's understanding of individual student performance, thereby increasing areas of specific emphasis of their pedagogical strategies.

A final category of transformation is to avoid representing the whole domain, but instead only tutor specific pathways for complex tasks that have multiple (or even uncountable) solutions. Scenario-based training, path-branching ITS, "what-if" scenarios, and example-tracing tutoring can provide this type of transformation (Lane et al., 2008; Aleven et al., 2009). Particularly when the goal of the tutoring is to build basic familiarity or to tutor skills that can be readily trained in that specific set of tasks and paths, this approach can sometimes be more efficient than a higher-fidelity model that accurately represents the probability of certain events. For example, part of the success of both the Sherlock tutor for technician training (Lesgold, Lajoie, Bunzo, & Eggan, 1991) and the Digital Tutor (Fletcher. 2011) was attributed to their ability to train rare diagnosis and repair tasks that occur within larger technological systems. Similarly, scenario-based learning for uncommon but important social and cultural competencies might achieve similar impacts.

Transforming Underlying Ambiguities

Different transformations are applied when experts can agree about the high-level tasks and general goals, but cannot agree about how to measure progress or struggle to compare the relative value of specific states or values. These transformations typically convert a task with a poorly-defined state space or valuation of states (e.g., utility) into a simpler set of features or detectors that can be used to drive tutoring.

Constraint-based modeling (CBM) is one of the most well established approaches to simplify tutoring a task with ill-defined state or action values. Rather than being concerned with what the student has done, the goal in CBM is to give the students as much flexibility as possible so long as the student doesn't reach a situation that is known to be incorrect. (Mitrovic et al. 2003). In CBM, the ill-defined domain is made more tractable by providing constraints that can create a boundary while still allowing the learner to make choices (Woods et al., 2015). This supports ITSs that act when specific flaws or out-of-bounds conditions are detected, even though the relative utility of different states is difficult or impossible to assess. A CBM ITS only needs features of acceptable solutions in order to work properly (Mitrovic & Ohlsson, 2015), which is advantageous in terms of ill-defined domains because it can solve open-ended problems. Mitrovic has spent the better part of the last two decades analyzing CBM learning through use of a SQL based tutor. SQL is a good candidate as an ill-defined domain because there can be multiple correct solutions without a clear algorithm to guarantee success. To assist with the ill-defined nature of SQL queries, students are taught search strategies using example problems, which they can then use as reference points to apply those strategies to actual problems (Mitrovic & Weerasinghe, 2009).

Instance-based inference, such as case-based reasoning (CBR), is a second approach to dealing with ambiguity, and is particularly useful when the set of useful or agreed-upon features for a state space are not fully known. These techniques are applicable when cases can be assigned to specific groups (e.g., good/bad examples; expert/intermediate/novice; expert1/expert2/expert3). For example, Goldin, Ashley, and Pinkus (2006) used case-based methods to tutor case analysis as part of legal ethics. Combining a context classifier that detects which concepts have been defined and a CBR manager that compares these concepts to identify similar cases did this. It then reports what concepts have been defined in those cases and how they have been applied. Similar approaches where tagged instances are directly leveraged have also been observed for well-defined domains to identify expert blind spots (e.g., features that experts no longer see; Matsuda, Cohen, Sewall, Lacerda, & Koedinger, 2007). Instance-based inference relies on the ability to apply pattern matching to find similar instances or to identify clusters, even though specific rules or expert guidelines have not been developed. If the domain is fundamentally ill-defined, instance based clustering (e.g., good/bad labels) might actually show multiple clusters that represent the convolution of both the expert raters' attitudes and the instance features. As such, this approach can be useful both as a solution and as a diagnostic tool for fundamental ambiguities.

Similar approaches have been applied to user trajectory data (i.e., each instance is a pathway taken by a user during a certain task session), rather than a corpus of static cases. These approaches rely on data mining to compare ongoing results to trajectories of data and adapt to the user (Baker & Corbett, 2014). Lazar and Bratko (2014) created an intelligent tutor using this technique to tutor Prolog programming. They do this by counting the number of times a line and associated line edits appear. By searching the potential paths that are known to get to a successful solution, user edits are ranked and hints are generated. These approaches are analogous to case-based approaches, except that each combination of user-task session represents a case.

Alternative Assessments: Measurements Beyond Performance or Mastery

When the goals of a domain are not well defined (e.g., uncertainty over pedagogical goals or standards), it is difficult to provide learners with guidance to meet those goals; FitzPatrick. Hawboldt, Doyle, and Genge (2015) ran focus groups and document analysis that found that over half of the assessment tasks did not tie in with the respective objectives in terms of Bloom's revised taxonomy (Krathwohl, 2002). Not only were the assessment tasks were not matched to objectives, and over half the assessment questions were categorized at the lowest level ("remember") despite little to no learning objectives targeting that level (FitzPatrick et al., 2015).

In many cases, seemingly ill-defined domains can sometimes be implemented in ITS by following solid curriculum design principles that help instructors structure the domain and assessments rigorously. This can improve alignment of pedagogical goals to the target domain tasks. In this case, the pedagogical domain knowledge comes from the teacher's personal learning experiences while they were in school, the teacher's education and professional development, and the teacher's experience while teaching others

(Friedrichsen et al., 2009). Even if the domain is lacks consensus in some areas, expert teachers may actually still use effective techniques that have been refined over multiple classes, which might only need to be verified using validated assessments that carefully capture the learning objectives.

However, in many cases, objective and criterion-based assessments are not available. Sometimes, this limitation is only due to practical issues (i.e., it only needs to be done). However, in other cases, objective assessment criteria do not exist because experts or instructors cannot agree about the value of different tasks, states, or actions for a domain. From the standpoint of a classical ITS that attempts to support error-correction feedback and next-step hints, this is an insurmountable hurdle: how can one provide feedback if there is no way to detect a right answer? Unfortunately, this is somewhat true: certain tasks simply do not have objective solutions (e.g., subjectivity) and others defy the ability to suggest a next step (or steps) that will guarantee better outcomes (e.g., emergent results).

While such tasks are not amenable to classical ITS interventions, they are not impossible to measure or to deliver meaningful adaptive pedagogy. Research on assessment literature has identified a variety of alternative educational assessments. An assessment has at least four key facets: its reference basis, intended usage, objectivity, and relationship of the learner to the assessed concept (Gipps, 1994). The first three are listed in Figure 4. The reference basis represents what the assessment compares the learner against. A criterion-based assessment assumes some particular task or universal benchmark (e.g., running a 5-minute mile). Alternatively, a norm-based assessment compares the learner against some peer group. Ipsative assessments compare the learner against themselves, either at different time points or between different skills/domains. The usage of an assessment can primarily formative (i.e., help adapt to the learner) or summative (i.e., assess the state of the learner, such as for reporting or communication). Finally, the objectivity of the assessment means that it can be either objective (i.e., universally-agreed or mandated) or subjective (i.e., depends on viewpoint or rater characteristics).

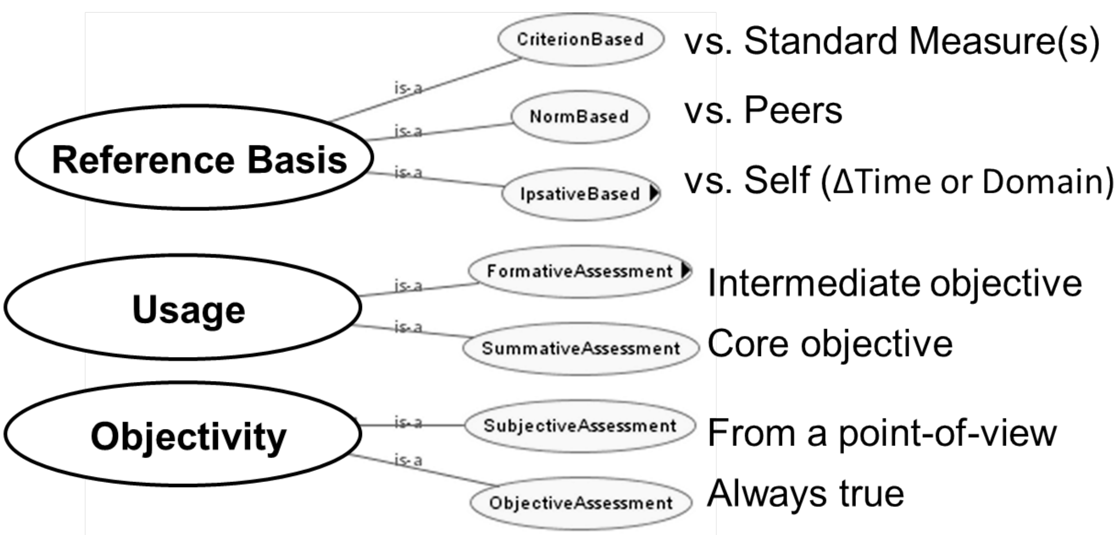

Figure 4: Key Assessment Characteristics

Building on these characteristics, it is possible to assess qualitatively different relationships with constructors and knowledge. The vast majority of assessments focus on measuring task performance, mastery, and learning (i.e., change in mastery). However, assessments of academic emotions (e.g., engagement with a topic), attitudes (e.g., motivation to continue studying a topic), frequency of use (e.g., how often a learner tends to use a skill spontaneously), and a variety of other relationships for the learner to knowledge or skills are likely essential for long-term learning and retainment.

Expanding the range of assessments significantly expands the scope of feedback than can be provided. Figure 5 notes a list of common pedagogical interventions, as limited by the type of assessment involved. To interpret this table, an assessment can be categorized in terms of Objectivity x Usage x Frame of Reference. If a checkmark exists in each one of these fields, then the assessment should always be able to power the given intervention. Error feedback, providing a right answer, hinting at the next optimal step, and other common interventions (VanLehn, 2006) rely on a well-defined problem structure where objective criteria are established. These can be used for well-defined tasks where traditional right/wrong feedback and hints would be effective. Comparative feedback based on norm-based or impassive assessments fills a complementary role and can be used more broadly, indicating how the user's performance differs from other performances. Alternative approaches used for ill-defined domains are domain non-specific approaches, such as inquiry learning, dynamic collaboration, and Socratic questioning (Fournier-Viger et al., 2010). Inquiry learning appears well suited to consuming impassive assessments, because learners build explanations in a stepwise manner. Consistency between steps is important, so examining a user's next steps in light of their prior steps seems most appropriate. Robust approaches include Socratic questioning (e.g., “Why…”) and self-reflection prompts, which can employ knowledge from nearly any type of assessment. Selecting or suggesting collaborative groupings is also robust with respect to the knowledge available, but requires a learning environment that supports collaboration.

Pedagogical Intervention	Objectivity		Usage		Frame of Reference		
	Objective	Subjective	Summative	Formative	Criterion	Norm	Ipsative
Error Feedback	✓		✓	✓	✓		
Correction Hint	✓			✓	✓		
Next-Step Hint	✓			✓	✓		
Comparative Feedback	✓	✓	✓	✓		✓	✓
Inquiry Learning Prompt	✓	✓		✓			✓
Reflective / Socratic Prompt	✓	✓	✓	✓	✓	✓	✓
Collaborative Grouping	✓	✓	✓	✓	✓	✓	✓

Figure 5: How Common Interventions Depend on Assessments Types

While this is not intended to be a universal guide to mapping alternative assessments to interventions, it demonstrates the possible pedagogy that can be leveraged by mapping alternative assessments to alternative ITS interventions. Future research will hopefully map out a deeper space of guidelines for this type of mapping, which could adaptively select or suggest interventions that match the available assessments.

RECOMMENDATIONS FOR HANDLING ILL-DEFINED COMPONENTS OF DOMAINS

Based on this discussion, a particular set of steps are recommended when building an ITS for a domain that may have ill-defined components. These steps can be integrated into existing approaches to understanding domain knowledge, such as cognitive task analysis (Schraagen, Chipman, & Shalin, 2000) and curriculum design methodologies (Wiggins & McTighe, 2005). Considering Figure 1, it is assumed that instructional designers for ITSs would first attempt to understand the domain expertise (Fig 1, left side) and then would attempt to understand and implement domain pedagogy (Fig. 1, right side). As noted previously, it is also assumed that a designer would often attempt to identify set of tasks, the features and actions of those tasks, the relationships between tasks, the goals and relative values for task outcomes/actions, and the assessments that measure progress toward goals. During this process, it should be possible to identify and address ill-defined components. The following steps are recommended for handling ill-defined components:

1) Authority Simplifications to the Task Set,
2) Segmenting Tasks into Well-Defined and Different Ill-Defined Categories,
3) Transforming Ill-Defined Tasks into Well-Defined Pedagogical Tasks (e.g., ones where appropriate feedback and real-time interventions are possible), and
4) Assessing Traditional and Alternative (Subjective, Ipsative, etc.) Relationships to Domain Knowledge.

Authority Simplifications to the Task Set

If the set of tasks that constitute domain expertise is ill defined, authority-based simplifications are usually applied. Ultimately, education and training is a socially constrained process: skills are trained because some need exists. As such, if experts disagree about what tasks constitute expertise, the stakeholders who are adopting the ITS (e.g., schools, companies, governments) can act as an authority to simplify this set based around their perceived needs. This is important even for well-defined domains, since there is little point in building high-quality activities that stakeholders would not want to use.

Split Tasks into Well vs. Different Ill-Defined Categories

Once the set of tasks is identified, determine the level of definition that is available for each task. As noted, this involves considering the task state features, state space, action space, and goals/utility of states and actions. Next, bin tasks into different categories based on the ability to specify these components, both for domain expertise tasks and for learning activities. Well-defined tasks have the widest set of possible interventions available. Nye et al. (2015) notes the types of ITS interventions that are possible for different categories of tasks.

Transform Ill-Defined Tasks into Well-Defined Pedagogy

To expand the range of pedagogy that is available for a task, it is possible to transform certain ill-defined components into more tractable forms. This is the most complex step, since it depends on the specific types of ill-defined tasks that must be trained. Constraint-based modeling is often applied to design tasks where the space of invalid designs may be known but the qualities of optimal good designs are not. Case-based modeling and comparison against archetypes is applied when tagged examples are available (e.g., good/bad cases), but features or actions are unclear. Example-tracing tutors and branching tutors can also be applied where the state space or dynamics are large (or possibly not fully known), but tutoring a limited path or scenario through the space is still useful. Authority simplifications can also be applied to task features, actions, or goals as well, if necessary. When applying simplifications or transformations, the most important consideration is that the learning objectives should ideally not be simplified. For example, if a key domain task is to search and choose between a wide array of options, a tutor that only allows one path would be unable to train that skill.

Assess both Traditional and Alternative Relationships to Domain Knowledge

Particularly where task goals and value functions for states and actions are not agreed-upon or hard to measure, alternative assessments may be necessary. Alternative assessments tend to either be decoupled from the specific states of the domain (e.g., measures of learner affect) or parameterized. Subjective assessments, for example, allow comparing learner performance against different subjective benchmarks (e.g., experts applying different theoretical frameworks). Norm-based assessments allow comparing a learner against other learners, even if a “right” answer is unknown (e.g., identifying outliers or clusters of learner types). Parameterized assessments can enable additional types of feedback (e.g., comparative feedback) that would be unavailable otherwise. While alternative assessments have not been traditionally used in ITSs, they open up a wide range of possibilities for addressing ill-defined domains (e.g., explicitly tutoring where experts disagree due to different schools of thought).

By applying these four high-level strategies and implementing lower-level strategies noted in the prior section, it should be possible to provide pedagogically useful instruction to a wide variety of seemingly intractable domains.

CONCLUSIONS AND FUTURE DIRECTIONS

Overall, this conceptualization of ill-defined domains makes them appear significantly more tractable for an ITS than they might initially appear. A core goal of this analysis was not to fall back on the "God of the Gaps" definition for ill-defined domains, where a domain or its tasks are ill-defined because existing tutoring systems cannot handle that domain. By first establishing a working definition for a domain and its components, it becomes clear that nearly every domain can be supported by adaptive ITS behaviors. While these ITS behaviors might not match the interventions characteristic of traditional ITS behavior (i.e. VanLehn, 2006), they can provide a wide variety of interventions that can be powered by alternative assessments or representations of the domain. When used systematically, hybrid ITS methodologies mean that multiple pedagogies can be combined that use different information and interventions for different tasks in the same domain.

This makes the concept of an "ill-defined domain" somewhat misleading. Even in "well-defined" domains, such as mathematics, problems such as Bertrand's Paradox (i.e., calculating probabilities based on integrals of a circle) continue to be debated and studied (Aerts & Sassoli de Bianchi, 2014). This makes the concept of trying to label domains as ill-defined or defining continuums of ill-definedness for a domain is not very productive. By comparison, identifying and categorizing different types of ill-defined tasks within a domain appears quite valuable because this allows building different types of interventions and user adaptation to the available information about each task. We advocate focusing analysis of ill-defined components and behaviors at the task level for future research. Explicit modeling of each component should help identify correlations or patterns that grow into generalizable principles for both ITS design and other types of instructional design.

Such future research will certainly be needed for ill-defined domains, since there are a large number of unresolved questions about how to handle ill-defined domains. Some of these questions are theoretical, such as proposed guidelines about how to tutor domains which have many interconnected tasks that are all impacted by the same actions. Another major question is how to identify stopping rules for building an ITS for a domain: when is a domain so fundamentally ill-defined that its scope or very existence should be challenged (e.g., tasks and/or goals are arbitrarily and unstably-related)? Tasks that are ill-defined due to multiple interacting learners (e.g., team learning) also need further study and integrating findings from computer-supported collaborative learning research. Other questions are practical, such as what pedagogical techniques can be efficiently authored and maintained for ill-defined domains (particularly ones where the required skills change rapidly). The long-term goal of exploring ill-defined tasks should be to move beyond the general label of "ill-defined" to establishing the specific information and affordances available for user assessment and pedagogical strategies. This mapping of task characteristics to available pedagogy would be a significant advance for the field, enabling instructional designers to rapidly understand how an ITS can interact with a learner (or learners) based on the tasks that are taught. Such a mapping would be a major asset for authoring tools and for automating ITS interactions, since it would establish boundary conditions for pedagogical strategies and clarify the possible roles of an ITS in different strategies and learning tasks.

REFERENCES

Aerts, D. & Sassoli de Bianchi, M. (2014). Solving the hard problem of Bertrand's paradox. *Journal of*

Mathematical Physics **55**: 083503, doi:10.1063/1.4890291

Aleven, V., Mclaren, B. M., Sewall, J., & Koedinger, K. R. (2009). A new paradigm for intelligent tutoring systems: Example-tracing tutors. *International Journal of Artificial Intelligence in Education, 19*(2), 105-154.

Alexander, P. A. (1992). Domain knowledge: Evolving themes and emerging concerns. *Educational Psychologist*, *27*(1), 33-51.

Ackoff, R. L. (1981). The Art and Science of Mess Management. *Interfaces*, *11*(1), 20-26. doi:10.1287/inte.11.1.20

Baker, R. S., & Corbett, A. T. (2014). Assessment of robust learning with educational data mining. *Research & Practice in Assessment, 9*(2), 38-50.

Bellman, R. (1957). A Markovian Decision Process. *Journal of Mathematics and Mechanics 6*, 679-684.

Chesler, N. C., Arastoopour, G., D'Angelo, C. M., Bagley, E. A., & Shaffer, D. W. (2013). Design of professional practice simulator for educating and motivating first-year engineering students. *Advances in Engineering Education*, *3*(3), 1-29.

Easterday, M. W., Aleven, V., Scheines, R., & Carver, S. (2011). Using Tutors to Improve Educational Games. In G. Biswas, S. Bull, J. Kay, & A. Mitrovic (Eds.), *Artificial Intelligence in Education* (Vol. 6738, pp. 63-71): Springer Berlin Heidelberg.

FitzPatrick, B., Hawboldt, J., Doyle, D., & Genge, T. (2015). Alignment of Learning Objectives and Assessments in Therapeutics Courses to Foster Higher-Order Thinking. *American journal of pharmaceutical education*, *79*(1), 1-8.

Fletcher, J. D. (2011). DARPA Education Dominance Program: April 2010 and November 2010 Digital Tutor Assessments. *IDA Document NS D-4260*. Alexandria, VA: Institute for Defense Analyses.

Fournier-Viger, P., Nkambou, R., & Nguifo, E. (2010). Building Intelligent Tutoring Systems for Ill-Defined Domains. In R. Nkambou, J. Bourdeau, & R. Mizoguchi (Eds.), *Advances in Intelligent Tutoring Systems* (pp. 81-101): Springer Berlin Heidelberg.

Fournier-Viger, P., Nkambou, R., Mayers, A., Nguifo, E., & Faghihi, U. (2011). An Hybrid Expert Model to Support Tutoring Services in Robotic Arm Manipulations. In I. Batyrshin & G. Sidorov (Eds.), *Advances in Artificial Intelligence* (Vol. 7094, pp. 478-489): Springer Berlin Heidelberg.

Friedrichsen, P. J., Abell, S. K., Pareja, E. M., Brown, P. L., Lankford, D. M., & Volkmann, M. J. (2009). Does teaching experience matter? Examining biology teachers' prior knowledge for teaching in an alternative certification program. *Journal of Research in Science Teaching, 46*(4), 357-383.

Gipps, C. V. (1994). *Beyond testing: Towards a theory of educational assessment*: Falmer Press.

Grele, R. J. (1981). Whose Public? Whose History? What is the Goal of a Public Historian? *The Public Historian, 3*(1), 40-48.

Goldin, I. M., Ashley, K. D., & Pinkus, R. L. (2001). Introducing PETE: computer support for teaching ethics. In *Proceedings of the 8th international conference on Artificial intelligence and law, St. Louis, Missouri, USA* (pp. 94-98). ACM.

Gutierrez-Santos, S., Mavrikis, M., & Magoulas, G. (2010). Layered Development and Evaluation for Intelligent Support in Exploratory Environments: The Case of Microworlds. In V. Aleven, J. Kay, & J. Mostow (Eds.), *Intelligent Tutoring Systems* (Vol. 6094, pp. 105-114): Springer Berlin Heidelberg.

Hoffman, R. R. (1998). How can expertise be defined? Implications of research from cognitive psychology. *Exploring expertise*, 81-100.

Ichikawa, M., Koyama, Y., & Deguchi, H. (2010). Virtual City Model for Simulating Social Phenomena. In K. Takadama, C. Cioffi-Revilla, & G. Deffuant (Eds.), *Simulating Interacting Agents and Social Phenomena* (Vol. 7, pp. 253-264): Springer Japan.

Jacovina, M. E., Snow, E. L., Dai, J., & McNamara, D. S. (2015). Authoring Tools for Ill-defined Domains in Intelligent Tutoring Systems: Flexibility and Stealth Assessment. In R. A. Sottilare, A. C. Graesser, X. Hu, & K. W. Brawner (Eds.), *Design Recommendations for Intelligent Tutoring Systems: Volume 3: Authoring Tools and Expert Modeling Techniques*, (pp. 109-121). U.S. Army Research Laboratory.

Kalman, R. (1959). On the general theory of control systems. *IRE Transactions on Automatic Control*, *4*(3), 110-110.

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into practice*, *41*(4), 212-218.

Lane, H. C., Hays, M. J., Core, M., Gomboc, D., Forbell, E., Auerbach, D., & Rosenberg, M. (2008). Coaching intercultural communication in a serious game. In *Proceedings of the 16th International Conference on Computers in Education (ICCE 2008)*, (pp. 35-42).

Lazar, T., & Bratko, I. (2014). Data-driven program synthesis for hint generation in programming tutors. In *Proceedings of the 12th International Conference on Intelligent Tutoring Systems* July 2014, *Honolulu, HI*. (pp. 306–311).

Le, N. T., Loll, F., & Pinkwart, N. (2013). Operationalizing the Continuum between Well-Defined and Ill-Defined Problems for Educational Technology. *IEEE Transactions on Learning Technologies*, *6*(3), 258-270.

Lesgold, A., Lajoie, S., Bunzo, M. & Eggan, G. (1991). SHERLOCK: A coached practice environment for an electronics troubleshooting job. In J. Larkin & R. Chabay (Eds.), *Computer-assisted instruction and intelligent tutoring systems: Shared goals and complementary approaches* (pp. 201-238). Hillsdale, NJ: Lawrence Erlbaum Associates.

Lynch, C., & Ashley, K. (2014). Empirically valid rules for ill-defined domains. In *Educational Data Mining (EDM) 2014*, (pp. 237-240). London, U.K.

Lynch, C., Ashley, K.D., Pinkwart, N., & Aleven, V. (2009). Concepts, structures, and goals: Redefining ill-definedness. *International Journal of Artificial Intelligence in Education, 19*(3), 253-266.

Matsuda, N., Cohen, W. W., Sewall, J., Lacerda, G., & Koedinger, K. R. (2007). Evaluating a Simulated Student Using Real Students Data for Training and Testing *User Modeling 2007* (pp. 107-116), Springer.

Mitrovic, A., Koedinger, K., & Martin, B. (2003). A Comparative Analysis of Cognitive Tutoring and Constraint-Based Modeling. In P. Brusilovsky, A. Corbett, & F. de Rosis (Eds.), *User Modeling 2003* (pp. 313-322). Springer Berlin Heidelberg.

Mitrovic, A., & Ohlsson, S. (2015). Implementing CBM: SQL-Tutor After Fifteen Years. *International Journal of Artificial Intelligence in Education*, 1-10. doi:10.1007/s40593-015-0049-9

Mitrovic, A., & Weerasinghe, A. (2009). Revisiting ill-definedness and the consequences for ITSs. *Proceedings of the 14th International Conference on Artificial Intelligence in Education* (AIED 2009). V. Dimitrova and R. Mizoguchi (Eds), 375-382. IOS Press.

Nash, G. B., Crabtree, C. A., & Dunn, R. E. (2000). *History on trial: Culture wars and the teaching of the past* (2nd ed.). New York, NY: Vintage.

Norman, D. A. (1983). Some observations on mental models. *Mental models*, *7*(112), 7-14.

Nye, B., Bharathy, G., Silverman, B., & Eksin, C. (2012). Simulation-Based Training of Ill-Defined Social Domains: The Complex Environment Assessment and Tutoring System (CEATS). In S. Cerri, W. Clancey, G. Papadourakis, & K. Panourgia (Eds.), *Intelligent Tutoring Systems* (Vol. 7315, pp. 642-644): Springer Berlin Heidelberg.

Nye, B. D., Goldberg, B., & Hu, X. (2015). Generalizing the Genres for ITS: Authoring Considerations for Representative Learning Tasks. *Design Recommendations for Intelligent Tutoring Systems: Volume 3: Authoring Tools and Expert Modeling Techniques* (pp. 47-64). U.S. Army Research Laboratory.

Ososky, S. J. (2013). *Influence of Task-Role Mental Models on Human Interpretation of Robot Motion Behavior* (Doctoral dissertation, University of Central Florida, Orlando, Florida).

Pinkwart, N., Aleven, V., Ashley, K., & Lynch, C. (2007). Evaluating legal argument instruction with graphical representations using LARGO. In *Frontiers in Artificial Intelligence and Applications, 158*, (pp. 101-108). Springer: Berlin.

Pinkwart, N., Ashley, K., Lynch, C., & Aleven, V. (2009). Evaluating an intelligent tutoring system for making legal arguments with hypotheticals. *International Journal of Artificial Intelligence in Education*, *19*(4), 401-424.

Richards, N., Moriarty, D., & Miikkulainen, R. (1998). Evolving Neural Networks to Play Go. *Applied Intelligence, 8*, 85-96, Rittel, H. & Webber, M. (1973). Dilemmas in a general theory of planning. *Policy Sciences 4*(2), 155-169 (1973)

Ritter, S. (2015). Authoring for the product lifecycle. *Design Recommendations for Intelligent Tutoring Systems: Volume 3: Authoring Tools and Expert Modeling Techniques* (pp. 137-144). U.S. Army Research Laboratory.

Roberts, N. (2000). Wicked problems and network approaches to resolution. *International Public Management Review 1*(1), 1-19.

Scheuer, O., McLaren, B. M., Loll, F., & Pinkwart, N. (2012). Automated analysis and feedback techniques to support and teach argumentation: A survey. *Educational technologies for teaching argumentation skills*, 71-124.

Schraagen, J. M., Chipman, S. F., & Shalin, V. L. (2000). *Cognitive task analysis*. Mahwah, NJ: Lawrence Erlbaum Associates.

Silverman, B.G., Bharathy, G.K., Nye, B.D., Kim, G.J., Roddy, M., & Poe, M. (2010) M&S methodologies: A systems approach to the social sciences. In Sokolowski, J.A. & Banks, C.M. (eds.) *Modeling and Simulation Fundamentals: Theoretical Underpinnings and Practical Domains*, (pp. 227-270). Wiley & Sons, Hoboken, NJ.

U.S. Army (2006). *Counterinsurgency Field Manual 3-24*. Department of the Army.

Vanlehn, K. (2006) The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education 16*(3), 227-265 (2006)

Waldmann, M. & Dieterich, J. (2007) Throwing a bomb on a person versus throwing a person on a bomb. *Psychological Science 18*(3), 247.

Wiggins, G. P., & McTighe, J. (2005). *Understanding by design* (2nd Ed.): Pearson.

Woods, A., Stensrud, B., Wray, R. E., Haley, J., & Jones, R. M. (2015). A Constraint-Based Expert Modeling Approach for Ill-Defined Tutoring Domains. In *Florida Artificial Intelligence Research Society (FLAIRS) 2015* (pp. 469-474). AAAI Press: Hollywood, FL.