May 01, 2006

Shallow acquisition of domain specific concepts

David F. Baldwin

Northeastern University

Shallow Acqusition of Domain Specific Concepts by Naive Learners

David F. Baldwin

April 23, 2006

1 Introduction

Quantitative problem solving is a highly valued - and hard earned - skill within a number of different fields and domains. Whether used in scientific insight, engineering prowess or otherwise, the ability to solve novel and complex problems is an important and unique aspect of human thought. The breadth and depth of what we can learn over the course of time and practice is enormous. Most interesting, however, is how, with increasing knowledge, humans are able to better generalize and adapt to a given problem or context.

Skilled reasoners in a variety of domains - be it mathematics, literature, architecture or physics - are able to abstract a given problem from its irrelevant features, represent the problem by mapping it in terms of an adaptation some known schema and then solve the task. To say this process is non-trivial, even for the most knowledgeable and intelligent, would be a severe understatement. Learning the process of solving problems within a particular area takes years of practice, repetition and failure. Many, even with a considerable amount of training through schooling or otherwise, fail to develop these skills beyond a certain rote competency.

A great deal of problem solving studies have focused how subjects perceive and attack a problem. Classic problem solving studies studies(Chi et al., 1981) established the idea of 'deep structure' in a problem - loosely defined as the underlying principles by which the problem can be solved. Their finding that unskilled undergraduates categorize physics problems by their irrelevant surface features while more experienced students tend to group by the underlying physical concept(e.g. centrifugal force) showed a stark difference in the conceptual usage of these two groups.

Recognizing the deep structure of a problem is extremely difficult. While Chi et al.(1981) delineated this distinction, several further studies have shown how truly difficult this recognition can be. Students in and out the laboratory are often easily confused by distracting features in simple problems. For example, subjects tested on isomorphic instances of the same problem often fail miserably as the distance(e.g. the transformation required to map the technique) between novel problems and known instance increases (Kotovsky et. al, 1985).

The problem then, becomes how to facilitate this behavior. Chi et al. (1989), among others, found that students are most successful in learning when they explicitly attempt to compare and observe the relations between different problems(Kurtz, 2001, Pirolli,1994). Despite manipulations based on these ideas, however, the majority of students rarely spontaneously activate the cognitive processes needed to discover the associations between novel problems and previously learned material(Sandberg et al. 1997). Even when explicitly told that a problem is an instance of a previously seen concept, subjects often have difficulty recognizing how the current problem can be framed in the light of some known concept. (Pretz el al., 2003, Gentner et al. 2004).

The difficulties described above are a reflection of learners' propensity to integrate knowledge with irrelevant features. To understand why this phenomena is so prevalent, it is important to comprehend what types of knowledge might facilitate deeper problem solving ability. This begs the fundamental question - how does one's knowledge affect how a novel task is examined. One prevalent and intuitive view is that there is an incremental prerequisite structure - certain knowledge is required or helpful to learn novel topics (e.g. one must master basic mathematics before attempting set theory).

It is less clear, however, whether knowledge outside of these prerequisite topics affects learning. In other words, does 'related' knowledge affect conceptual acquisition as well? For example, how does one's knowledge of physics, biology or even chess influence how they learn about mathematical concepts.

To begin to answer this question, several goals must be accomplished. First, in section 2.1 a more rigorous framework for describing knowledge and its topology(i.e. 'related' vs. 'necessary' knowledge alluded to above) is developed. One way of describing this organization of knowledge is through the idea of domain specificity. This idea postulates that knowledge is domain specific and encapsulated - e.g. knowledge about physics is generally not applied to tasks other than physics.

Using this framework, we use the relational structure of concepts to further describe the differences between experts and novices in section 2.2 . Within a particular domain, it is shown that experts have structured and organized concepts that guide abstracting a task from it distracting aspects. Novices, however, tend to lack this cohesive structure between different concepts, leading to their reliance on surface features.

Finally, we describe in section 2.3 analogical transfer and how it operates over this relation structure. By doing so, additional parallels are drawn between expertise and relational structure. Furthermore, we describe one of the mechanisms underlying the generalization and application of existing knowledge to novel contexts.

These three tools allow us to develop a novel hypothesis on how concepts in a particular domain influence those in other domains. Specifically, we postulate that the presence of rich conceptual structure outside of a domain may facilitate transfer to concepts within this domain. This characterization allows for the development of an empirical test of these cross domain influences.

The rest of this study is structured as follows. Section 2 further explores the theoretical framework discussed above. Section 3 introduces a methodology for examining this phenomena in novice reasoners. The results of the experimentation are then presented in sections 4 and 5 followed by a discussion of their implications and ties to problem solving, conceptual structure and expertise in general in section 6.

2 Background

In discussing the idea of the influences of different domains, it becomes useful to introduce a concise notation to refer to knowledge within and outside particular domain. The primary domain in question will be referred to as D_1. A particular secondary set of knowledge outside that domain is referred to as D_2. For example, if discussing the influence of concepts in the domain of physics on those in the domain of math, D_1 refers to math and D_2 to physics. This convention is used throughout the paper.

2.1 Domain Specificity and Expertise

The multifarious contexts in which we think and reason can be considered even more incredible given that we must not only encode, but retrieve and use a vast amount of information. A particular context or task leads us to cognize some minute subset of that information. By and large, the mind is agile at accessing the knowledge germane to the particular task at hand. When mentioning some token object, such as 'pepperoni pizza', or drawing a strange analogy (e.g. 'street pizza'), it takes little effort to identify context and access only the particularly salient knowledge that might aid in comprehension. While far from perfect, concepts and information are, by and large, fluidly retrieved from memory, applied and perhaps altered due to learning.

This incredible ability has lead to one of the most profoundly difficult questions - how, in fact, does the mind go about accessing different information given some context. This, known more generally as the frame problem(Dennett,1984), has troubled legions of scholars in philosophy of mind, artificial intelligence and psychology. While the frame problem has several definitions across these studies, its essence lies in two issues. First. the mind deals with a incredible amount of information and it is not plausible that the everything can be considered at any given period of time. Second, attempting to devise a procedure that solves every conceivable problem - or even a limited set - is extremely difficult. Many prominent attempts (e.g. (Newell and Simon,1972) among many others) have ended in abject failure and no current computational metaphor has the power to explain such an ability.

There are myriad of responses to the frame problem, but the general response has been to deny that the existence of these sorts of domain general processes. Rather, instead, knowledge is organized in to modules or domains situated to solve only a small portion of these tasks. It is, for the most part, widely accepted that both knowledge and thought have some degree of domain specificity rather than some very general process that subsumes most human activity.

Domain specificity, then, implies that certain types of knowledge are used for different tasks. In other words, there must exist some degree of information encapsulation within a domain - the concept of 'fish' will most likely not be activated in the same context as 'titanium'. While there is little support for the existence of concrete and mutually exclusive domains - it is ostensibly the case that only some subset of one's knowledge is activated at a given time.

Individual theories of domain specificity differ greatly in their scope and definition of a domain. Modularity theory(Chomsky, 19–; Fodor, 19–) specifies innate and immutable(unchanged relative to experience) perceptual modules. Regardless, there is strong evidence of domain that domain specific reasoning exists in a variety of non-perceptual contexts. The examples are numerous and usually controversial - theories are often reliant on evidence from domain-specific contexts - for example, physics principles aren't applied to mental states (McCloskey, 1982), neurological localization of navigational behavior(Burgess,1996), and even moral sense(Greene,2001).

Furthermore, expertise is another prominent source of evidence for domain specific reasoning. The expert, be in chess or botany, develops an extremely intricate and related web of knowledge. It is obvious that the botany expert will have little expertise in chess and vice-versa. Expertise is necessarily limited to some narrow scope and individuals have great variability of cognitive ability across particular domains. Several studies have explicated this further - e.g. experts' memory in their domain of expertise is far higher than for raw digit sequences or tasks other domains(Chi,1978). Domain-specific expertise may extend to lower-level perceptual processes as well, such as spatial abilities (Sims,2002) and face perception (Kanwisher,2000).

It is clear from such work that expertise is acquisition of domain specific knowledge. As such, a domain gives a specific characterization of the knowledge that is likely to be activated or brought to memory given some task in that domain. To frame the present research question in terms of domain specificity, we wish to characterize the boundaries(or, the degree of information encapsulation) of a domain(D_2) as a function of expertise to determine how it might affect some other, closely related domain(D_1).

It is fairly obvious that domains are not information disjoint. However, if information encapsulation within a domain is soft - to what degree can knowledge within a domain generalize to knowledge outside of that domain? While the information supporting botany and chess are close to mutually exclusive, this may not be the case with domains containing a higher degree of cross-coherence(similar domains). More concretely, can the concepts associated with the domain of e.g. canines(D_1) be transfered to something in the domain of fish? automobiles? physics? (D_2)

Throughout this paper, the word 'domain' is used liberally. In this context, it is useful to define a domain as any subset of concepts within an ontological system. This is most certainly congruent with many of these multitudinous definitions of a domain. Whether a domain is discussed in terms of domain-specific theories(e.g. the theory-theory (Gelman and Hirschfeld,1994)), modularity or otherwise, each posits some specific assumptions about the nature of a domain. It is easy to see however that this extremely broad definition subsumes whatever segmentation these previous theories discuss. This is done not to provide an overarching explanation, but for ease of discussion and independence of the nuance of each particular theory. Additionally, this assumption motivates the experimental task, one that might be applicable to more concrete definitions.

2.2 The importance of relational structure

Concepts are generally thought to be domain specific and activated when appropriate to the context or task, However, concepts do not exist in isolation. When acquiring new information, it must be associated with other, existing knowledge to have relevance. Whether learning a simple procedure or a complex task, information must be associated and the how this occurs plays a major role in future use. Whether a subject relates a physics problem involving a inclined plane to a principle of motion or its shape of a wedge of swiss cheese, it is related to previously acquired knowledge and integrated into the ontological structure. In other words, the interrelatedness of concepts (also referred to as 'relational categories' in some contexts) are exceedingly important for the sort of abstract reasoning necessary in problem solving.

How knowledge is related in a conceptual system is unclear and there exists no simple definition of a 'relation'. It has been suggested that there is a spectrum defining how well concepts may be understood in relation to other concepts. On one side of this spectrum lies 'entity' concepts - those comprehended well by themselves or with the use of relatively few other concepts. The other extreme involves 'relational' concepts highly dependent on associative structure for meaning (Goldstone 1996). Independently, Gentner(1981) and Kloon and Sloutsky(2004) have made a distinction between relationally dense and sparse concepts. For example, 'acceleration' and 'bird', respectively(roughly corresponding to relational and entity concepts, respectively). Other research(Markman & Stilwell, 2001) attempted to delineate four different types of categories, one of which were relational, corresponding to both dense/relation concepts.

While concrete characterization of this spectrum's extremes and intermediate levels is still in progress, it is apparent that certain concepts acquire their meaning based on how they are associated with other concepts. There is strong support that these are especially prevalent in scientific and mathematical domains such as problem solving(Anggoro et. al, 2005).

The ABSURDIST model(Goldstone and Rogosky, 2002) has further elucidated the informativeness of relations in cross-system(e.g domain, person, time, etc.) conceptual alignment. Through a series of computational experiments, the model is able to determine a mapping between two conceptual systems(e.g. a novel set of concepts and existing knowledge, or two domains such as D_1 and D_2) based solely on degree of relatedness between each concept. That is, mappings between a source and target domain can be determined in absence of any sort of entity or relational labeling whatsoever. Because the model works solely on relational aspects, a sparse conceptual structure is less accurately mapped than a strongly connected one. Knowing labels for each of these concepts would only make this process more fluid, but the experiment clearly shows the importance of relational information, especially in cases of uncertainty, such as during learning.

Of course, this structure can change dramatically and most likely will during the course of learning. Whether by relation to a new piece of information acquired at a later time, or through large scale conceptual changes(Carey, 1981) - ontological changes of various scales obviously occur during learning. Indeed, expertise is not simply the accretion of knowledge, but how it is structured.

Additionally, there is strong evidence that context sensitivity in experts is due to dense and salient relations. Ross and Murphy(1999) explicated how subjects reason about food and found they are easily able to use context sensitive relation reasoning. Experts in folk-biological domains also exhibit far greater context sensitivity than the general population(Shafto and Coley, 2003). It appears to be the case that expert knowledge is relationally rich as to support such context sensitive inferences. Furthermore, this structure may affect the salience of such relations when reasoning about a particular context.

From this specification, it is possible to see the relation between this structure and expertise. Consider once again subjects solving physics problems(e.g. Chi et al. 1983). While they have already been taught the domain knowledge necessary to attack a particular problem, they are unable to abstract the problem from its surface features. One possible interpretation is that that they lack the precise relational structure to map a novel problem, veiled by other features, to their existing knowledge. Experts, on the other hand, have dense, interconnected knowledge, which allows them to access the

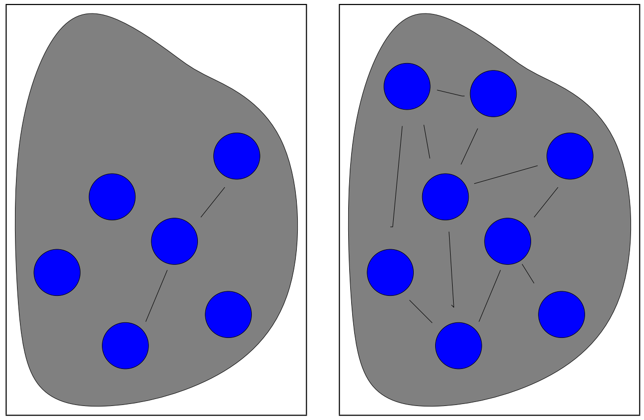

Figure 1: Comparing the conceptual structure of the expert and the novice. The novice(left), not only lacks certain pieces of knowledge within some domain and also has a sparse relational structure. The expert(right), has a dense web of interrelated concepts, which can be accessed in a context sensitive manner.

appropriate mapping in a fluid and context sensitive manner.

Figure 1 gives a rough graphical depiction of the differences between experts and novices. The grey circle represents some domain. The blue verticies and their associated edges represent concepts and relations, respectively.

Relational information is clearly an underpinning of expertise. When one begins to learn about that particular domain, their knowledge of it is fairly sparse - it exists somewhat in isolation. Thus, as new facets of those concepts are learned, the only grounding available is that of the surface features. As expertise increases, not only does one's breath of knowledge grow, but we also learn to associate these learned concepts together in new and novel ways. As seen above, the density and organization of these structural relations may directly predict skill in a particular domain.

We now turn to the mechanism that operates over these mappings - analogical transfer - in order to develop a more precise definition of how relational structure within a domain aids transfer of known concepts to some novel task - another behavior clearly indicative of skill in a domain.

2.3 Analogical transfer

Thus far, it has been shown that knowledge is likely to be organized in to specific domains. Additionally, the previous section has shown how knowledge is organized relative to other concepts in that domain may be indicative of expertise within that domain. This section further develops this notion by examining how relations between concepts is exploited via analogical transfer.

Analogical transfer - the mapping of existing knowledge to new domains - serves as one of the core components of conceptual mapping. While it is suggested that analogy is not the entirety of the process, it is an extremely useful framework for presenting the idea that expertise allows further generalization of knowledge. Analogical transfer also provides an excellent way to test the predictions made later in this section.

In the most general sense, analogy involves aligning a *target* conceptual system[1] to a *base* one - the latter corresponding to existing knowledge and the former to some novel instance. Roughly speaking, problem solving by analogical transfer takes place in the following fashion. First, mental representations of a relevant source domain and the target are brought to memory and attention. By finding a mapping between the features of each system, a correspondence between existing and novel knowledge allows previously learned concepts to be utilized (Holyoak,1985, Gentner, 1982).

Analogical transfer is particularly relevant in scientific domains. Indeed, some of the most important achievements in fields such as biology have come from successfully transferring concepts from other domains(Dunbar, 1995). Gentner (2002) showed that analogical reasoning is key to successful scientific thought by analyzing historical accounts of scientific discovery. However, the use of analogy is hardly limited to scientific domains - complex reasoning in other domains, e.g. politics (Dunbar,1997), has been demonstrated as well.

While analogy is an extremely broad area of research, its most important aspects for this discussion is its ability to provide relational structure to a novel concept. When acquiring a new piece of information or concept, learners are made aware of its features and properties. However - as hinted at in the introduction - they are given far less detail about how these concepts relate to each other and one's prior knowledge (Gentner and Wolff, 2000). If the base knowledge in a particular domain transfers to the novel concept,

[1] This is usually referred to as a domain outside this paper, but in order to disambiguate this from the larger scale domains we speak about here, we use the present terminology

the structure of the target knowledge can be inferred.

Most theory on analogical transfer postulates that general schema are abstracted from problems during learning(Holyoak, 1984; Holyoak, 1985), rather than storing specific exemplars. These schema are essentially interrelated concepts. Experts are able to relate these new concepts with previously acquired knowledge while novices generally simply associate them with surface features. Furthermore, use of relational concepts show greater variability based on prior knowledge than entity concepts(Gentner,2005). This suggests that the relational structure of a particular domain is highly related to its ability to be generalized - a richly related system of concepts is likely to be applicable to tasks beyond that which it is learned.

Several experiments on children's comprehension of metaphor support this idea.

First, comprehension of one metaphor between two different domains seems to greatly inform later metaphors between the same two domains. That is, once children learn how these domains are associated through an exemplar, they are able to use that to transfer further information between those domains(Kelly et al. 1986). The authors claim that this provides evidence that the large amounts of information must be juxtaposed between the two domains. Another slightly different interpretation is that only once a sufficient amount of relational knowledge between these domains(as a whole) is acquired can they be aligned and amalgamated in a new metaphor. Kelly and Keil(1987) also suggest that these sorts of bridges between two previously disparate domains increases the perceived similarity between them.

Expertise, specifically through relational structure, clearly plays a large role in the success or failure of analogical transfer. For analogical transfer to be successful, the structured relations between concepts must be present and salient. Consider when an expert learns a novel concept in their domain of expertise. They can integrate the concept within the variety of existing knowledge in the group of concepts related to that domain. Later, when attempting to use that system of concepts, there exists a wealth of structural information between the new problem and the rich system of concepts associated with it(figure 2.3a).

Now consider a novice reasoner. It is feasible to suspect that one's 'mundane' knowledge of the world - e.g. folk-physics, object identification - is commensurate with some level of expertise. This is partially why surface level features are so distracting. In particular, with novices, the salience of surface features that map to known concepts may be the only mapping

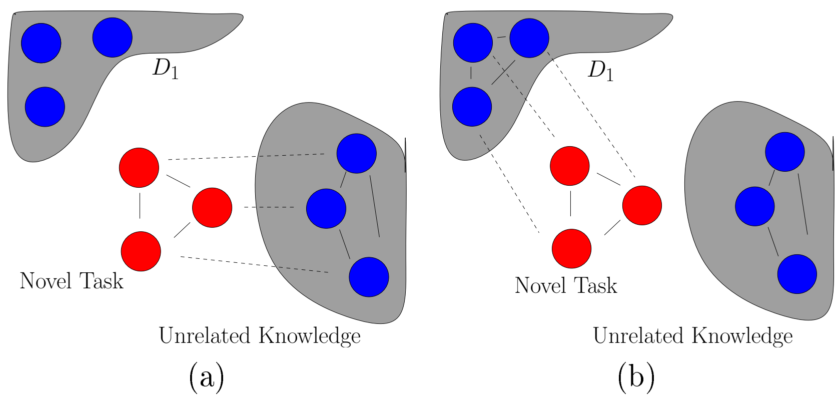

Figure 2: Difference in novice(a) and experts(b) when performing analogical transfer from a novel task(red) to the task domain D_1. Dashed edges represent an analogical mapping.

available or may be so salient that they discourage further consideration of alternate mappings(figure 2.3b)

The importance of structured knowledge is clear - it organizes information in fashion that allows it to be accessed, mapped and applied to novel domains in an efficiently and tractably. As has been shown above, the specific knowledge gleaned from experience - and the association of that knowledge in a the larger system of concepts - has a large influence on processes such as analogical transfer.

2.4 Analogy as a means of cross-domain transfer: one possible explanation

As previously demonstrated, when a student learns a novel concept, their expertise in D_1 plays a fundamental role in how well that concept is learned and later generalized. When later approaching a novel problem, the density and sophistication of both prior knowledge and it associated structure will determine how the novel is mapped to the. This mapping, in essence, is analogical transfer. It is clear that propositional relatedness of existing knowledge in D_1 will directly affect how successful this transfer is. The primary goal of this study, however, is to determine if other D_2 domains also affect this transfer.

While relatively few previous studies have attempted to address this specific problem, previous work does provide a good intuition when cross domain transfer might occur. Dunbar(1997) explored the use of cross-domain analogy in science and showed more than 90% of analogies used in the biology laboratories are within-domain(biology) and demonstrated that cross-domain reasoning is fairly rare in these situations. It does appear fairly rare that successful transfer of knowledge occurs between domains, though notable examples show highly significant scientific discover as cross-domain transfer(Gentner, 2005).

It appears that both experts and novices make relatively few cross-domain analogies(as novices are apt to make surface level analogical mappings) when attacking a new problem. This may be an illusory distinction - consider an expert in some domain. A biologist has an extremely rich set of knowledge within D_1(here, biology). The accumulation of years of experience and acquiring the argot of the discipline makes it facile for an expert to integrate a new idea with their existing domain knowledge. In other words, they can associate the new information with a rich, salient conceptual structure and rarely need to go outside the particular domain in question. When expertise is high, many of the related tasks are strictly within the particular domain of study. Only when there is some dissonance between their conceptual framework and some observed evidence does cross-domain thought become necessary. [2].

Novices, as noted in the previous section, lack this structure within D_1. The most salient aspects of a problem, then, may be the surface features, which can be associated with a salient domain simply by virtue of being something that might be encountered naturally through daily interaction with the world.

If experts rarely exhibit cross-domain transfer, and naive novices are unable to attack particular tasks rarely make the correct transfer due to their lack of structured knowledge, when might such learning occur? Cross-domain transfer may indeed be rare, but it may also be the case that it occurs when rich conceptual structure exists in a related domain(D_2). Suppose that a novice in one domain has particular expertise in some other domain. One possible way that the novel information could be correctly mapped is *through* D_2.

[2] It is interesting to note that this is somewhat related to the notion of incommensurability(Kuhn, 1970).

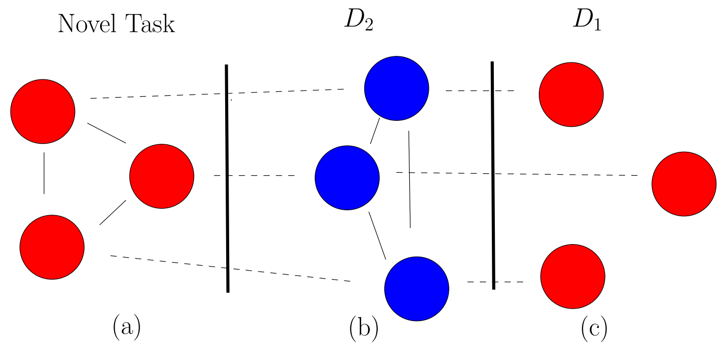

Figure 3: Mapping between two conceptual systems((a) and (c)) via a third(b). In the context of this study, (a), (b) and (c) refer to some novel concept, D_2 and D_1, respectively

To illustrate this, consider the mapping between three different conceptual systems in figure 3. Suppose there is a mapping between concepts in systems (a) and (b), as well as between (b) and (c). It may be the case that, in isolation, transfer between (a) and (c) is not feasible. However, mapping the structure of (b) on to the analogous concepts in (c) might allow for this to occur. In essence, (b) acts an an intermediate stage which promotes transfer between the previously unmappable conceptual systems.

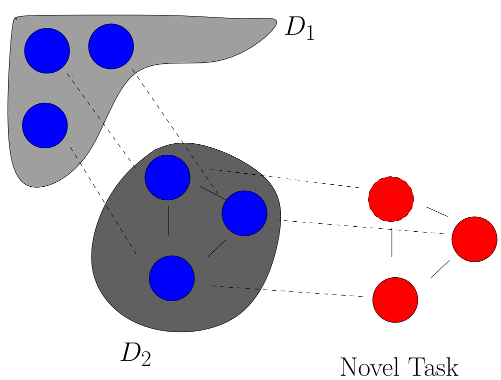

Figure 4: The mapping of two conceptual systems with additional D_2 knowledge

This is only one hypothetical fashion in which this might occur, others may exist. The primary idea, however, is that knowledge outside a particular domain or conceptual system may certainly influence the mapping of some instance to that domain. The correspondence to domain specific reasoning is clear - a rich conceptual system, D_2, *outside* the domain of the problem(D_1) may influence transfer *to* that domain(Figure 2.4). As alluded to previously, the relatedness of these two domains plays a enormous role in how successful this transfer might be - again, it is extremely unlikely such a mapping would occur between chess and botany. It is unclear whether this is due to the fact that 'unrelated' domains are structurally distinct or that spontaneously making these mappings is unlikely.

Even if there is some degree of information encapsulation, unless knowledge is segregated in the mutually exclusive sense, there is some degree of transfer between two domains. This section has given a concrete example - though analogical transfer - of how such mappings might occur. The experiment that follows is a task to verify the existence of cross-domain transfer - verifying this particular explanation is beyond the scope of this study.

3 Method - Testing the strength of cross domain influences

3.1 General Overview

The previous sections have provided some evidence that there may be cross-domain influences in analogical reasoning and problem solving. We now devise a procedure to observe the effects of such transfer. The primary goal, then, is to first demonstrate that extra-domain concepts can influence transfer on certain tasks. We do this by testing differentially knowledgeable subjects on a domain novel to their experience.

Intuitively, it is more likely that cross-domain affects will be seen in subjects with a structured set of information about a given domain - the knowledge usually found in experts. One possible way to perform this experiment is to take two pools of experts - one with expertise in a primary domain(D_1) another with expertise in secondary domain(D_2) Their performance in the primary domain D_1 could then be compared on some measure of skill. This could be repeated for a number of different domains to develop a better quantification of the gradient of cross-domain influence in experts.

This study, though, focuses on naive learners as an initial test of these hypotheses. By naive learners, it is meant that neither of our two groups have extensive knowledge in domain D_1. By teaching them a concept in D_1 and then measuring their performance as a function of their knowledge in D_2, we can accomplish a similar goal.

This is done for two primary reasons. First, this study is motivated by pedagogical concerns, so our natural test bed should be students rather than sophisticated experts in a particular domain. Second, for this initial study, it is suspected that it may be easier to tease such cross-domain behavior from unsophisticated subjects. Expert knowledge is likely to be highly variable, whereas little or no knowledge of D_1 is found in our subjects.

To test the efficacy of cross-domain influences on naive learners, it is necessary to find a set of tasks that is accessible enough for novice subjects to comprehend and succeed in without exhaustive training. Simultaneously, the tasks must be sufficiently complex so that the more experienced reasoners are not performing at a ceiling level for all tasks.

This goals appear contradictory, but it is possible, as shown below, to devise a procedure that accomplishes these goals. This study uses simplified combinatorial optimization as is primary domain(D_1) and discrete math as D_2. The experiment compares an 'expert' group of students with general background in discrete math with a 'novice' grouping of subjects from the general population.

After training subjects on a concept in D_1, subjects are asked to solve 6 problems of increasing difficulty. Three different training conditions aim to dissociate the differences between cross domain effects and simply the presence of extra experience in the experts. The problems then, in essence, aim to determine the depth to which the concept has been learned. By measuring performance as a function of both training and expertise, a more complete picture of cross-domain influence between these two domain is painted.

Each of the subsequent sections overviews each these items in turn followed by a description of the experiments themselves.

3.2 Optimization problems

To test these hypotheses, it is necessary to use a set of stimuli that can be related to the experts' body of knowledge while still being comprehensible and solvable by the 'novice' subject pool. Discrete combinatorial optimization problems have several advantageous aspects. First, they are easily grounded in realistic problems and are often based on idealizations of everyday tasks. That is, comprehension the problem itself should be of little trouble to college-aged subjects independent of their quantitative ability or other knowledge.

For example, many graph theoretic problems have obvious applications. Finding the shortest or longest path in an undirected graph might be related to navigation through a city. One might also relate needing the minimum spanning tree of a graph to discovering the minimum amount of cable required to connect several appliances. Fortunately, most optimization problems are easily adapted to some tangible and familiar representation.

Solving such problems is strongly related to basic algorithmic knowledge such as recurrence relations and asymptotic analysis(topics typically covered in a discrete math course - this experiment's D_2) and it is expected that experience in such areas would lead to more principled methods of problem solving. However. it is also not unrealistic to expect subjects without this knowledge to attack the task successfully(even without instruction) especially in small instances of the problem where total enumeration of the state space can be computed easily.

Also, the generality of optimization problems are fundamental in measuring performance and obfuscating the underlying principle of a problem. Not only is it possible to easily change the problem description to obscure the technique necessary to solve the task, making it more difficult to apply a particular strategy or algorithm, but certain optimization problems can be altered in subtle ways to change how they must be solved.. Returning to a previous example, common algorithmic strategies for finding the shortest path in a graph(Johnson, 1977) are inapplicable to the longest path in the same graph, a problem which is computationally intractable (Garey and Johnson, 1979). Other simple manipulations, such as changing the size of the problem instance, provide simple and independent ways to compute how ability is related to a specific problem.

Finally, the performance of subjects directly solving these tasks is easily quantified . Human performance in optimization of well-defined problems(e.g. those containing full knowledge of the underlying state space) can be compared to the optimal answer. While optimization problems are far from ubiquitous in problem solving research, several recent studies involving human performance of NP-complete optimization have suggested that performance on such problems correlates with problem solving ability(Vickers, 2003).

As briefly mentioned above, the experiment must make use of a specific concept related to optimization. This study uses the *greedy algorithm*. a concept applicable to certain optimization problems. In short, optimization problems can be solved with a greedy algorithm if they can solved optimally by choosing the best local solution at each step. For example, in the context of shortest paths, the path with the lowest cost is selected first. There may be several possible greedy algorithms for a given optimization problem, only some of which may be correct. Furthermore, in other optimization problems greedy choice leads to non-optimal solutions(Cormen et al, 2001).

Greedy algorithms are part of a far more elaborate theory of matriods (Edmonds, 1971) , which are founded in the larger theories still of optimization problems. Greedy choice is somewhat intuitive and comprehension of a shallow version of its idea - choose the immediate best choice - should be achievable by most of the novice subjects. More importantly, greedy algorithms are subsumed by a technique known as dynamic programming, which will be important when describing how the problems are structured - they permit subtle manipulations to determine how well subjects have learned.

3.3 Training

Since we wish to test subjects in a primary domain D_1 in which they have little experience as well as measure how they integrate a novel concept with their existing knowledge, it is necessary to train subjects on the greedy algorithms in which they will be tested.

Subjects are first trained using one of three possible techniques - no training and two differential types of training. It is necessary to compare performance to the no training baseline to ensure that subjects have little domain knowledge in D_1 without training. Moreover, we hope to compare two different types of training to further delineate the difference between the expert population having additional experience in D_1.

Two distinct training regimens are utilized. The *procedural* training describes a procedure to solve each of the training problems. That is, the explicit algorithm to solve each problem is presented to subjects. The *conceptual* training actually teaches the idea of a greedy algorithm rather than a rote procedure. If there are actually cross-domain effects, subjects should perform equally well regardless of expertise on the no training . Additionally, the procedural training should show little differences. However, the critical condition, the conceptual training, should show an interaction between performance and expertise.

One valid concern is the development of these training techniques and controlling whether differences are due to the type of training rather than the training itself(e.g. the procedural training is simply 'better'). It should be noted that tuning these training methods would be a study in of itself, so we make several simplifying assumptions. The conceptual and procedural trainings are such that, whenever possible, the same wording is used. Both also contain identical diagrams and both training methods are of roughly the same length. The full training for each group is provided in Appendix A.

3.4 Stimuli

Six increasingly difficult problems are used in order to determine how the subjects generalize what they have learned. The problems are designed be orthogonal in a three dimensional feature space. Each dimension represents a facet of the problem, a superficial dimension and two 'deeper' aspects. The particular dimension which the subject attends to should determine how well they perform on the task.

This design is a natural extension of the multitude of experiments exploring expertise by e.g. Chi et al(see section x for a summary). Instead of distinguishing(and, implicitly, solving) problems by some 'expert' level features that indicate recognition of deep structure, there are three different levels of abstraction. This allows us to better distinguish our two subject groups as well as ensure neither are performing at a ceiling.

More specifically, the design involves three tiers of problems, each more difficult than the previous(relative to the training). Difficulty is manipulated by changing the orthogonal features of each problem so that it is possible to expose what particular aspect of the problems the subjects are using. The three features are as follows. A 'surface metaphor' is used to veil the problems in a real world metaphor used to distract subjects from the true nature of the problem. Next, the 'algorithmic method' describes the nature of the specific algorithm to be used. Roughly speaking, it is an adaptation of this procedure that must be used to solve the problem. Finally, the problems in each level may or not be solvable by greedy methods - the hardest problems are not. The problems and their orthogonal features are summarized in table 3.4 [2] .

[2] Note that 2 elements of the 3 dimensional feature space are omitted - i.e. the non-greedy problems with swapped metaphors. It was assumed(and verified) that the hardest levels are sufficiently difficult

Difficulty	**#**	**Metaphor**	**Method**	**Greedy?**
Easy(Training)	1	Packing	Knapsack	Yes
	2	Scheduling	Interval Scheduling	Yes
Medium	3	Packing	Interval Scheduling	Yes
	4	Scheduling	Knapsack	Yes
Hard	5	Packing	Knapsack	No
	6	Scheduling	Interval Scheduling	No

Figure 5: A list of the problems and their orthogonal features.

One of the most interesting aspects of this design is that attention to the features necessary to solve the hardest problems can be used for all problems. That is, if a certain subject distinguishes the greedy problems from the non-greedy ones, they should have a high probability of solving all of the problems. We describe the effects of each level in turn to provide a concrete explanation of how this may occur.

The first level of difficulty(easy problems) are strictly new instances of the training problems. They require little generalization other than following the procedure or instructions given in the training.

The second level of difficulty(medium problems) present new problems that require subjects to determine the proper greedy choice rule to solve the problem. These problems can be solved using similar techniques, but require more than simple rote following of the example problems - in other words, the subjects must use the concepts discussed in the training.

Finally, the most difficult problems(hard problems) are not solvable by the greedy choice property. Subjects here will fail to obtain the optimal solution with any local(greedy) solution. They are forced to recognize this fact and then develop an alternative means of solving the problem(e.g. a simplified version of dynamic programming). Solving problems in this category would represent a large conceptual leap, and it would be extremely surprising to find this behavior in naive subjects.

Table 3.4 summarizes the problems and their orthogonal dimensions. The medium problemset is particularly interesting as it is essentially a replication of the design of Chi et al(1984). Because the metaphor used in the training problems is switched, if subjects are simply following some shallow mapping of the training to the surface features of the problem, they will fail. Only if they actually recognize this change - via some deeper mapping - will the training be applicable to these problems.

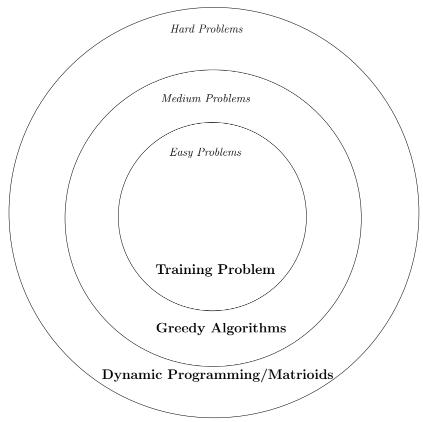

Figure 6: A Venn diagram of the problems(italics) and the strategies(bold) associated with each.

Figure 3.4 shows these different levels and the strategies associated with them. Note that each strategy(from the center out) is increasingly more general and subsumes the previous levels. For example, the easiest problem can be solved by any of these strategies while hardest problem can be only be solved though dynamic programming(which could be used for any of the problems). By structuring the problems in such a fashion, we are able to obtain a quantitative determination of how well subjects generalize what they have learned.

Taken as a whole, the experiment consists of two between subject factors - subject expertise and training as well as one within subject factor - the problem difficulty. The study measures the subject's performance as a function of these three factors. Qualitatively, performance as problem difficulty increases should be monotonically non-increasing, allowing verification that these problems follow the same ordinal trend for each condition (e.g the expert/procedural/easy condition should not have a higher success rate than the expert/conceptual/hard condition)

In general, to support the hypothesis,both groups should perform at an equal level when no training is provided, indicating lack of domain knowledge in each group. Furthermore, experts and novices should also be commensurate with each other during procedural training, since this mechanistic training does not elucidate the method to solve a larger family of problems.

The critical conditions are the expert and novice conceptual training. We expect the expert population to achieve greater performance gains on the conceptual task than novices of the same, indicating the presence of cross domain influences on this type of problem solving.

4 Experiment 1

Before the main experiment was attempted, it was necessary to tune the size of optimization problems to achieve our goals. Optimization problems, unlike typical tasks used in problem solving research, are general problems and can be created for a particular set size. That is, we can increase the computational complexity of the problem without changing the procedure needed to solve it.

This is advantageous for several reasons. First, with any sort of problem solving task, there is a chance that subjects will guess under uncertainty. The goal of the training is to force subjects to devise their own procedure to a novel problem. By manipulating the set size of an optimization problem, a method is used to find a reasonable level of difficulty such that a) subjects not creating a formal procedure(as taught in the training) have a high probability of failing b) subjects following the correct procedure will complete the problem without error.

This experiment, then, seeks to find a suitable cutoff for our optimization problems. The subjects are given the procedural training instance and asked to solve instances of same problem with varying set sizes.

4.1 Method

Participants Twenty Northeastern University undergraduate students in exchange for course credit.

Materials All subjects were given a packet of problems to complete. The packet consisted of the two training problems in the procedural training materials(available in Appendix B) and 16 additional problems - several instances 'easy' problem - in random order.

Design

Participants were asked to first read the training material, and then start solving the problems. Subjects were asked to solve 16 problems, four of each of the following set sizes - 8,12,16,20.

Participants were tested alone or in groups in a quiet room. They were given 30 minutes to complete the problems after opening the packet. Subjects were allowed to use scratch paper or write on the test itself, but no additional aids, including calculators, were allowed. After this time elapsed, subjects were asked to stop and any unanswered questions were consider incorrect. When subjects completed the task or time expired, they were debriefed and given credit.

4.2 Results

We wish to score the proportion of subjects that successfully and consistently attempted the problem - i.e. those that devised a novel procedure. Rather than measure the raw proportion of each set size, per subject, we assign the four problems in each set size a single, binary score. If subjects solved 3 or 4 of the associated problems, they were scored as 'correct', otherwise, the set size for that subject was marked as incorrect. Thus, we obtain a proportion of subjects that consistently solved the problem of that set size.

It is possible that subjects could fail at some set size i also succeed at set size j where $j > i$. After averaging, however, these sorts of discontinuities in performance were present in none of the subjects. Thus, the independent proportion correct on these set sizes can be considered a good measure of increase in difficulty.

The proportion of subjects responding correctly(by our definition above) was .80,.65,.60,.40 for set sizes 8,12,16 and 20, respectively. This categorical data was analyzed by χ^2 significance test. The difference in proportions is significant($\chi^2_{(}1, 20) = 10.883$, $p < .02$). Due to the small sample size, Fisher's Exact Test was conducted pairwise on each of the set sizes. 8 and 20($p < .001$), 12 and 20 ($p < 0.02$) were significant.

4.3 Discussion

The set size 16 appears to offer a suitable balance of preventing guessing behaviors while also not exceeding a threshold of complexity such that careless mistakes may obfuscate strong performance. It appears that in the lower sets sizes, a large proportion of subjects are able to solve the problem. Only at

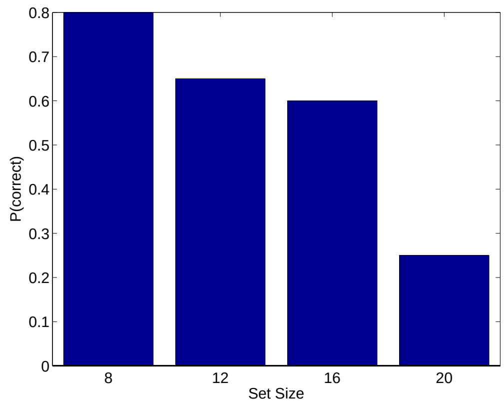

Figure 7: Experiment 1 Results

later tasks does the problem become sufficiently difficult to require the development of an explicit procedure. This is the size we use for all subsequent problems in experiment two.

5 Experiment 2

5.1 Method

Participants: 72 Northeastern University undergraduate students in exchange for course credit or ten dollars compensation. The novice group, consisting of 48 subjects was broken down into three training groups with 16 students each. Similarly, 24 experts were broken down into three training groups of 8 students each.

The expert and novice groups were selected on the basis of their knowledge in the domain of discrete math. A second criteria to minimize individual differences required all subjects to have an quantitative SAT score of 600 or greater (approximately one standard deviation from the average). The novices came from the general population and were screened to ensure they did not fit the expert profile. The experts were taken from a pool of computer science and math majors having taken 2 general courses in discrete math or programming.

Materials: A questionnaire was given to subjects before the experiment began. Subjects were ask to list any discrete math or computer science courses they had taken in college and their SAT scores in both the verbal and quantitative sections. All subjects met this criteria.

All subjects were given a packet of problems to complete. The packet consisted of the two training problems and the training material appropriate to their training condition(available in Appendix B). 6 testing problems(available in Appendix A) were provided to the students.

Design:

Participants were asked to complete the questionnaire. If subjects met the criteria, they were given a packet containing the materials listed above.

Subjects were asked to read the training material, and then start solving the problems. Subjects were then asked to solve the 6 problems in any order they chose. Subjects were allowed to use scratch paper or write on the test itself, but no additional aides, including calculators were allowed.

Participants were tested alone or in groups in a quiet room. They were given 45 minutes to complete the problems after opening the packet and reading the training material. After this time elapsed, subjects were asked to stop and any unanswered questions were consider incorrect. When subjects completed the task or time expired, they were debriefed and given credit.

5.2 Results

Problem difficulty:

We wished to ensure that problem difficulty was as expected. That is, the P(correct) on each of the three problem levels should be significantly different to ensure that the problems indeed did follow the predictions.

Considering all the problems individually(as opposed to the different levels of difficulty), there was a significant effect of problem $(\chi^2_(5, N = 432) = 147.481$, $p < .001)$. Merging the two problems at each level shows a similar trend. Data from both the expert and novice conditions, as well as all training conditions was tested. The results were highly significant$(\chi^2_(2, N = 216) = 70.807$, $p < .001)$ indicating that the problem did indeed predict the qualitative trend described in the design.

Novice performance: The novice analysis involves testing the testing the proportion of subjects successfully solving each problem. A 3x3(training by problem level) ANOVA was conducted. There was a main effect of both training($F_{(2,135)} = 15.86$, $p < 0.001$) and problem level($F_{(2,135)} = 99.31$, $p < 0.001$). There was also a significant interaction between problem type and training($F_{(4,135)} = 2.78$, $p < 0.05$). This interaction may be due to

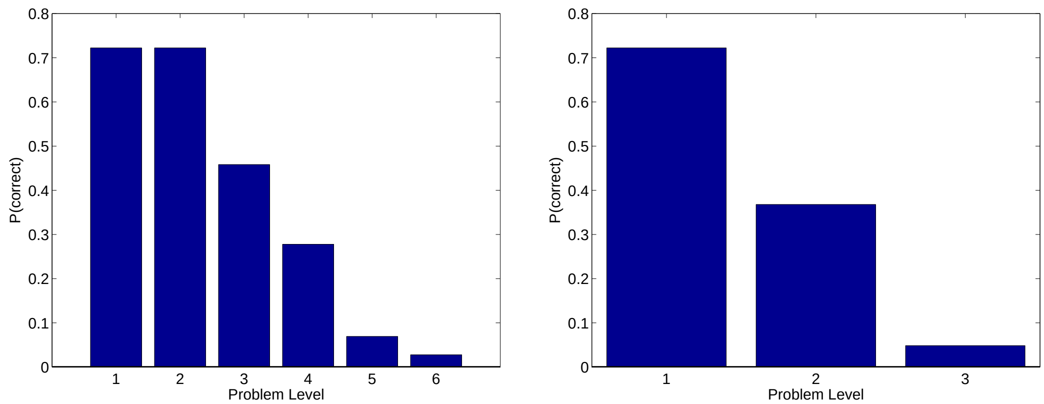

Figure 8: Effects of problem difficulty level. Left - Six problems individually; Right - Three problem levels.

novices hitting a floor in the no training condition. However, it may also signify a non-linearity with the hardest problems. In other words, transferring knowledge at the level of abstraction necessary for the hard problems is a rather large. The training conditions(procedural and conceptual) alone produce no interaction ($F_{(2,95)} = 0.66, p > .5$);

Tukey post-hoc tests over the training show that performance on the no training condition was significantly different than both the procedural and conceptual trainings. The latter two, however, were not different. Similar tests over problem level showed significant difference between all levels.

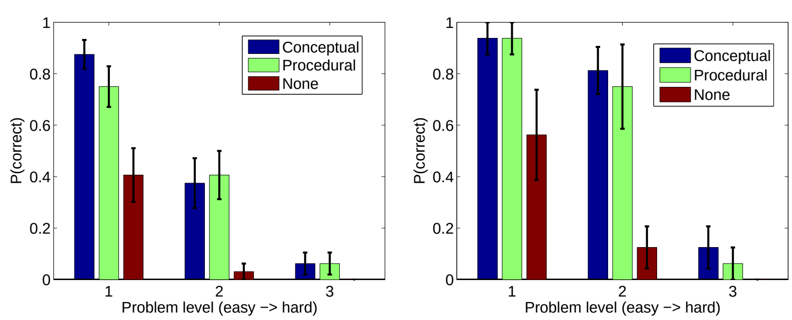

Figure 9: Separate analyses of expert and novice scores

Expert performance: The expert analysis is conducted similarly to the novice training. A 3x3(training by problem level) ANOVA was conducted. There was a main effect of both training($F_{(2,63)} = 13.6$, $p < 0.001$) and problem level($F_{(2,63)} = 99.31$, $p < 0.001$). There was also a significant interaction(similar in nature to that in the previous novice analysis) between problem type and training($F_{(4,63)} = 3.03$, $p < 0.05$).

Tukey post-hoc tests showed a similar trend to the expert behavior. Performance on the no training condition was significantly different than both the procedural and conceptual trainings. The latter two, however, were not different. Similar tests over problem level significant difference between all levels.

General performance A 3x3x2 ANOVA(training,problem and population ,respectively) was conducted over the dataset, showing significant main effects of expertise($F_{(1,198)} = 12.62$, $p < 0.001$), training ($F_{(2,198)} = 28.4$, $p < 0.001$) and problem difficulty ($F_{(21,198)} = 102.75$, $p < 0.001$). Additionally, there were significant interactions between expertise and problem difficulty ($F_{(2,198)} = 4.42$, $p < 0.02$) as well as training and problem difficulty($F_{(4,198)} = 5.12$, $p < 0.01$).

To present a clearer picture of the above analysis, we analyze each of these in turn by problem level. For simplicity, we now consider each of the problems in turn. Each of the following are a 3x2(expertise, training) ANOVA over the respective problems.

Easy problems There was no significant effects of expertise ($F_{(1,66)} = 2.67$, $p > .1$). However, the effect of training was significant ($F_{(2,66)}$ =10.4, $p < .001$). The interaction was not significant. Clearly, both training groups are around equal level. Training, however provided a significant advantage over lack thereof. The task also appeared to be fairly simple, in that both experts and novices had high($> .75$) success rates.

Medium problems There was a significant effect of expertise ($F_{(1,66)} = 12.95$, $p < .001$) and training ($F_{(2,66)}$ =17.56, $p < .001$). Training effects, as notice previously, were not significant between the procedural and conceptual training as found by Tukey post-hoc tests. The interaction(between training and expertise), again, was not significant($F_{(2,66)} = 1.5$, $p > .25$). However, the data(see medium vs. hard or easy problems in figure 5.2) clearly shows a large interaction between training and expertise. This is examined in further detail in the next analysis.

Hard problems No significant main effects existed, either in expertise ($F_{(1,66)} = 0$, $p = 1$) or training($F_{(2,66)} = 0.27$, $p > .25$). Subjects were uniformly poor at solving these problems that required a rather large jump in ability, as expected.

Training vs. no training: Because the analysis above suggests that there were no differences in effects between the two training conditions(conceptual and procedural), we wished to examine more closely the effects of training vs none. The same qualitative trends shown in the previous sections are replicated here, with one notable exception. In the medium problems, the interaction between training and expertise is significant($F_{(2,66)} = 3.06$, $p < .07$), while the interactions in the other problem types(easy and hard) were not(easy: $F_{(2,66)} = 0.03$, $p < .8562$, hard: $F_{(2,66)} = 0$, $p = 1$). This further exposes the differential abilities of experts on sufficiently difficult problems.

Performance and test scores: To determine any significant correlations between some general mathematical ability and performance on these tasks, the relationship between quantitative test scores and the number of tasks successfully solved was investigated. For the novice group, the correlation was not significant(r = 0.249, $p > 0.05$). Similarly, in the expert group, the correlation was not significant(r = 0.193, $p > 0.1$).

5.3 Discussion

Several trends are evident in the analysis above. First, it is shown that, in general, the materials used(problems and training) were effective in their intended goals of teaching subjects a concept and the presenting them with incrementally difficult problems. The training also gave large boosts to both expert and novice reasoners, There is strong evidence that the problems used in the experiment were verifiably situated on a gradient of difficulty. Both problem type and problem level showed significant differences across both subject groups.

Additionally, the training materials provided a significant boost in performance to both groups of subjects. Both procedural and conceptual trainings lead to large jumps in performance relative to the no training position. This is clear in both the easy and medium problems - e.g. whereas subjects in the no training condition rarely solved the medium problem, 80% and 40% of the experts and novices solved the problem, respectively (see figure 5.2). As a whole, however, these results indicate a well selected problem set that measured skill across a gradient of increasingly difficult problems.

Contrary to predictions, the two different training types had little effect across different subject groups. It is not immediately clear why the two

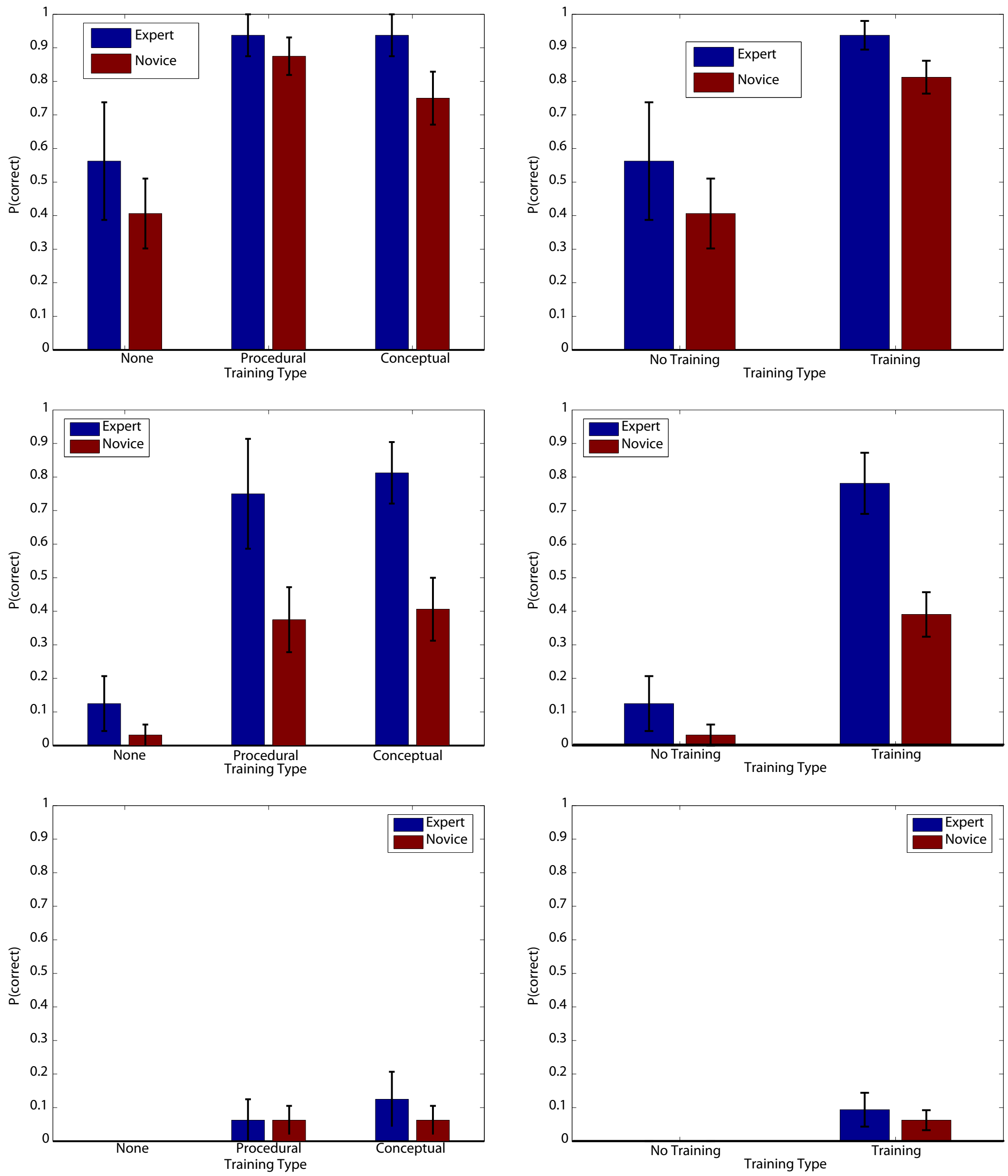

Figure 10: Expert and Novice scores for the Easy, Medium and Hard problems. The left column is across all training conditions, the right compares both training conditions(combined) with no training

training conditions were identical. Despite the fact that the types of training were vastly different - a rote procedure versus a conceptual description of the general procedure - both subjects subjects perform on the same level. One would expect that the procedural would not give these subjects the ability to generalize their knowledge on the hard problems. This clearly is not the case.

It is possible that the training for the procedural was implicit of the greedy choice property. That is, the attempts to standardize the length and content of the two training procedures may have lead to some conceptual overlap. Another interesting possibility is that the subjects were able to glean the conceptual features necessary to solve the medium problems with little effort. This appears to be somewhat unlikely, but reveals interesting vectors for future work.

Regardless, performance on the problems was indicative of the nature of the subject pool. It is clear that the hard problems show both subjects are naive relative to the domain of testing - few were able to solve the the problems even with training. More importantly, the analysis also shows that subjects were on roughly equal footing before training. That is, the simple presence of experience in D_2 does not give the expert subjects any additional advantage.

The striking result, then, is the interaction between expertise and training seen in the medium problem. Only with training do the expert subjects have additional leverage over the novice subjects. The medium problem is precisely where deeper structural must begin to be used for successful performance. It is here that we see the D_2 knowledge of the experts causing significant gains in performance.

6 General Discussion

As a whole, a simple picture emerges - on sufficiently difficult problems, trained naive experts are significantly better than trained novices. Without training, however, these subjects perform at equal levels.

This is a particularly interesting generalization of the result of Chi et. al.(1984). The original study showed that novices fail to distinguish problems by their deeper structural aspects and we have replicated this result in a highly different context. In Chi, experts had a rather large degree of experience in physics (a D_1 domain, in our terms) relative to the novice subjects. The same qualitative trend has been shown here with subjects with *equal*(i.e. none) D_1 knowledge. What distinguishes subjects in our study is the presence or absence of the related D_2 knowledge, which appears to aid experts in certain circumstances.

The differential performance of experts and novices was only seen in a critical condition - the medium problems with training. Akin to the Chi et. al. study, these problems were constructed such that over reliance on the surface features of the problem metaphors would lead to incorrect strategies. Both studies suggest that experience is critical in abstracting a task from distracting features. Surprisingly, the data suggests that experience outside of particular problem domain can influence this behavior.

The influences of D_2 domain knowledge cannot be simply attributed to additional experience. Experts in the medium level problems only outperformed novices in the training condition. This begs a simple question - why does training(in actuality, two seemingly distinct methods) allow the expert to achieve what the novice does not? Section 2 suggests related D_2 domains might share the same sort of structural similarity as the D_1 domain, allowing them to act as a bootstrapping mechanism during learning. Training may allow the naive expert to frame novel D_1 concepts in light of D_2 schema.

This perhaps permits naive experts to better contextualize these types of problems. Intuitively, if a task can be decomposed into something more familiar via instruction, it can be better understood. Of course, this may also hinder progress if the D_2 domain is not mappable to the D_1. This may be precisely why novice subjects falter - they lack the deep structure in D_1, and seemingly similar aspects(e.g. surface features) become a highly salient alternative.

This result specifically highlights the relational underpinnings of the acquired concepts. For this D_2 information to aid problem solving by analogical transfer, it must have some degree of interconnectedness with the newly acquired information. While this does not facilitate large differences in skill(it did not help the naive experts solve the most difficult problems), it does reveal subtle differences in performance.

There are alternate explanations for this behavior, however. We attempted to select the training and problems such that they could be understood with D_1 domain experience, but it may be the case that subjects in the novice condition had difficulties understanding the training. Clearly, though, they did glean something from it. However, it is difficult to disambiguate whether the experts indeed have effects of cross-domain influence or whether training in general was more suited to their understanding(via terminology, etc rather than cross-domain knowledge).

An apparent lesson in this sort of testing is that concocting stimuli requires significant effort. Despite heavy pretesting, there still exist problems(i.e. the 'hard' problems are too hard). Using a more rigorously tested training and problem set might alleviate some of disambiguation problems seen above. Furthermore, a greater sample size for both groups may be necessary due to binary(correct or incorrect) measure of performance as well as the volatile nature of human performance. Ideally, one could construct instances of optimization problems where the specific value obtained gives evidence as to what strategy was followed.

Furthermore, using naive subjects necessitates using simplified versions of complex domains. For a more transparent study, there must be a way to test these sorts of effects with a with two mutually exclusive domains. Optimization and discrete math may be too close to be meaningfully distinct, even in naive learners. It would be especially interesting to replace discrete math with some other, more 'distant' domain to see if similar results are achieved.

Despite these flaws, the primary result of this study is compelling - different levels of knowledge associated with a particular domain affect performance. Specifically, we have shown that, when given salient cues via training, subjects with only very basic level understanding in D_2 perform better than those without such knowledge. Whether the knowledge can truly be considered cross-domain remains an open question, but having some level of expertise in associated knowledge certainly aids transfer.

If students are as poor at transferring their knowledge as the literature discussed in the introduction suggests, these results may be indicative of one of its many causes. While it reinforces the need to explicate the interrelatedness of learned materials, it also suggests that a wide breadth of exercises and topics may be helpful in building the intuition that leads to structured conceptual knowledge. Whether cross-domain or otherwise, a breadth of knowledge may be essential for bootstrapping further depth in understanding.

Ostensibly, the interaction between knowledge and its structure - across different levels of generality, domain and otherwise - is one of the primary reasons human thought is so unique. Understanding how and when this interaction occurs is crucial to uncovering why it is so difficult to learn abstract problem solving and a myriad of other tasks. This study has provided an initial exploration of these dynamics. While further research is necessary to present a more thorough picture of cross-domain effects, we have shown that even small amounts of expertise in some D_2 aid in generalizations to novel D_1 contexts.

References

A Newell and HA Simon, (1972) Human problem solving, Prentice-Hall, Englewood Cliffs

Burgess N., and O'Keefe J (1996) Neuronal Computations Underlying the Firing of Place Cells and Their Role in Navigation, Hippocampus, 6

Chi M. T. H. (1978) Knowledge structures and memory development, in Childrens' thinking: What develops?, Lawrence Erlbaum Associates

Chi, M. Bassok, M.W. Lewis, P. Reimann, and R. Glaser (1989) Self-explanations: How students study and use examples in, Cognitive Science, 13

Cormen T., Leiserson, C., Rivest, R., Stein (2004) Introduction to Algorithms, MIT Press

D. Vickers, M. Heitman, M. D. Lee and P. Hughes (2001) Optimization And Intelligence: Individual Differences In Performance On Three Visually Presented Problems, Personality and Individual Differences, 23

Dennett, D. C. (1984). Cognitive Wheels: The Frame Problem of AI. In C. Hookway (Ed.), Minds, machines, and evolution (pp. 129-151). Cambridge: Cambridge University Press.

Dunbar, K. (1995) . How scientists really reason: Scientic reasoning in real-world laboratories. In R. J. Sternberg J. E. Davidson Eds. , The nature of insight p. 365-395 . Cambridge, MA: MIT Press.

Gelman SA. Hirschfield LA(1994) Mapping the mind: Domain specificity in cognition and culture, Cambridge University Press

Gentner D.(2002) Analogy in scientific discovery: The case of Johannes Kepler Model-based reasoning: Science, technology, values, New York: Kluwer Academic/ Plenum Publisher.

Edmonds, J. (1971) Matroids and the greedy algorithm, Mathematical Programming, 1

Garey, M. D. Johnson (1979) Computers and Intractability; A Guide to the Theory, W. H. Freeman

Gentner, D. (1983) Structure mapping:A theoretical framework for analogy., Cognitive Science, 7

Gentner, D., Markman, A. B. (1997) Structure mapping in analogy and similarity, American Psychologist, 52

Gentner, D., Loewenstein, J., Thompson (2004) Analogical encoding: Facilitating knowledge transfer, Proceedings of the Twenty-sixth Annual Meeting of the Cognitive Science Society,

Gick, M. L., Holyoak, K. J. (1983) Schema induction and analogical transfer, Cognitive Psychology, 15

Greene J. D., R. B. Sommerville, L. E. Nystrom, J. M. Darley, and J. D. Cohen (2001) An fMRI Investigation of Emotional Engagement in Moral Judgment, Science,

Holyoak, K. J. (1984) Analogical thinking and human intelligence, in Sternberg R.(ed), Advances in the psychology of human intelligence, Eribaum

Holyoak. K. J. (1992) The pragmatics of analogical transfer, The psychology of learning and motivation, 19

Kuhn, Thomas S. 1970. The Structure of Scientific Revolutions. second edn. Chicago, IL: The University of Chicago Press.

Kotovsky, K., Hayes, J.R., Simon, H.A. (1985) Why Are Some Problems Hard? Evidence From the Tower of Hanoi, Cognitive Psychology, 17

Kurtz, K. Chun-Hui M., Gentner, D. (2001) Learning by Analogical Bootstrapping, Journal of the Learning Sciences, 10

M McCloskey (1982) Naive theories of motion, in D. Gentner(Eds.) Mental models, Lawrence Erlbaum Associates, Inc

M.T.H. Chi , P.J. Feltovich R. Glaser (1981) Categorization and representation of physics problems by experts and novices, Cognitive Science, 7

N Kanwisher (2000) Domain specificity in face perception, Nature Neuroscience

Ormerod, T. C., Chronicle, E. P. (1999) Global perceptual processes in problem solving: The case of the traveling salesperson, Perception Psychophysics,26

Pirolli P., Recker M (1994) Learning strategies transfer in the domain of programming, Cognition Instruction, 15

Reeves, LM.,Weisburg, RW. (1994) The Role of Content and Abstract Information in Analogical Transfer, Psychological Bulletin, 115

Ross, B. H. (1996) Category learning as problem solving, The Psychology of Learning and Motivation, 19

Sandberg, J. Barnard, Y. (1997) Deep Learning is Difficult, Instructional Science, 9

Sims, Valerie K. and Mayer, Richard E. (2002) Domain Specificity of Spacial Expertise: The Case of Video Game Players, Applied Cognitive Psychology, 16

Appendix A - Training Materials

Appendix B - Problems