Psychological Bulletin
2002, Vol. 128, No. 4, 612–637

0033-2909/02/$5.00 DOI: 10.1037//0033-2909.128.4.612

When and Where Do We Apply What We Learn? A Taxonomy for Far Transfer

Susan M. Barnett and Stephen J. Ceci
Cornell University

Despite a century's worth of research, arguments surrounding the question of whether far transfer occurs have made little progress toward resolution. The authors argue the reason for this confusion is a failure to specify various dimensions along which transfer can occur, resulting in comparisons of "apples and oranges." They provide a framework that describes 9 relevant dimensions and show that the literature can productively be classified along these dimensions, with each study situated at the intersection of various dimensions. Estimation of a single effect size for far transfer is misguided in view of this complexity. The past 100 years of research shows that evidence for transfer under some conditions is substantial, but critical conditions for many key questions are untested.

Disagreement at the beginning of the 20th century

> Every experience has in it the possibilities of generalization. (Judd, 1908, p. 38)
>
> There is no inner necessity for improvement of one function to improve others closely similar to it, due to a subtle transfer of practice effect. (Thorndike & Woodworth, 1901b, p. 386)

. . . and disagreement at the end

> Numerous studies have shown that critical thinking . . . can be learned in ways that promote transfer to novel contexts. (Halpern, 1998, p. 449)
>
> Reviewers are in almost total agreement that little transfer occurs. (Detterman, 1993, p. 8)

Can a ninth-grade honors math student apply knowledge of geometry to estimate the square footage of the family's new home? Can a military recruit in basic training apply a textbook lesson on the principles of radar to troubleshooting a radar system when it malfunctions in the field? Can a baker apply what was learned about fractions in school to an analysis of how to divide the ingredients for a recipe?

These questions all involve the concept of transfer of learning and knowledge. This concept is not a new one; the importance of this phenomenon for both everyday functioning and theory has been documented for more than a century. But, what is really known about transfer of knowledge? What do scholars agree on, and on what do they disagree? In fact, there is little agreement in the scholarly community about the nature of transfer, the extent to which it occurs, and the nature of its underlying mechanisms. In this article, we propose a first step toward making sense of the vast literature on transfer by focusing on one aspect of the topic, namely, the extent to which transfer occurs. To do this, we need to interpret and unify our understanding of what is meant by the occurrence of transfer. We focus on what is transferred and when and where the transfer occurs and attempt to better define the term *far transfer*. In doing so, we offer a taxonomy for transfer and show that previously published studies may be better understood if viewed within this framework. Specifically, we propose that this literature may be organized from the perspective of considering transfer along a number of contextual and content dimensions. Applying the taxonomy, we offer evidence of successful transfer along some of these dimensions and highlight notable gaps in the research literature.

In this article we do not address the additional, but equally important, questions concerning how to ensure that knowledge is learned well in the first place—that is, questions concerning the ways in which teaching can be optimally tailored to promote transfer—and the mechanisms underlying the transfer process. Thus, the taxonomy does not cover details of the teaching regimen (issues of variability of practice, feedback, and training instructions) and characteristics of the individual learner (age, intelligence, and related knowledge) that could clearly play a role. In our view, theories concerning the optimal way to teach for transfer and the microgenetic analysis of the processes underlying transfer cannot be effectively evaluated without a framework for comparing their success. Thus, we focus exclusively on the prerequisite task of developing a framework for evaluating when far transfer occurs.

Recently, sophisticated processing models have been put forward to account for certain observations, particularly surrounding analogical reasoning and transfer (e.g., Forbus, Gentner, & Law, 1995). Such processing models account for the differential efficacy of certain types of cues and surface-versus-deep-structure differences. However, these processing models were not developed to explain the sorts of data we shall be addressing in this article. Specifically, they were not formulated to inform us about whether far transfer is as rare as Detterman (1993) claimed or as pervasive

Susan M. Barnett and Stephen J. Ceci, Department of Human Development, Cornell University.

Early thinking on this topic by Susan M. Barnett was supported by a Graduate Research Fellowship from the National Science Foundation. We thank Ross Barnett, Doug Detterman, Diane Halpern, Keith Holyoak, David Klahr, Barbara Koslowski, Amy Masnick, Ulric Neisser, Wendy Williams, and Jack Young for helpful comments on previous versions of this article.

Correspondence concerning this article should be addressed to Susan M. Barnett, Department of Human Development, Cornell University, Ithaca, New York 14853. E-mail: smb11@cornell.edu

as others alleged (e.g., Brown, 1989; Halpern, 1998). Moreover, such models, notwithstanding their formal elegance, are silent regarding many of the issues that motivate this discussion, such as whether and when reasoning skills might transfer from school to home or work. Therefore, we do not discuss such models here.

Our emphasis is on studies that provide empirical evidence regarding human transfer of thinking and reasoning from one context to another. This means we do not review studies that deal with transfer by animals, nor do we cover transfer of perceptual learning (e.g., letter cancellation tasks; see Martin, 1915, as cited in Woodworth & Schlosberg, 1954, for the earliest example), attitudes and moral development (e.g., Sunday school ethics transferring to everyday moral behavior; see Humphreys, 1951), and motor skills (e.g., mirror drawing studies; see Starch, 1910, for the earliest example), except insofar as such studies may illustrate principles that are germane to the issue of far transfer of cognitive skills, which drives the present framework.

This article is divided into three main sections, each further divided into subsections. In the first section we discuss the existing state of knowledge about the success of transfer and explain why making sense of this body of research has been so difficult. In the second section we describe a proposed taxonomy of transfer to add structure to this messy debate and apply this taxonomic approach to seminal studies. We discuss in the third section the broader implications of the proposed taxonomy. We conclude with a discussion of what is known and what remains to be determined about transfer in light of this more systematic treatment of the data, as well as a discussion of the theoretical implications of this way of thinking.

Existing State of Knowledge

In this section, we suggest two reasons why transfer of knowledge is such an important topic, discuss the historical literature that frames the debate, and review and contrast findings of key studies.

Importance of the Topic

Can we transfer what we learn? How similar does the learning context have to be to the transfer context? Is this independent of the content we wish to apply? These are a few of the questions that have animated psychologists' long-standing interest in the topic of transfer.

The nature of transfer and the frequency and context of its occurrence have been the subject of considerable empirical and theoretical research for the past 100 years, resulting in thousands of articles, chapters, and books dealing with various aspects. There are two reasons why this topic has attracted so much attention throughout the past century, one theoretical, the other practical.

Theoretically, transfer provides an important test-bed for models of learning and performance, a point cogently made by Singley and Anderson (1989): The definition and assessment of performance models often turn on whether learned behaviors are permanent and, if so, whether they are applicable in novel contexts. The question of whether learned behaviors are applicable in novel contexts also has implications for the domain specificity of expertise (see, e.g., Chi, Glaser, & Farr, 1988; Ericsson & Smith, 1991) and whether it is possible to develop so-called adaptive or transferable expertise (Barnett & Koslowski, in press; Hatano & Inagaki, 1984). Furthermore, as we revisit in our concluding comments, transfer carries with it ramifications for theories of general intelligence, specifically, the basis of cross-task correlations and *g*, or so-called general intelligence.

Practically, there is also a lot at stake in the outcome of transfer research, in terms of both money and time invested in education. Many of the dates, names, concepts, and theorems students are confronted with in school are unlikely to be of great relevance to their future lives.[1] Indeed, the world is changing so fast that even some of the most seemingly directly applicable components of the school curriculum—for example, how to use a computer—may be outdated before the student leaves the educational establishment. In his classic educational psychology text, Klausmeier (1961) asserted, "A main reason for formal education is to facilitate learning in situations outside school" (p. 352). Consequently, much of the financial and human investment in education has been justified on the grounds that formal schooling helps inculcate general skills that transfer beyond the world of academia and thus help students become more productive members of society. It is assumed that through repeated exposure to logically equivalent problems, children distill the underlying reasoning schemes and develop meta-cognitive insights into the inner workings of their cognitive systems and that these insights and schemes transfer to solving problems they confront outside school (Van de Vivjer & Hutschemaekers, 1990). As Resnick (1987) has pointed out, "The goals of increasing thinking and reasoning ability are old ones for educators, . . . such abilities have been the goal of some schools at least since the time of Plato" (p. 7).

It is these underlying basic processes, insights, and modes of cognizing that are reputed to have enduring applicability beyond the specific lessons in which they are taught (see Sewell, Hauser, & Featherman, 1976; Walberg, 1982; Wiley, 1976). Hence, it is a presupposition of educators that a student taught to permute a set of items in school will transfer this skill to sets of items confronted outside of school; a student taught arithmetic will transfer this knowledge to calculate a bowling average; a student taught to organize items hierarchically will transfer this skill to answering questions on IQ tests (Ceci, 1991). The transferability of learning is of prime importance in evaluating these educational claims, a point not lost on scholars in this area: "Given that universities attempt to hone critical abilities and see this in fact as one of their major tasks, our results [concerning the transfer of coursework] paint an optimistic picture" (Lehman & Nisbett, 1990, p. 959).

In addition to the large financial investment in formal education, significant investment is made in workplace and military training that is (presumably) more directly relevant to the job at hand. Yet, even in these cases, the intention is usually not just to teach trainees how to succeed on training course examples (e.g., principles of a radar system taught to recruits during basic training) but

[1] This is not meant to imply that nothing of enduring value is learned in school. It is clear that skills such as learning to read, calculate, and understand the fundamental rules of nature and society are acquired as a result of formal teaching. However, if this were the only justification for formal schooling, such content could be taught in a fraction of the time (and money) devoted to education. For instance, summer academic camps routinely elevate children's mathematic achievement by an entire school-year level in only 3 weeks of camp (Benbow & Lubinski, 1996).

to build skills that then transfer to myriad variations that may be encountered on the job (e.g., trouble-shooting a problem in a radar system during combat). Again, the transferability of skills is key: If the skills developed by such efforts do not transfer beyond the training context, much of the investment may be considered wasted, as noted in a National Research Council report on enhancing human performance (Druckman & Bjork, 1994).

Nonetheless, this is precisely the criticism that a number of scholars have made about the failure of transfer studies to document that training in one context or on one type of problem generalizes to related problems in different contexts. Consider these examples:

> Transfer has been one of the most actively studied phenomena in psychology. . . . Reviewers are in almost total agreement that little transfer occurs. (Detterman, 1993, p. 5, p. 8)

> The question for which we do have some empirical answers has to do with how generalizable cognitive training is from one subject area to another. As of now, the answer is not very much. (Schooler, 1989, p. 11)

If critics are correct in asserting that transfer very rarely happens, then the justification for educational and training expenditures may need to be reevaluated, as Detterman (1993) noted when he said, "Cognitive psychologists, and other people who should know better, continue to advocate a philosophy of education that is totally lacking in empirical support" (p. 16).

One might think, therefore, that what is needed is a meta-analysis of the corpus of transfer research over the past century to determine definitively whether transfer happens. However, such an analysis would be misleading because of the lack of common structure to the studies. Our thesis is that the disagreement among proponents and critics of transfer is the result of a lack of structure in the transfer debate and a failure to specify the various dimensions that may be relevant to determining whether and when transfer occurs. Although there are definitions of transfer such as "the carrying over of an act or way of acting from one performance to another" (Woodworth & Schlosberg, 1954, p. 734) and "the ability to extend what has been learned in one context to new contexts" (Bransford, Brown, & Cocking, 1999, p. 39), there is no clear, agreed-on definition of what constitutes "carrying over" or "a new context."[2] As a consequence of this lack of a clearly operational definition, protagonists often seem to be talking at cross-purposes—comparing apples and oranges. This is because the notion of transfer encompasses many things that are unstated in the definition, such as the relevant dimensions.

To resolve this debate over whether transfer occurs, it is necessary to specify a framework of relevant dimensions, identify where along these dimensions the pivotal questions regarding the success or failure of transfer lie, and map the findings of the conflicting studies against this taxonomic framework. Only then would it be possible to differentiate between genuinely contradictory studies and those that merely reflect perceived disagreements due to ill-defined terminology, failure to specify dimensional characteristics, or both. If we succeed at providing a taxonomic framework and definition, then rigorous experimental tests can be designed to resolve substantive disputes and to fill in remaining important gaps in our knowledge about transfer, such as when it occurs and what are the constraints with which it operates. That there is clearly a need for a taxonomic framework has been argued by others. In a National Research Council report, part of which was devoted to the issue of transfer, Reder and Klatzky (1994) commented that "efforts to understand transfer would clearly be helped by greater efforts to characterize the nature of relevant variables. A valuable addition would be a task taxonomy" (p. 52). In this article we propose just such a taxonomy. It is intended to provide structure to the century-long controversy over whether and when transfer occurs.

To adumbrate the framework, we argue that, at a minimum, the following dimensions are needed: (a) the nature of the skill to be transferred, the performance change measured for this skill, and the memory demands of the transfer task used to measure it and (b) the distance between the training and transfer contexts along multiple dimensions (knowledge domain, physical context, temporal context, functional context, social context, and modality).

First, however, we provide an abbreviated historical background to the current debate over whether transfer occurs. Following this, we provide a tour of the modern transfer literature, focusing on the seminal and most important studies (operationalized as those studies relied on by current writers on this topic), rather than an exhaustive literature review that would necessarily require book-length treatment. The inclusion of omitted studies would not significantly change the picture we present below.

The Historical Backdrop to the Current Transfer Debate

The history of transfer research goes back almost 100 years to the debate between Thorndike and Judd about the implications of their various experimental results. In 1908, Charles Judd reported on the sketchy details of a rough experiment carried out many years earlier by himself and Scholckow in which fifth- and sixth-grade boys were instructed to throw small darts at a target submerged under 12 in. of water. Some participants were provided with a theoretical explanation of the principle of optical refraction that made the underwater target appear skewed, and others were not. On this initial task, there was no difference in rate of hitting the target as a consequence of the explanation of refraction. However, when the target's underwater depth was changed from 12 in. to only 4 in., there was a large superiority for the boys trained in refraction, even though the transfer task (throwing darts) was very different from the training context (theoretical explanation of refraction).[3] Adherents of Judd's ideas argued that this finding supported the view that teachers should focus on teaching broad principles and generalizations rather than specific facts, skills, and beliefs.

Although he was not a proponent of the so-called doctrine of formal discipline, Judd's (1908) study was cited as evidence for the proposition that certain types of learning have pervasive and enduring effects on the mind and foster generalized thinking that

[2] Singley and Anderson (1989) presented a mathematically formal definition that relates the degree of improvement in a transfer task to the improvement that might occur in equivalent practice on the initial task. However, this does not address issues concerning the content or context of the transfer test.

[3] Later researchers were to replicate this study by showing the superiority during transfer of participants trained in the theoretical explanation of refraction, but they also demonstrated that the theoretical explanation facilitated the original learning (e.g., Hendrickson & Schroeder, 1941).

goes beyond the specific training provided. According to early adherents of the doctrine of formal discipline, the particular content did not matter as long as the mind was exercised in mastering it. Thus, students were made to memorize long lists of digits in the belief that this strengthened memory in general; learning an onerous assignment was thought to strengthen will power and attention.

Toward this end, American and British proponents touted the presumed benefits of instruction in the classics, geometry, logic, Latin, and chess—to name a few of the types of activities that were assumed to enhance general thinking skills and result in transfer to contexts outside of formal education. Indeed, at about the same time that American and British scholars were touting the superiority of the doctrine of formal discipline, Binet appeared to have adopted a position similar to it, arguing that learning should be predicated on training in basic processes that he believed strengthened the mind's faculties:

> What they should learn first is not the subjects ordinarily taught, however important they may be; they should be given lessons of will, of attention, of discipline; before exercises in grammar, they need to be exercised in mental orthopedics; in a word they must learn how to learn. (Binet, 1908, cited in Gould, 1981, p. 154)

However, this viewpoint was difficult to reconcile with Edward Thorndike's conclusions regarding his failure to find much evidence for such transfer throughout his long career. Beginning around the turn of the 20th century, Thorndike and his colleagues reported numerous experiments finding poor or uneven transfer across disparate tasks that entailed similar operations. For example, following instruction in the principles of estimating the area of geometric shapes (e.g., 100-cm^2 rectangles), participants did not transfer their learning to solve other problems concerning estimating geometric area, such as estimating the area of other rectangles and triangles (see Thorndike & Woodworth, 1901a). These results led Thorndike (1906) to posit that for transfer to occur, it was necessary for the elements present in the original learning context to be present in the transfer context. Given the rarity of this happening, he was led to make his provocative claim that transfer rarely occurs, saying,

> The mind is so specialized into a multitude of independent capacities that we alter human nature only in small spots, and any special school training has a much narrower influence upon the mind . . . than has commonly been supposed. (Thorndike, 1906, pp. 246–247)

This claim was to sway an entire generation of scholars. Thus began the century-long debate over the nature, contexts, and prevalence of transfer—a debate that continues today (e.g., Brown, Kane, & Long, 1989; Detterman, 1993; Halpern, 1998; Perkins & Grotzer, 1997; Singley & Anderson, 1989). Outside of the transfer debate, the same concept of identical elements versus generalized schemes found currency. For instance, in cognitive psychology, the concept of *encoding specificity* emerged as a rationale to explain the superiority of matching the memorial context at the time of retrieval with that which prevailed at the time of encoding (Tulving & Thomson, 1973).

Given the theoretical and practical importance of the transfer debate, researchers have conducted countless studies designed to find out whether transfer happens. As we demonstrate, however, these studies have not yielded clear answers. In fact, the debate may be as unresolved today as it was 100 years ago, and the issue of whether generalizable reasoning skills transfer to reasoning contexts outside of formal schooling remains an open question in the opinions of leading researchers. Consider the opinion of Resnick (1987): "The issue of transferability of thinking and learning skills is . . . still open" (p. 19). Consider also this question posed by Singley and Anderson (1989): "What then is the current status of the notion of general transfer? Is it dead, or very much alive?" (p. 25).

Findings of Key Modern Studies

Rather than provide a comprehensive review of this large and diverse body of research, we focus on those studies that are seminal or unique in the points they make. It is important to note that we focus on the studies that concern so-called *far transfer* (transfer to a dissimilar context) as opposed to *near transfer* (transfer to a more similar context), as this is the type of transfer that educators and policymakers are most concerned about. We include both studies in which the authors make claims that far transfer has occurred and studies that other researchers and reviewers have used to support conclusions about the transferability of education. Our decision to focus on far transfer also derives from the fact that far transfer is most relevant to questions about how to best train for transfer, because such questions arise out of a desire to ensure that what is taught in schools is generally applicable over time and contexts, not just immediately in similar contexts.

Research relevant to the question of cognitive transfer is an interdisciplinary endeavor that has been conducted under a number of guises, including (a) studies of analogical transfer, (b) investigations of the doctrine of formal discipline, (c) attempts to teach intelligence and "higher order skills," and (d) evaluations of the effects of schooling. We turn to each of these areas next.

Analogical Transfer

Analogical transfer studies involve training on one task followed by testing on a novel task that is an analogue of the first, to see if the training transfers to the analogical task. In the modern era, studies of analogical transfer have yielded myriad, ostensibly conflicting results. Many of these studies have succeeded in showing transfer, but a number of somewhat similar studies have failed. We describe here key studies of both kinds.

For example, Gick and Holyoak (1980) showed somewhat successful transfer in a prototypical analogical transfer experiment using the classic Duncker (1945) tumor radiation problem and a military analogy.[4] In this series of studies, college students were

[4] The military problem was as follows:

> A general wishes to capture a fortress located in the center of a country. There are many roads radiating outward from the fortress. All have been mined so that while small groups of men can pass over the roads safely, a large force will detonate the mines. A full-scale direct attack is therefore impossible. The general's solution is to divide his army into small groups, send each group to the head of a different road, and have the groups converge simultaneously on the fortress. (Gick & Holyoak, 1980, p. 309)

presented with one or more stories—in some conditions in the guise of a memory study—followed, either immediately or after a few minutes' break, by the radiation problem. Those who were trained with the closest analogy as a source from which to transfer were more likely to come up with the target solution for the transfer problem. Although most trained participants apparently did not transfer their prior learning without a hint, virtually all did eventually solve the transfer problem when they were explicitly told to think about the training problem.

On the other hand, Reed, Ernst, and Banerji's (1974) prototypical analogical transfer study failed to demonstrate transfer on most performance measures. They investigated transfer between the missionaries and cannibals problem (how to get safely across a river in a limited-capacity boat without having the cannibals in the group ever outnumber the missionaries) and an analogous problem substituting wives and jealous husbands.

Brown and her colleagues (Brown, 1989; Brown & Campione, 1990; Brown & Kane, 1988; Brown, Kane, & Echols, 1986; Brown, et al., 1989) conducted a series of experiments designed to study analogical transfer in young children and showed that transfer occurred but only under certain circumstances. For example, Brown's (1989; Brown & Kane, 1988) experiments showed that children could transfer principles such as "hide using mimicry as a defense mechanism" from one animal to another. Children transferred most successfully when they understood events at a causal level rather than merely learned to replicate particular behaviors. That is, they transferred when they developed a deep, rather than surface, understanding.

A number of other researchers have reached similar conclusions and have investigated the ways in which training can be conducted to most readily facilitate such a deep understanding. These various approaches focus on getting participants to work with the training materials at a deep, structural level. For example, Catrambone and Holyoak (1989) used comparison questions to promote induction of a general schema from multiple examples, which then improved transfer. Also, Cummins Dellarosa (1992) found that interproblem processing (focus on comparison questions) promoted more transfer than intraproblem processing (focus on specific wording or details). Similarly, Needham and Begg (1991) found that problem-oriented training (e.g., trying to explain) resulted in more transfer to a problem-solving task than did memory-oriented training, and Halpern, Hansen, and Riefer (1990) enhanced participants' ability to draw inferences from a studied passage by including far analogies in their training materials, presumably encouraging a focus on deep, structural processing. Halpern et al.'s participants did not derive the same benefit from a near analogy, which the authors suggested may have been because participants did not have to exert as much effort to make sense of the near analogy as they did for the far analogy. Also, Reed and Saavedra (1986) showed that a task involving more concrete and effortful processing, termed the *discovery method* (running a computer simulation with feedback), improved performance more than a passive task (observing a computer-generated graph) and a more abstract method (performing calculations).

In a different analogical reasoning environment, Schliemann and Nunes (1990) studied poorly educated Brazilian fishermen and also demonstrated successful transfer—from one fishing problem to another. Even the subset of fishermen who had received formal schooling, who could often solve the problems within the domain of fishing, did not use the procedure they had learned in school. Rather, they developed intuitive heuristics to solve such problems.[5] Schliemann and Nunes also studied a sample of school students who did not transfer very successfully from math class (where they were taught the procedure) to the seafood yield problem that the fishermen solved, although the same students did transfer their own everyday approach to solving proportionality problems, which the authors presumed was developed from repeated shopping transactions.

One possible explanation for the success of Judd's (1908) and Gick and Holyoak's (1980) experiments in finding transfer, the failure of Thorndike and Woodworth's (1901a, 1901b, 1901c) and Reed et al.'s (1974) experiments, and the mixed results of Schliemann and Nunes (1990) may be Brown's (1989) distinction between the presence and absence of an understanding at a deep level, leading to an ability to transfer the principle. Perhaps Reed et al.'s (1974) participants, although able to successfully solve the missionaries and cannibals problem, did not understand why their solution was successful at a strategic (deep) level and therefore were unable to transfer, whereas Judd's and Gick and Holyoak's participants learned a (deep) principle—refraction or convergence leading to concentration, respectively. Thorndike and Woodworth's (1901a, 1901b, 1901c) participants were not engaged in problem solving but instead were making perceptual judgments for which there was no deep principle to transfer. Schliemann and Nunes's fishermen may have understood their own heuristics but may have merely rote-learned the procedure taught in school. This is essentially the position of those who argue that transfer depends on the flexible deployment of abstract general principles.

In addition, Brown (1989) concluded that knowledge about the area to which the principle is to be transferred is also key to successful transfer. Children cannot apply learned causal schemas to a subject about which they know nothing. When these criteria were met, Brown found that children successfully transferred a general principle to a novel context. Further evidence to support this comes from a study by Schliemann and Magalhães (1990), in which uneducated Brazilian cooks were given mathematical proportionality problems in three different domains, or bodies of knowledge (prices, recipes, and medicines), using similar relations and quantities. They performed perfectly on the price problems, a domain in which they were familiar with using precise proportionality calculations. However, they performed less well on the recipe problems, an area in which they were familiar with performing only rough proportionality calculations. They performed even more poorly on the medicine problems, a totally unfamiliar topic,

The radiation problem described an analogous situation in which a type of ray could be used to kill a cancerous tumor; however, in the dosages needed it would also kill surrounding tissue. The solution is to spread the rays around the patient's body as they converge simultaneously on the tumor from different locations, thus not killing the surrounding tissue.

[5] Elsewhere, these researchers reported similar results. For example, unschooled street vendors used intuitive heuristics to calculate costs of coconuts and were 98% accurate. However, when they were tested on the identical mathematical operations, their accuracy dropped to 37% (T. N. Carraher, Carraher, & Schliemann, 1985).

despite the similarity of the mathematical calculations required for performance in this domain and the other two domains.

In other work regarding successful transfer of principles, Kosonen and Winne (1995) taught American college students the "law of large numbers" (the effect of sample size on statistical reliability). They found that the learning transferred to new problems on other topics (hiring people with particular qualifications, conducting a survey of attitudes, playing a recreational board game, etc.) and in other formats (probabilistic, objective, and subjective). Also, Chen and Klahr (1999) reported a demonstration of successful transfer of the "control of variables strategy," a scientific reasoning principle, by elementary school children. In a systematic study investigating transfer performance at different ages and on a variety of tests, children were asked to evaluate the rigor of simple scientific experiments. We return to this study later, as it provides a basis for discussing several of the dimensions we propose to classify transfer studies.

In an alternative approach, Novick (1990) investigated undergraduates' transfer of a method of problem representation. She found that exposure to a matrix representation in a training problem increased the probability of using a matrix to help solve a transfer test problem.

Finally, there are cases in which participants actually perform worse on the transfer task than they would have if they had not been exposed to the initial training task in the first place. Such effects, called *negative transfer*, have been reported throughout the past century for both insight and analogy problems (see Woodworth & Schlosberg, 1954, for a review of early studies of negative transfer; see also Singley & Anderson, 1989, for a review of modern studies, including their own). A well-known example of negative transfer in an insight problem comes from a study by Luchins (1942). He gave his participants a series of water jugs that contained different volumes. Participants were asked to figure out how to measure a specific amount by using the three jugs. Luchins discovered that if participants were previously given training problems that required them to use more elaborate arithmetical processes (e.g., to get 18 oz. they had to calculate $B - A - 2C$), they were unlikely to solve a simpler transfer problem in the most straightforward manner. In other words, they transferred a nonoptimal process. According to feature overlap theory (see Halpern et al., 1990), this type of overtransfer is more likely to occur in near contexts than in far ones because the former may share both surface and deep features, making it more likely that the participant will detect a surface match between the formulas needed in the two situations and thus go no further. In contrast, when a far context is involved, the source and target may share no surface features, only deep ones, prodding the participant to effortfully process the source and target until an underlying common feature is detected, resulting in the most economical solution, rather than the dyseconomic solution based on surface features.

Thus, a cursory review of the results of selected analogical transfer experiments suggests a mixed conclusion, with many successes but some notable failures, and many studies showing mixed results depending on the circumstances. We return to the sources of this conflicting state of affairs after reviewing studies from other bodies of research relevant to transfer.

Formal Discipline

Experiments designed to test the doctrine of formal discipline, described earlier, have also provided evidence of successful and unsuccessful transfer. Lehman and Nisbett and their colleagues (Fong, Krantz, & Nisbett, 1986; Lehman, Lempert, & Nisbett, 1988; Lehman & Nisbett, 1990; Nisbett, Fong, Lehman, & Cheng, 1987) have conducted a group of experiments to evaluate the doctrine—namely, the claim that education can build general, transferable reasoning skills. In a study investigating the benefits of graduate education, Lehman et al. (1988) found successful transfer for some disciplines but not for others and concluded that "a version of the formal discipline hypothesis is correct" (p. 431). They interpreted their results in terms of pragmatic inferential rules shared by academic fields. They hypothesized these rules to be less general and abstract than the rules of formal logic but not tied to a specific domain of knowledge. For example, relevant reasoning schemas might include "causal schema and their associated evidence-checking procedures, and contractual schema, including the permission and obligation schema" (Lehman et al., 1988, p. 439).[6]

VanderStoep and Shaughnessy (1997) conducted a similar study with undergraduates in which they found a benefit of taking a course in research methods for tests of real life methodological and statistical reasoning. They also interpreted their findings as a positive test of teaching transferable, somewhat general thinking skills.

In a rare demonstration of transfer to the home environment, Fong et al. (1986, Experiment 4) tested students who were enrolled in a statistics course on everyday reasoning versions of statistics questions. This was done over the phone at home, and the test was disguised as a household survey about sports. Thus, it involved a different medium as well as a different ostensible purpose. Students who were contacted at the end of the semester performed better than those tested at the beginning of the semester, showing a benefit of attending the course that was successfully transferred to a very different environment, though only for some questions. The authors noted that they did not have an explanation for why some questions showed successful transfer and others did not.[7] In summary, many of these experiments investigating the doctrine of formal discipline have generally been successful at demonstrating transfer, though some studies have found mixed results.

[6] Unfortunately, the validity of the tests used to assess transferred skills in these studies has been questioned. Koslowski et al. (2000) have argued that the conclusions may not be justified, particularly for the statistical skill measure, in which the correct answers (defined by the experimenters and validated by outside experts in statistics and methodological reasoning but not in the subject matter being examined) appear to be based solely on the use of statistical principles without reference to underlying theory or mechanism, even though in some cases the correct answer may arguably be ambiguous without further background information. For example, participants asked why a batter who is hitting .450 after the first 2 weeks of the season cannot maintain that unheard of average throughout the entire season may choose options other than the one based on the law of large numbers for legitimate reasons based on their real-world knowledge of baseball. However, the experimenters claimed validation of their tests by correlations of their measures with IQ tests.

[7] Unfortunately, the Fong et al. (1986) article does not describe the questions in detail, nor does it specify whether the questions were equated for initial difficulty prior to transfer or whether the materials covered in the training course were equally applicable to the various transfer test questions, making it difficult to evaluate why the results were mixed.

Teaching Intelligence and Higher Order Skills

Transfer of reasoning skills has also been demonstrated somewhat successfully in studies focusing on teaching intelligence and higher order skills. A number of researchers have attempted to teach intelligence and related meta-cognitive skills, including “critical thinking” (e.g., Halpern, 1998; Herrnstein, Nickerson, de Sanchez, & Swets, 1986; Perkins & Grotzer, 1997; Sternberg, 1985; Williams et al., 1996; Winne, 1995). To the extent that efforts to teach intelligence are attempting to teach generally applicable thinking skills rather than cynically teaching to the test (i.e., to merely improve test scores, as some test-cramming programs are purported to do), they can be regarded as tests of general transfer of reasoning skills. Similarly, training in meta-cognitive and critical-thinking strategies is typically intended to be generally applicable and therefore also an experiment in transfer. Indeed, the National Research Council Committee on the Development of the Science of Learning (Bransford et al., 1999) concluded that transfer can be facilitated by training students in meta-cognitive awareness through various activities, such as reciprocal teaching, that encourage introspective awareness and self-monitoring.

One of the most well-known theories of intelligence involving meta-cognitive skills is Sternberg’s triarchic theory (see Sternberg, 1985). Sternberg and his colleagues have developed several different curricula designed to teach students how to think from an analytical, creative, or practical perspective. One such curriculum focused on teaching high school psychology (Sternberg, Ferrari, Clinkenbeard, & Grigorenko, 1996); another focused on improving middle school students’ academic success in all subject matters by teaching from a practical-thinking perspective (Williams et al., 1996). Transfer was found to regular academic skills such as reading and writing. Sternberg (1988) also described tests of other aspects of his theory, including successful tests of a program to teach so-called learning-from-context skills to adults and a program to teach insight skills to children. The latter program included a test of successful transfer to novel problems with different content.

Halpern’s (1998) critical-thinking program is another attempt to boost general reasoning skills by applying findings from cognitive psychology “to promote the learning of transcontextual thinking skills and the awareness of and the ability to direct one’s own thinking and learning” (p. 451). Training includes skills such as verbal reasoning, argument analysis, hypothesis testing, probability, and decision making. Training to promote transfer focuses on awareness of which skills to use when and provides practice with a wide variety of examples, combined with corrective feedback and elaborative questioning, to develop rich, interconnected knowledge structures—the deep understanding that Brown (1989) suggested is important for transfer. Halpern (1998) mentioned evidence for “gains in adult cognitive development . . . , superior responses to novel open-ended questions, . . . and changes in the organization of information” (p. 451) but acknowledged that the true test of her program’s success would require evidence of transfer. Unfortunately, transfer of skill improvements to other situations has not yet been evaluated.

Other researchers have investigated the benefits of other meta-cognitive training programs, such as self-explanation and self-regulation strategies. They have shown that training programs do increase strategy usage and that increased strategy usage improves performance on trained tasks (e.g., Bielaczyc, Pirolli, & Brown, 1995). However, this does not constitute transfer. In one notable study, Adey and Shayer (1993) did find evidence of transfer from science class to English examinations taken 2 years after their intervention to train meta-cognitive and higher order thinking skills. Also, Herrnstein et al. (1986), in a comprehensive evaluation of a program to teach thinking skills to seventh-grade Venezuelan school children, found that program benefits transferred to a creative design task and a practical reasoning question about personal health. Neither of these transfer tasks was covered by their training program. Degree of transfer depended on a number of factors, including the closeness of the transfer task context to the training context, in terms of aspects such as modality (multiple choice test vs. practical design task). In addition, Hamers, de Koning, and Sijtsma (1998) trained Dutch third graders on an inductive reasoning task focusing on generalization, discrimination, and cross-classification of attributes and relations among objects. These investigators found transfer from their training to performance on Raven’s (1958) matrices (a visual type of IQ test that also entails visual cross-classification of multiple dimensions), which they referred to as far transfer. Thus, there is some evidence of successful transfer of higher order, intelligence-related reasoning skills.

Impact of Schooling

The fourth area of research relevant to the question of transfer is that of the impact of formal schooling. Schooling does show some transferable benefits, although the exact nature of the skills transferred is unclear.

If schooling has positive effects on measures other than those directly taught, this could be construed as evidence of transfer. Benefits of schooling have been demonstrated in a number of more and less controlled ways. Ceci and Williams (Ceci, 1991; Ceci & Williams, 1997) have shown that schooling improves IQ independently of the fact that higher IQ tends to lead to more schooling. They examined studies of individuals with more or less schooling for reasons out of their control or for choices not confounded with IQ, such as lack of access to school due to remote location of homes, differential school access due to enrollment cut-off dates, schools closed due to war, and staying in school to avoid the draft. Results indicated that being out of school when others of your cohort are in school decreases your IQ. Also, schooling appears to increase income over and above its effect on IQ (Ceci & Williams, 1997), suggesting transfer to an out-of-school environment is perhaps driven by an impact on a wider range of individual characteristics, such as motivational and temperamental factors.

Although nonexperimental evidence such as this does not reveal what the particular skills are that are learned in school and transferred to later IQ tests and jobs, it suggests that something enduring is transferred from school-based instruction. Whether that benefit is cognitive, motivational, or simply due to educational credentials is unclear.

Discussion of These Four Areas of the Existing Literature

From the foregoing review of key studies in four areas of transfer, it can be seen that there is often evidence of successful transfer of one sort or another from studies using a variety of methodologies, contents, and participants. To be sure, however, there are gaps in such studies, and there are some noteworthy failures to observe transfer. Despite this array of evidence, researchers familiar with this literature have been particularly negative about the pervasiveness of transfer. Consider these examples:

> Significant transfer is probably very rare and accounts for very little human behavior. . . . There is almost no evidence to support the educational philosophy of formal discipline. . . . The goal of education would not be to develop higher order processes. . . . I view education . . . as the learning of information. (Detterman, 1993, pp. 17–21)

> Flexible use of knowledge is often cited as a hallmark of human intelligence. At the same time, transfer is thought to be an elusive quality. (Brown, 1989, p. 369)

> Transfer is hard to come by, particularly far transfer. (Perkins & Grotzer, 1997, p. 1129)

> By and large, subjects fail to solve the target [transfer] tasks unless prodded to use previous solutions. (D. Carraher & Schliemann, 1998, p. 4)

Given the wealth of apparent evidence, it is surprising that so many researchers have expressed such pessimism. The answer to this seeming riddle may lie in the gaps and failures noted above, leading to the qualifiers used in these quotes, for example, “significant,” “flexible,” “far,” and “unless prodded.” What do these terms mean, and just how significant, flexible, unprodded, and far does transfer have to be to count?

If transfer is the standard used to justify the human and financial investment in education and training programs, then it must apply well beyond the environment of training—that is, far transfer is required. Finding evidence of transfer from today’s math class to tomorrow’s math class is not sufficient. Indeed, Detterman (1993) dismissed all examples of near transfer as being trivial and instead focused on failed attempts to demonstrate far transfer. Perkins and Grotzer (1997) also focused on the issue of near versus far transfer. They reviewed programs intended to teach intelligence and to improve thinking by teaching strategies and heuristics and meta-cognitive awareness. On an initially positive note, they stated, “Many studies have demonstrated that targeted interventions can teach people to think better within particular subject matters and in some general ways as well, with transfer beyond the kinds of tasks used in instruction and moderate persistence” (p. 1125). However, as mentioned above, these authors concluded with the pessimistic comment that transfer, particularly far transfer, is hard to come by.

What, then, is far transfer, and do any of the studies in the literature show it? Unfortunately, defining the terms *near* and *far* is no simple matter, as they are usually based on the intuitive notion of *similarity*, which is itself ill defined (Murphy & Medin, 1985). Using the number of elements two domains have in common as a precise measure of similarity, as suggested earlier by Thorndike (1906) and more recently by various processing models of analogical transfer (e.g., Forbus et al., 1995), requires that all the elements of the two domains have been mapped and manipulated, which is generally not practical in the complex, real-world domains we address in this review. Of course, the notion of similarity is still not totally avoided by such approaches, as the degree of similarity between individual elements then needs to be assessed. Furthermore, these theories do not explain why aspects of the context other than knowledge domain might have an effect. Given these concerns, the approach we have taken is to reduce the summary measure—near versus far—to a number of underlying dimensions, as we reveal later. Judgments of distance along these dimensions still involve the notion of similarity, but at least we can be specific about which dimension the similarity or dissimilarity is related to, even if we cannot precisely operationally define similarity itself.

With this in mind, it is possible to argue that many demonstrations of successful transfer involve only near transfer. For example, Brown’s (1989; Brown & Kane, 1988) mimicry experiments mentioned earlier only involve transfer within the same domain—from one animal to another. Far transfer is often said to be transfer between domains (although note that the notion of *domain* itself is also ill defined; Sternberg, 1989). According to this criterion, transferring this concept to the design of military hardware might constitute far transfer. Similarly, Hamers et al.’s (1998) findings may also be considered near transfer, as the training exercises they used appear very similar to those involved in solving their transfer test (Raven’s, 1958, matrices). The fact that these authors considered their experiment to demonstrate far transfer highlights the confusion in the definition and use of this term.

Indeed, the terms *near transfer* and *far transfer* have been used to mean very different things by different researchers. For example, Campione, Brown, and Bryant (1985) operationalized near transfer as going from learning a “spy” code of a scrambled message that had a specified periodicity (e.g., two-letter intervals between code letters) to reading another scrambled message that had the exact same periodicity. Conversely, they defined far transfer as switching to a different periodicity such as three-letter intervals. On most measures, this would not be considered far transfer on any dimension.

Two more convincing examples of far transfer are those mentioned earlier by Chen and Klahr (1999) and Fong et al. (1986). Chen and Klahr’s demonstration of what they termed *remote transfer* of the control of variables strategy by elementary school children involved a transfer test that was conducted several months later, by different experimenters, using different domains and a different testing format. In Fong et al.’s experiment, the training context was a university course and the transfer task was conducted at home in the guise of a household phone survey. It could perhaps be argued that both tasks were in the same domain because both involved the application of statistical understanding (although the superficial domain of the transfer test was sport), but that could not be avoided as statistical reasoning skills were what was being transferred. If far transfer is defined in terms of the goal of gaining skills that transfer from school to contexts outside the academic environment, this surely meets the criterion of far transfer. Similarly, the impact of schooling on IQ might constitute far transfer, depending on where and when the IQ test is conducted, if the skills measured are assumed to be in a different domain from school lessons, and the impact of schooling on future income might also suggest some sort of far transfer. However, whether what is transferred is a problem-solving skill, a change in motivation, or some sort of accreditation effect (i.e., employers hire highly schooled individuals for the best-paying jobs) is unclear from these data alone. Additional candidates for the label *far transfer* are the studies by Adey and Shayer (1993) and Herrnstein et al. (1986). The former involved transfer from meta-cognitive training in science class to English examinations 2 years later, while the latter found transfer from a thinking-skills training program to a creative design task. As can be seen, the sense in which the transfer examined in these studies is far varies considerably, from differences in domain to differences in physical location, delay interval, purpose, and modality. These are some of the elements that form the core of our taxonomy.

In addition, there are concerns about the lack of spontaneity of the transfer in many of the studies. Transfer from one situation to another, only after the experimenter prods a participant with a hint to indicate the isomorphism between the two situations (e.g., "this stock market task resembles the racetrack task you already know," Ceci & Ruiz, 1993, p. 177), can be dismissed as merely following directions, rather than spontaneous and flexible deployment of knowledge to a distant domain. Correspondingly, Detterman (1993) raised the valid concern that many experimental demonstrations of successful far transfer are trivial because they were not spontaneous (and therefore do not count as significant) and came about only after the provision of hints, which are not likely to be provided in the real world. According to Detterman, all of the analogical reasoning tasks that are cited as evidence for transfer fail on this criterion, as they all involve some sort of hint or instruction to apply the learned knowledge. To quote Detterman, "Telling subjects to use a principle is not transfer. It is following instructions" (1993, p. 10). Thus, in addition to determining whether a study tested far transfer, it is also important to distinguish between those studies in which the transfer is spontaneous and those in which it is prodded.

Gick and Holyoak (1980) and Reed et al. (1974) both investigated the effectiveness of pointing out the connection between training and transfer problems in their studies. The former merely used a hint that the training problems might be helpful, whereas the latter actually explained the parallels between the two situations. In both cases transfer was significantly improved. Gick and Holyoak also investigated the impact of having to recognize which training example out of three provided was relevant to the transfer task in comparison with a situation in which there was only one training example, with unclear results. Brown and Kane's (1988) series of studies also investigated some aspects of the issue of transfer spontaneity. In a first study, they presented matched pairs of problems to draw attention to the similarities of solution procedures that could then be transferred. In contrast, in a second experiment they presented three training problems followed by three transfer problems (e.g., training A, B, C followed by transfer A, B, C rather than paired problems: training A, transfer A, training B, transfer B, etc.). In this second experiment, the children had to recognize which of the three solution methods was appropriate in each case, not just repeat a previously successful approach. Thus, these latter results show a slightly more complete form of transfer.

This difference between being told what to do during a transfer task and having to recognize the appropriate approach oneself underlies many of the spontaneity issues raised. Thus, the spontaneity question may be better characterized as a measure of the memory demands of the test—that is, the difference between merely being able to execute a problem-solving approach or procedure versus being able to recall it and/or recognize when it is applicable.

In addition to these questions of near versus far transfer and of the memory demands of the task, our earlier review of research suggested a third issue—the nature of the skill being transferred. In particular, the research suggested it might matter whether the skill is a specific fact or procedure or a deeper, more general principle. For example, the procedure "when you add two numbers together, carry one over to the next column if they add to more than nine" is somewhat specific, whereas the principle "break down the problem into subproblems" is much more general. A procedure for getting the missionaries safely across the river is more specific than statistical reasoning principles, such as the law of large numbers, which are in turn more specific than meta-cognitive skills, such as "look at a problem from multiple angles."

This dimension—the nature of the skill being transferred—is often conflated with the near versus far dimension. Detterman (1993) set up the distinction between the two but then later equated them in his arguments against the existence of far transfer (equating general transfer with far transfer). However, there are potentially important differences between the two that deserve to be noted. Near versus far relates to the similarity of the training and transfer situations; in contrast, the dimension "specific versus general" relates to how generally applicable the learned information is—that is, specific facts or procedures versus general skills, principles, or strategies. Specific facts or procedures, described in terms of superficial aspects of the problem, are applicable to specific situations only, whereas principles, described in terms of deep, structural aspects of the problem, are usually more generally applicable. Far transfer, as defined here, can theoretically be either specific or general, and general skills can be transferred to a near or a far context; thus, the two dimensions must be separated.

It is clear from the comments mentioned above that a useful taxonomy must not just draw attention to the differences between experiments to reconcile differences in results but must also identify gaps between what the studies have demonstrated and what consumers of the research require to justify the generalizations they wish to make. This is the question of external validity. If there is a gap between the situations investigated in the research and the situations to which generalizations are to be made, the nature of this gap must be documented. Next, it is important to determine whether the gap should be a source of concern or can be dismissed as immaterial. If there is reason to believe (whether from transfer studies or other research) that some components of this gap may be material, that is, if there is reason to believe transfer success may be affected, then the gap between what has been demonstrated and what needs to be demonstrated would constitute a call for research that must be addressed if the transfer debate is to be resolved. We now turn to our proposed taxonomy, which reflects the dimensions discussed.

A Taxonomy of Transfer

> The essence of any scientific procedure is classifying and quantifying in such a way as to reveal order in the data. (Meehl, 1995, p. 269)

In this section we propose a taxonomy of the various dimensions along which studies may be organized, explain why these dimensions were chosen, discuss possible interactions between dimensions, and apply the taxonomy to the transfer literature.

Dimensions of the Taxonomy

The preceding discussion of ways in which transfer studies differ raises three issues. First, it suggests that the question of whether a particular study tested near or far transfer is more complex than might be initially thought. *Near* and *far* can mean many different things, and researchers are not consistent in their usage. Second, it suggests that the memory demands of the task—the manner in which use of transferred knowledge is tested—may affect transfer success and thus may need to be explicitly considered. Third, it cautions that the issue of whether the skill to be transferred is specific or general should not be confounded with discussions of whether the task constitutes near or far transfer. They are logically separable factors that may both affect transfer. Thus, it highlights a need to consider separately the transfer of specific facts and procedures and of general principles. These three observations form the nucleus of our proposed taxonomy of transfer.

Given the myriad dimensions along which transfer experiments can be distinguished, we suggest that the characteristics of transfer can be thought of as breaking down into two overall factors: the content—that is, what is transferred, and the context—that is, when and where it is transferred from and to (see Figure 1). As we explain below, the former factor encompasses issues regarding spontaneity and specificity–generality. The latter factor both captures and extends the near-versus-far-transfer distinction. Armed with these two global factors, each of which is subdivided into multiple independent dimensions, we attempt to provide some degree of mental hygiene to the debate surrounding the question of central importance in this review, namely, does far transfer occur, and if so, under what conditions?

A Content: What transferred

Learned skill	Procedure	Representation	Principle or heuristic
Performance change	Speed	Accuracy	Approach
Memory demands	Execute only	Recognize and execute	Recall, recognize, and execute

B Context: When and where transferred from and to

	Near ←				→ **Far**
Knowledge domain	Mouse vs. rat	Biology vs. botany	Biology vs. economics	Science vs. history	Science vs. art
Physical context	Same room at school	Different room at school	School vs. research lab	School vs. home	School vs. the beach
Temporal context	Same session	Next day	Weeks later	Months later	Years later
Functional context	Both clearly academic	Both academic but one nonevaluative	Academic vs. filling in tax forms	Academic vs. informal questionnaire	Academic vs. at play
Social context	Both individual	Individual vs. pair	Individual vs. small group	Individual vs. large group	Individual vs. society
Modality	Both written, same format	Both written, multiple choice vs. essay	Book learning vs. oral exam	Lecture vs. wine tasting	Lecture vs. wood carving

Figure 1. Taxonomy for far transfer.

Content: What Is Transferred

The content factor can be further broken down into three dimensions: (a) the specificity–generality of the learned skill, (b) the nature of the performance change assessed, and (c) the memory demands of the transfer task. Not all of these apply to all situations.

Learned skill. The first dimension of transfer content is the specificity–generality of the learned skill (see Novick, 1990). That is, whether it is a specific fact or routinized procedure, a form of representation, or a more general problem-solving heuristic or principle. A specific procedure might be characterized as a set of particular steps described in terms of superficial features, whereas a general principle might be characterized as a deeper, structural, or causal understanding. Examples of a specific routinized procedure would be an algorithm for getting the missionaries safely across the river in the missionaries and cannibals problem (Reed et al., 1974) or an equation for calculating proportions (e.g., school-taught "rule of three"; Schliemann & Nunes, 1990). An example of transfer content that is a form of representation would be a matrix (Novick, 1990) or a tree diagram. Examples of more general problem-solving heuristics or principles would be the control of variables strategy (Chen & Klahr, 1999), statistical principles (Fong et al., 1986), hierarchical classification (Herrnstein et al., 1986), pragmatic inference rules (Lehman et al., 1988; Lehman & Nisbett, 1990), and study skills such as "check your work" (Williams et al., 1996).

Specificity–generality is a continuum, and gradations are possible within each of these categories. For example, it is possible that training programs intended to teach general principles might sometimes result in encoding the principles in the form of more specific facts or procedures tied to the examples used in training, as suggested by Ross (1987) and Reeves and Weisberg (1994). Indeed, Bassok and Holyoak (1989) found that the same basic algebraic formula could be learned in a more general form if taught in the context of algebra and in a more specific form if taught in the context of physics. The more general form showed superior transfer to novel contexts. Thus, it is important to ascertain what the participants learned before attempting to assess whether they could transfer what they learned because transfer performance might be expected to differ depending on whether they learned a specific fact or procedure or a more general principle.

To complicate matters, this dimension could also be characterized in other ways—for example, in terms of deterministic, well-defined transfer (roughly equivalent to transfer of procedures) versus probabilistic, ill-defined transfer (roughly equivalent to transfer of strategies, principles, or heuristics). Another name for this dichotomy might be *formal* versus *pragmatic transfer*, a distinction between, on the one hand, formal learning, which occurs in closed-problem spaces (the constituents are narrowly defined), and pragmatic (or probabilistic) learning, which involves heuristics for ill-defined, fuzzy-problem spaces. In formal-learning situations, the elements are (a) the boundaries between admissible and inadmissible states, (b) the starting state, (c) transformation rules, and (d) the end state. In addition to procedures such as those used to solve the missionaries and cannibals puzzle, examples of formal approaches to transfer could be the application of componential approaches, Piagetian analyses, and factor analytic approaches. On the other hand, in a pragmatic (or probabilistic) approach there is usually some degree of ambiguity with respect to the starting state and the transformation rules. Examples of pragmatic approaches include the estimation of subjective probabilities and the application of various reasoning heuristics (such as the law of large numbers) to situations that are complex. Unlike the pragmatic approach, the formal approach permits formalized problem representations (e.g., principles of logic).

In sum, in a formal approach, such as Piaget's pendulum problem (Inhelder & Piaget, 1958), the solution derives from a deterministic combinatorial analysis, whereas pragmatic solutions are often based on informal heuristics. Classic problems such as the Duncker (1945) tumor problem, as well as Tversky and Kahneman's (1974) representativeness heuristic, would be difficult to solve by writing a formal computer algorithm because the decomposition process involved in transferring the X-ray-tumor context to the army-destroys-the-fortress context cannot operate abstractly but must take into account the broader relational context (Holyoak, 1984, made this point). For instance, nowhere in a participant's lexicon does the term *fortress* convey the necessary information that it is also a "target" much the way the tumor is a target of the X-rays. Participants must realize that in transferring the solution from the tumor–target problem to the fortress problem the word *fortress* "is a target solely by virtue of its semantic role" (Holyoak, 1984, p. 210). It would be difficult for a formal computer algorithm to solve such problems without inputting exhaustive contextual knowledge.[8]

Performance change. The second dimension of transfer content, the nature of the performance change, refers to the measure against which improvement is expected; it could be the speed of execution of some activity, the accuracy and quality of execution, or what is executed (doing or not doing a particular thing). Most of the reasoning tasks discussed earlier focused on what is executed and how accurately it is done. For example, researchers looked at whether students used the law of large numbers when tackling statistical reasoning problems, and if so, how well they applied it (Kosonen & Winne, 1995). However, speed of execution is occasionally used as an outcome measure in reasoning tasks. For example, in Reed et al.'s (1974) work with analogical transfer, the time taken to reach a solution was one of the measures. In that study, transfer success was found to differ depending on the measure used—training with no hints resulted in a reduction in the number of errors made, whereas training accompanied by an explanation of the relevance of the training to the transfer task also resulted in improvement in terms of solution speed (if the training problem was the more difficult of the two). Depending on the situations to which research findings are to be applied or generalized, one or other of these measures of performance change might be more relevant.

Memory demands. The third dimension of transfer content, the memory demands of the transfer task, encompasses the issue of spontaneity that we discussed earlier. Specifically, does the transfer test require the individual merely to be able to execute a learned activity, prompted by hints as to the correct procedure to apply, or are participants also required to select the appropriate approach? In other words, do participants also need to recognize when it is appropriate to use a particular approach? Furthermore, if what is to be done is not specified, are participants required to merely pick from a selection offered, or do they need to spontaneously recall the approaches themselves? This is akin to the distinction between recognition and recall in memory research. Experiments without explicit or implicit hints that the training and transfer phases are related tap recall of a learned skill and its applicability, as well as the ability to execute. In contrast, experiments that hint that recently learned approaches may be useful in a subsequent transfer task tap either execution or recognition plus execution, depending on the number of different approaches learned previously. If the training phase used three different types of strategies, then the participant would have to recognize the correct one before executing it on the transfer test. If the training phase taught only a single approach, recognition of the correct one would not be required for success on the transfer task. The participant must merely execute the learned skill when prompted to do so.

Most of the formal discipline and higher order skills experiments did not prompt or hint which approaches were appropriate; therefore, they would be considered to test recall as well as ability to execute. In contrast, Brown and Kane's (1988) studies, discussed earlier, tested either execution alone or recognition and execution.[9]

Summary. In summary, the content tested in transfer experiments can be grouped into fundamentally different types of skills along a continuum from specific to general. In turn, improvements in performance of these skills can be measured in a variety of ways, such as speed, accuracy, and simply doing the right thing. Finally, any given skill can be tested for recall, recognition, or merely prompted execution. Figure 1A summarizes these variations of content.

[8] Another possible content dimension might be called *vertical versus horizontal transfer*. By *horizontal transfer* we mean transfer that involves two instances or tasks at the same level of complexity (e.g., going from learning one text editor to learning another in Singley & Anderson's, 1989, study). By *vertical transfer* we mean something like learning a memory strategy that can then be deployed in a wide array of tasks that depend on it, even if the tasks are very different in complexity—they all will fail if the strategy is not used. Another example is Piaget's (1952; Piaget & Inhelder, 1969) stages, in which once a child enters a stage all of the tasks in that stage that tap a given scheme are supposedly facilitated (e.g., a concrete operational child should theoretically be able to solve a very wide array of tasks, some entailing conservation, some entailing seriation, some entailing other feats). In other words, Piaget's stages can be seen as underlying knowledge structures that influence all activities above them (hence, as vertical forms of transfer). Note that the nature of such vertical structures could be either procedural or principled, and it could be deterministic and well-defined (i.e., procedural) or probabilistic. All that is assumed is that the problem space is organized so that all tasks falling below a given structure require it for their solution. Hence, Piagetian attainments can be thought of as vertical and procedural or principled. For simplicity we have not pursued this additional dimension here, as it was not explored by many of the studies described.

[9] There is a variant of this dimension that we have not pursued further because studies rarely provide the necessary classification information to do so, namely, whether the transfer is conscious or unconscious. This variant could also be termed *active* versus *passive transfer*. A few studies have tangentially investigated this issue. For example, Reed et al. (1974, Experiment 3) asked participants how much they had used the answer to the training task when working on the transfer task after being told the relationship between the two. They found that ratings did not correlate with performance on the transfer task, even though some transfer had occurred.

Context: When and Where Learning Is Transferred From and To

As was the case for the content of transfer, the context of transfer, or when and where learning is transferred from and to, can also be broken down into a number of dimensions. These include knowledge domain, physical context, temporal context, functional context, social context, and modality. We discuss each of these in turn.

Knowledge domain. The knowledge domain refers to the knowledge base to which the skill is to be applied, such as English class versus history class. For example, Adey and Shayer (1993) examined evidence of transfer from science class to English examinations. Physics to chemistry might be considered nearer transfer than physics to English, as more elements would presumably be shared. However, as mentioned earlier, the notion of domain is itself ill defined (Sternberg, 1989). Having broken down the notion of similarity of context into these various components, and to add structure, we now face the exact same issue with respect to defining distance along each of the component dimensions. Thus, Murphy and Medin's (1985) problem of defining similarity can be applied recursively at greater and greater levels of detail. As more research effort is directed at these issues, we hope that many of our dimensions are broken down again into subdimensions, adding even more structure and further reducing the subjectivity that ineluctably is part of our discussion of similarity.

Physical context. Both macroaspects, such as whether the training and transfer phases are conducted at school, in a research lab, in the home environment, and so on, and microaspects, such as whether the exact same room is used and whether the experimenter is the same, make up the physical context. Both types of physical cues could affect the success of transfer. For most of the studies reviewed here, transfer and training were conducted in the same macrocontext, usually the school, although one study, that by Fong et al. (1986), did also test transfer to a home environment.

Temporal context. This dimension reflects the elapsed time between training and testing phases (e.g., a few minutes, a week, or years later). As was mentioned earlier, most of the studies discussed here tested transfer soon after training, with a few noted exceptions. However, if it is to justify the effort invested in education, ideally one would hope for transfer to last for several years after training. In addition, there are other aspects of the temporal context which may have an effect but which are usually ignored in transfer studies, such as the time constraints, if any, on the participants. It is possible that training conducted in a context with no time constraints would transfer to a non-time-constrained transfer test better than it would transfer to a time-constrained test and vice versa, for example.

Functional context. The function for which the skill is positioned and the mind-set it evokes in the individual can be referred to as the functional context. For example, is it positioned as an academic activity or one belonging to the "real world" outside academia? Is the transfer task explicitly a test, or is it embedded in some daily activity? Problem-solving tools learned and encoded for one purpose might not transfer equally well to another. This dimension is conceptually related to the notion of *functional fixedness* (Duncker, 1945), in which the use of tools is tied to their original purpose. Although functional context is potentially confounded with physical context, it can be differentiated—not all school-based tasks are academic. For example, a transfer task carried out in an after-school child-care program could be nonacademic in style or function even if it is in the physical context of school, whereas an exam-like test conducted at home would be academic in style but in a nonschool physical context.

Fong et al.'s (1986) study using a fake household phone survey as a transfer task for academic training is an example of a study that qualifies as far transfer on this dimension. Another is Ceci and Ruiz's (1993) study of transfer from racetrack handicapping to stock market analysis. The participants' handicapping skill, developed while gambling real money on horse races, served a very different function in their lives than the artificial transfer task posed by the experimenters. This is the essence of the functional dimension. Differences in functional context would likely also be accompanied by a difference in motivation, which could explain some of the performance differences found.

Social context. This dimension refers to whether the task is learned and performed alone or in collaboration with others. A skill acquired in a group setting might not be equally well applied when alone or vice versa. Studies reviewed here did not address this, and although there is an extensive separate literature dealing with collaborative learning (see, e.g., Gabbert, Johnson, & Johnson, 1986; Meudell, Hitch, & Boyle, 1995; Meudell, Hitch, & Kirby, 1992; Slavin, 1983, 1984; Tudge & Winterhoff, 1993; Tudge, Winterhoff, & Hogan, 1996), it is not well integrated with the transfer work. In their National Research Council report, Druckman and Bjork (1994) stated, "Little empirical attention has been given [in the collaborative learning literature] . . . to transfer to other tasks" (p. 12).

Modality. The final dimension of transfer context, modality also has both macro- and microaspects. At the macrolevel, the task can be visual or auditory, written or verbal, linguistic or hands-on (e.g., model building), and so on. At the microlevel, a task can be in multiple-choice format or essay format, and so on. (It is possible that participants might, e.g., apply training to check their work by reviewing choices in a multiple-choice format but not transfer the skill to a less structured essay format.)

Summary. In conclusion, we return to the question of how far apart the contexts of training and transfer have to be to count as far transfer. As we have discussed, there are (at least) six dimensions along which the differences between training and transfer contexts in transfer studies may be compared. A transfer task may satisfy the requirements for far transfer on some of these dimensions but not on others. To be classified as true far transfer, does a study have to qualify as far on all dimensions? For example, does a far transfer task have to involve training in domain A, at school, on a paper-and-pencil task, in academic mode, conducted alone, followed by a transfer assessment applied to domain B, embedded in a play activity, next year, at home, while chatting casually with some friends? We could decide to reserve the label *far transfer* for studies that satisfy all these criteria (which might be very rare), but it is perhaps more fruitful to avoid use of the summary term and instead specify whether the transfer situation is near or far along each dimension. In this way, more precise evaluations of study results could be performed.

Figure 1B illustrates how such dimensions might be plotted. There are many ways in which aspects of the training and transfer environments could be plotted in such a figure. This is just one example. Note also that at this point in our understanding there is no necessary scalar relationship between relative distance along different dimensions or between different points along the same dimension; that is, the scale is subjective and can be different for each dimension.

Figure 1B is intended to illustrate the multiplicity of dimensions along which contexts may differ. Development of operational measures of distance along each dimension requires substantial further research and is beyond the scope of this article. At this stage, our goal is to draw attention to these dimensions and to justify why they may be relevant to understanding and applying the transfer literature, as well as to document how they are often ignored or muddled, to hopefully generate rigorous and focused research on these issues. We make no pretence of mathematical purity or continuity of data or linearity of dimensions. Instead, our taxonomy is avowedly pragmatic and illustrative.

Rationale for Dimensions

Our rationale for these dimensions is derived from three sources: the transfer literature reviewed earlier, evidence from other areas of psychological research that might have implications for transfer, and factors important for the theoretical and applied questions to which reviewers and consumers of transfer research have attempted to generalize the findings.

In terms of content, various transfer studies discussed earlier showed that whether the learned skill is encoded as a specific fact or procedure or a deep, general principle affects transfer success. Transfer studies also showed that transfer success may differ depending on the performance measure used. Finally, the discussion of hints and spontaneity demonstrated both that transfer performance varies considerably depending on the memory demands of the task and also that the use of hints—and hence changing the transfer test from recall to mere execution or recognition of the learned skill—is a concern regarding the generalizations to which the results can be applied. Thus, all three of these content dimensions merit a place in the taxonomy.

Regarding context, it is almost a given among researchers in this field that knowledge domain and temporal context affect transfer (Ceci, 1996) and that these dimensions clearly have practical relevance to transfer from school classes to future life, as does physical context. In addition, the fact that participants can perform very differently in a different physical context was illustrated by a study by Ceci and Bronfenbrenner (1985) in which children were more efficient at the exact same task, estimating time to make cupcakes (or charge a motorcycle battery), at home than in an unfamiliar university laboratory. This was not strictly a transfer study in that no explicit training phase was involved, but in some sense it was testing the transfer of skills from wherever they were learned to two different macrolevel physical contexts. Results showed that physical context mattered. In the studies reviewed here, microaspects of physical context were rarely mentioned; however, a notable exception was the study by Chen and Klahr (1999) in which the remote transfer test was intentionally conducted by a different experimenter (in this case the class teacher) in a different room at the school. This suggests a concern by the researchers' that microaspects of the physical context might play a role in the success or failure of transfer. Spencer and Weisberg's (1986) work also provides support for this concern. They found that changing a small detail of the physical context (the experimenter) between training and transfer phases had a negative impact on transfer success.

Further evidence comes from Rovee-Collier's (1993) work with infants' memories for the association between kicking their feet and the movement of a mobile hung over their cribs. Minor changes in context as trivial as the change from a yellow crib liner with green squares to a yellow liner with green circles were found to completely disrupt the association at a later test of the kicking behavior. Ceci and Bronfenbrenner (described in Ceci, 1996) also found that their 10-year-old video game participants could only transfer their learning on a computer-based task from a video game format to an isomorphic, but less meaningful, geometric-tracking task format if the exact same physical environment was used for the transfer test (same room, same computer, same mouse). Changing minor details of the physical environment disrupted the transfer.

There is less empirical evidence for the importance of functional context, as researchers have rarely investigated it, and when they have (e.g., Ceci & Ruiz, 1993; Fong et al., 1986), it has been confounded with other differences. However, we included it in our taxonomy because the functional fixedness work, mentioned earlier, suggests that it might affect transfer performance, and the applied questions related to the justification of the investment in general education suggest it would matter if it did.

Regarding social context, many believe that the social aspects of learning are important: "Part of this [situated learning] view is that learning is fundamentally a social activity" (Reder & Klatzky, 1994, p. 33). Also, Dunbar (1995, 1997) has shown that a moderately diverse group of individuals in a group reasoning situation can result in fruitful transfer and generation of joint insights, in which ideas provided by one group member can prompt recall of analogies by another member and yield novel insights that would be unlikely to have been generated by any of the members working alone (although he does not specifically compare group vs. individual problem-solving success). Furthermore, in some cases, lack of attention to this characteristic of the training situation has caused problems with interpretation of study results. For example, Hamers et al. (1998) compared the effectiveness of two training procedures, one direct (explicit teacher guidance) and one indirect (less explicit guidance), and concluded that the direct approach was superior. However, this result may be a consequence of the fact that they taught their participants in groups. In the indirect group, the participants were supposed to figure out the target principles themselves without being explicitly led to them by the teacher, and they presumably moved ahead when the group (i.e., at least one individual in the group) grasped the necessary principle. In such a design, many participants could be left ignorant and thus show poor transfer when tested individually. If the teaching method had required each participant to grasp the necessary principles in an individual session, or if transfer had also been in a group, the results might have been different. Thus, careful attention to external validity issues is required here—generalizations from one set of social contexts to another may be invalid.[10]

[10] The issue of external validity is complicated by the fact that in actual classrooms that use so-called cooperative education approaches (i.e., small groups learn to solve problems together), group performance may also be disproportionately influenced by the contribution of the top student.

Finally, the modality dimension was not explored in most of the studies reviewed here, although it was partially addressed in the study by Herrnstein et al. (1986). These researchers were concerned that using only multiple-choice tests for their transfer measure might not capture all aspects of learning. Therefore, they also included open-ended written questions and questions read aloud by the teacher as well as a practical design task and an oral argumentation task in their transfer measures. Although benefits for the experimental group were found on most of these measures, the size of the advantage was found to vary between test formats, with the largest benefits generally being found on tests closest to the original training. Thus, we believe that all six of these context dimensions merit a place in the taxonomy, at least until more conclusive evidence regarding their influence can be gathered.

Interactions Between Dimensions

A complication that we have avoided discussing so far, for the sake of simplicity, is the possibility of interactions between the various dimensions we have described. For example, spontaneity and memory demands may be more of an issue for some types of transfer than others. Specifically, the added difficulty of free recall of a learned solution, over and above the difficulty of recognition or execution only, may be greater for particular facts or procedures than for general principles because the latter might be retrieved by a wider variety of cues. Thus, an interaction might be expected between memory demands and type of learned skill transferred. Therefore, it is important to make it clear whether the skill to be transferred is specific or general when considering the import of spontaneity-based challenges to external validity. Indeed, in a related finding, Blanchette and Dunbar (2000) showed that participants spontaneously generated analogies based on deep, structural similarity in a free-recall task (without even an explicit training phase), whereas they generated more superficial analogies when tested with what the authors called the *standard reception paradigm*, in which source analogies were provided and the connection between the source analogies and the test was explicit. Thus, the memory demands of the transfer task influenced whether superficial or deep, structural learning was used.

Also, Chen (1996) has shown that the closeness of knowledge domains affects the success of transfer only under certain circumstances. He found that similarity of both superficial features (salient but solution-irrelevant details, e.g., the protagonist is either a girl or a genie) and procedural features (specific operational features involved in applying an analogy, e.g., two tools are combined by being either attached or tied together) enhanced transfer, but only if the participant had not encoded the principle to be transferred at an abstract level. On the other hand, abstract encoding transferred independently of superficial and procedural similarity, presumably because the encoded knowledge was not tied to any of the details of the training examples. Depth of encoding was assessed by the types of similarity the children reported noticing between the different problems. Thus, effect of domain distance also interacts with type of learned skill.

Applying the Taxonomy to the Transfer Literature

Detailed Discussion of Two Studies

As a first step toward demonstrating the utility of this taxonomy to both evaluate existing evidence for far transfer and suggest avenues for future research, we apply it to two of the studies reviewed earlier that provide sufficient documentation for this enterprise. These studies were selected because, unlike most published studies, they include some systematic exploration of the dimensions in the taxonomy. In view of the finding that successful far transfer may be more likely for general, deep principles than for specific, superficial facts or procedures, both these studies involved the transfer of heuristics or principles. Within-study comparisons such as these are likely to yield clearer conclusions regarding the impact of the various dimensions than between-studies comparisons, because studies also differ in many other ways, such as the inherent difficulty of the task and the effectiveness of the training, making between-studies comparisons often confounded. The first study, conducted by Gick and Holyoak (1980; see Figure 2), involved adults, whereas the second study, conducted by Chen and Klahr (1999; see Figures 3 and 4), involved elementary school children.

As discussed earlier, the Gick and Holyoak (1980; Experiments 4 and 5) study used analogues of the Duncker (1945) radiation problem to assess transfer of a principle (convergence of many weak forces yielding a strong force at the intersection), which they assessed by counting the proportion of trainees who used the correct approach in the transfer test as well as by mea-

A Content: What transferred

Learned skill			Principle or heuristic
Performance change		Accuracy (completeness)	Approach (use of correct approach)
Memory demands	Execute only (hint condition, single training story)	Recognize and execute (hint condition, multiple training stories)	Recall, recognize, and execute (no hint condition)

B Context: When and where transferred from and to

	Near ←		→ **Far**
Knowledge domain		Military problem vs. medical problem	
Physical context	Same room		
Temporal context	Same session		
Functional context	Both experiment		
Social context	Both individual		
Modality	Both written/verbal		

Figure 2. Taxonomy for far transfer applied to Gick and Holyoak (1980, Experiments 4 and 5).

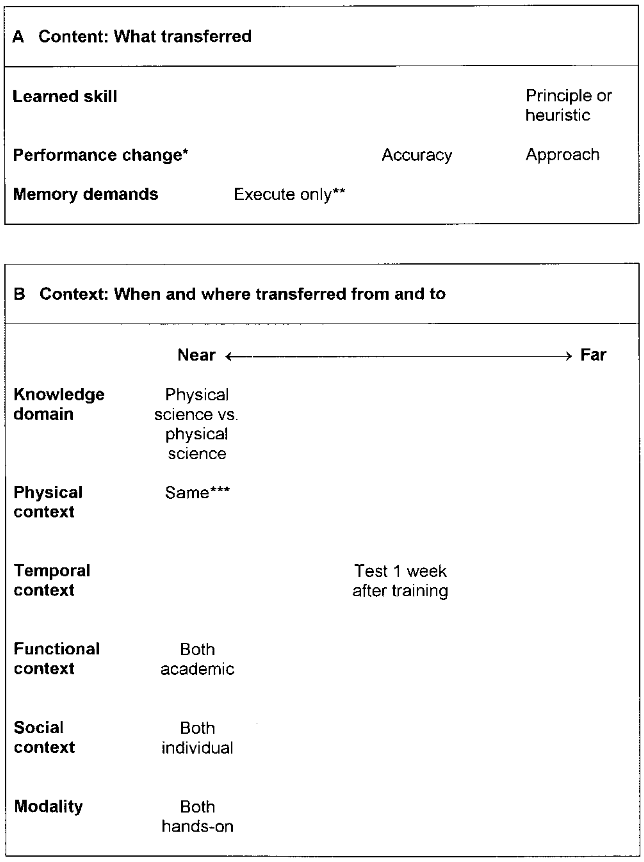

Figure 3. Taxonomy for far transfer applied to the near transfer tests of Chen and Klahr (1999). One asterisk notes that performance change was measured by the number of correct answers, where *correct* means an accurately applied learned principle. Two asterisks note that the authors implied that the relevance was made clear to the participants. Three asterisks note that the authors did not state the location within the school but implied it was the same or very similar.

suring how complete their usage was (see Figure 2). Of interest for the present discussion is the systematic comparison of the consequences of different memory demands (the third dimension of content) in a study in which at least one dimension of context involved reasonably far transfer. The study compared transfer success when (a) participants were required only to apply the learned principle, when (b) they also had to select which of three training examples to apply, and when (c) they were given no hint that the training was relevant to the transfer task and so had to recall the principle, recognize it as being relevant, and apply it.

There was no significant difference in success rates between the first two conditions, but the recall–no-hint condition (c) resulted in substantially less transfer. Even when there was no hint as to the relevance of the training stories (and therefore the participants were relying on free recall), having more than one training story reduced transfer in comparison with the case in which there was only one training story. If participants were retrieving the correct analogue from their complete portfolio of prior learning, the presence of these two distractors should not have had much effect because of the presence of many more distractors in long-term memory. However, in this case a look at the context suggests a reason—the transfer context was very near to the training context (on all dimensions except domain). Even though no explicit hint was provided, participants must have assumed that the answer lay in the training examples, and thus participants exposed to more than one example were misled by the examples that did not apply.

By systematically varying one dimension while holding all others constant (almost—the comparison above is not totally clean, as it combined two slightly different experiments), the Gick and Holyoak (1980) study is a good demonstration of how memory demands affect transfer. Unfortunately for those wishing to generalize from this research to other situations, all dimensions of transfer context were very near, except the domain of the problems. (Domain was quite distant, being from a military context to a medical context.) Therefore, it sheds little light on questions concerning what would happen if a similar training program were to be conducted in one context, such as a school lesson, and the transfer test were to be conducted in a distant context, such as at home during the summer vacation. For example, would having

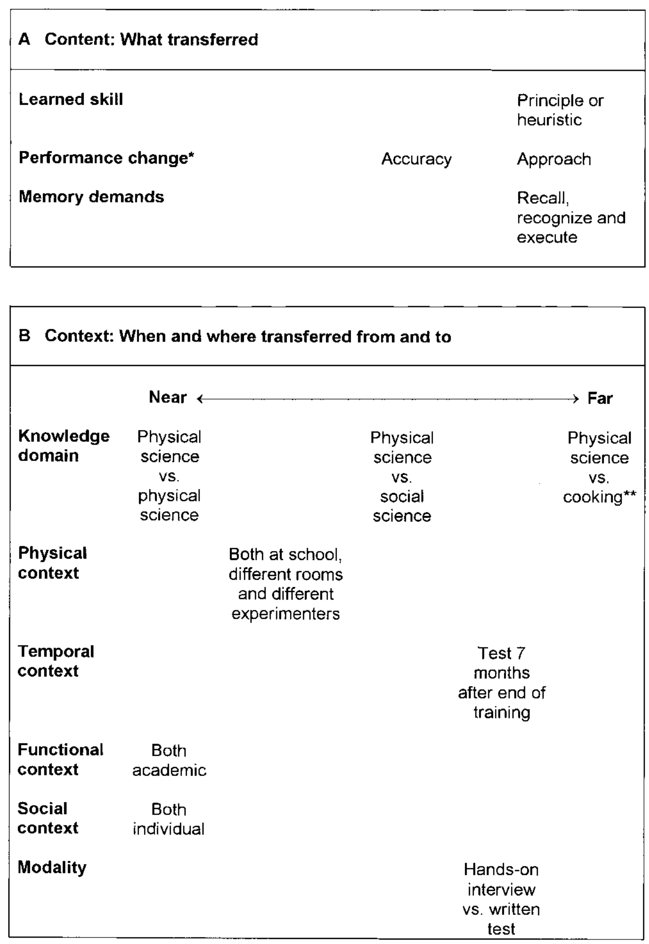

Figure 4. Taxonomy for far transfer applied to the remote transfer test of Chen and Klahr (1999). One asterisk notes that performance change was measured by the number of correct answers, where *correct* means an accurately applied learned principle. Two asterisks indicate the assumption that the task did not focus on the scientific principles behind cooking.

three training examples versus one matter if the training were more remote from the transfer test? (The reasoning above suggests not.) Would a hint make such a big difference to success if it concerned a training session conducted months earlier? Studies that manipulate this dimension under conditions of far transfer on other aspects of context are needed if such questions are to be answered.

The study by Chen and Klahr (1999; see Figures 3 and 4) is similar to the Gick and Holyoak (1980) study in terms of aspects of content, but it also included a number of different transfer contexts, a rarity in the research we reviewed. For their near transfer test (Figure 3), all dimensions except temporal context were very near, and even the temporal context change was only a 1-week delay. In the remote transfer case (Figure 4), domain was varied from near (physical science vs. physical science) to far (physical science vs. cooking), and both the temporal context and the modality were quite far apart, with a 7-month gap from training to transfer and a written transfer task following on from a hands-on training program. Chen and Klahr also made an effort to render physical context slightly different from training with the use of a different experimenter and a different room at the school. Other aspects of context were quite close.

It is unfortunate, for our present purpose, that Chen and Klahr (1999) did not report direct comparisons of these different contexts (near vs. far and different domains within far). Because their focus was on comparing training techniques and studying the microdevelopment of learning, they merely reported that the far transfer test showed some transfer; they did not directly compare it with near transfer (though they did report an age effect, that older children transferred further). Nor did they report the effect of the variations in domain distance that were explored within the far transfer test. Notwithstanding this, the study does constitute one of the few demonstrations we could find of successful transfer that is far along many dimensions of context. With the benefit of this taxonomy, perhaps future researchers will also be encouraged to extend this work and report comparison data such as those suggested above.

Comparison of Contexts of a Subset of Studies

We next attempt to make a more direct comparison between the contexts of a subset of studies by plotting them on a single figure, in terms of some of these taxonomic dimensions. Figure 5 shows the complexity when all six dimensions of the transfer context are combined. As this yields 64 cells, we separated the subsequent figures, showing three dimensions on each figure. Figure 6 shows the dimensions of knowledge domain, physical context, and temporal context. These are arguably the most important of the six context dimensions for educational applications of transfer research because schools would hope that the reasoning skills they teach their students would ideally transfer to different domains of knowledge, to physical contexts outside the school, and to temporal contexts several years later. (However, the functional dimension might also be considered critical for educational as well as for noneducational purposes, such as transfer of military training. For example, one would hope that hypothetical school-based learning would transfer to practical decisions at work and that military training would transfer to the very different functional context of a real combat situation, for which the goals may be very different.) As can be seen, in the subset of well-known transfer studies plotted in Figure 6, each investigated a unique combination of the first three dimensions of transfer context—the simple characterization “near” versus “far” does not capture the richness of the patterns of context investigated, indicating the need for this more detailed and more structured approach.

Given that all these studies investigated different contexts, differences in results could be attributed to differences in context, if all else were equal. Unfortunately, with a group of studies such as these that were not designed as matched comparisons, all else rarely is equal. The studies may also vary in other aspects of context or content as well as training regimen and difficulty of skills taught, not to mention aspects of the participant groups themselves. For example, although all the studies found some successful transfer, the Fong et al. (1986) study found mixed results, with successful transfer only occurring for some questions and not others. Perhaps, one might suggest on the basis of Figure 6, this is due to the difficulty of transferring to a remote physical context, as opposed to the design of the studies by Gick and Holyoak (1980) and Chen and Klahr (1999), in which, although the domain was also far, the physical context was near. However, some of the studies also varied along other dimensions of context—functional context, social context, and modality—as shown in Figure 7, in which the same group of studies is plotted against the remaining three dimensions. This figure demonstrates one of the problems in interpreting differences found in comparing studies from the previous figure. A look at Figure 7 shows that the Fong et al. study also involved a distant functional context, whereas the Gick and Holyoak and Chen and Klahr studies involved a near functional context, rendering interpretation difficult, even before other factors are considered.

Unfortunately, with so many potentially relevant dimensions that have been overlooked by many researchers, pairs of studies that happen to constitute clean, unconfounded comparisons are rare. However, with this taxonomy in mind, future researchers should be able to design the studies required to rigorously explore the effects of these different dimensions on transfer, especially if within-study comparisons are made, thus controlling for effects of training and task difficulty.

Although the existing research does not permit unconfounded comparisons between studies (to reliably draw conclusions about the effect of the various dimensions of the taxonomy), it does enable us to identify gaps between what has been investigated and the content and contexts that need to be investigated so that the desired generalizations about the achievability of far transfer can be made. Hence, one aspect of Figure 6 is worth particular note—the rightmost cell is empty. We simply did not find any well-controlled studies testing transfer to a far domain, in a far physical context, and in a far temporal context. This is unfortunate, given that this cell is of particular relevance to the educational applications of transfer research. Again, we hope that this taxonomy might provoke future researchers to fill this gap.

Even among the cells for which studies were available, finding a critical mass of studies in any given cell to permit robust conclusions was also difficult. One cell that has received considerable attention from researchers, and is far on at least one of these three dimensions, is that representing experiments involving transfer to a distant domain but with near physical and near temporal contexts. Figure 8 shows this sample on the same three dimensions as in Figure 6. (However, note that, as shown in Figure 9, these

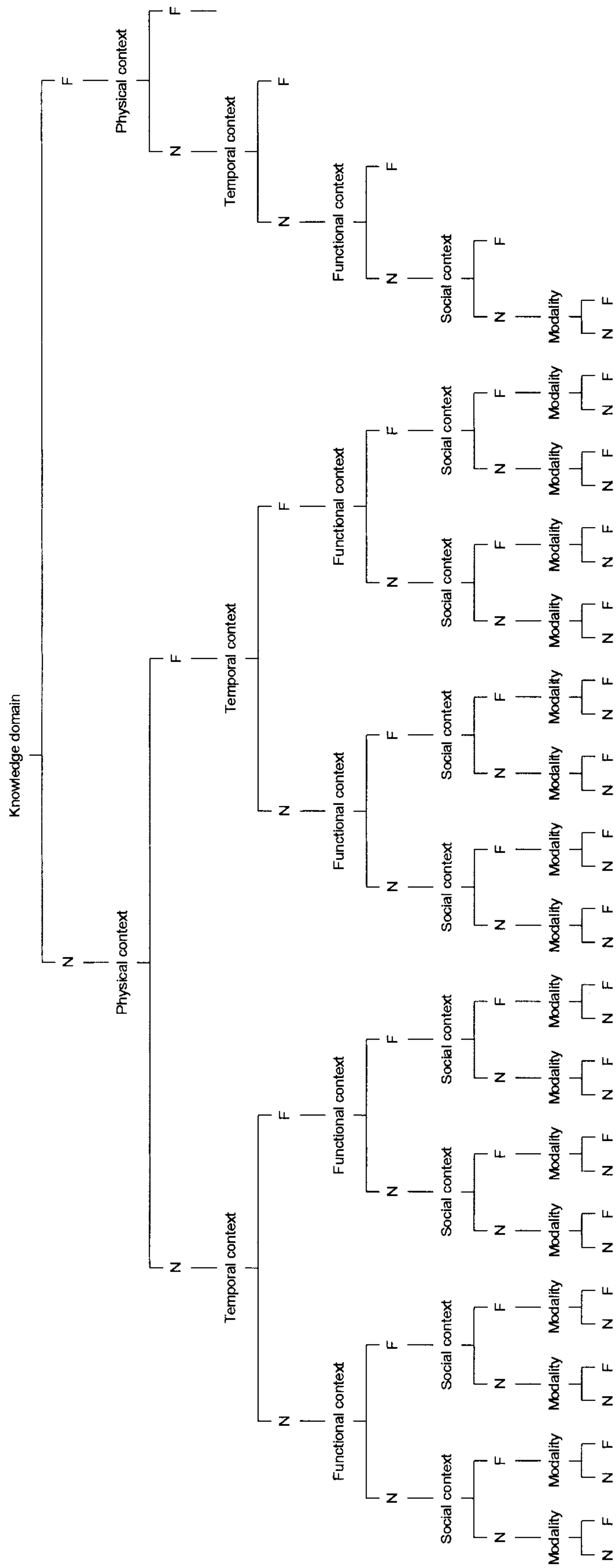

Figure 5. Dimensions of context. N = near; F = far.

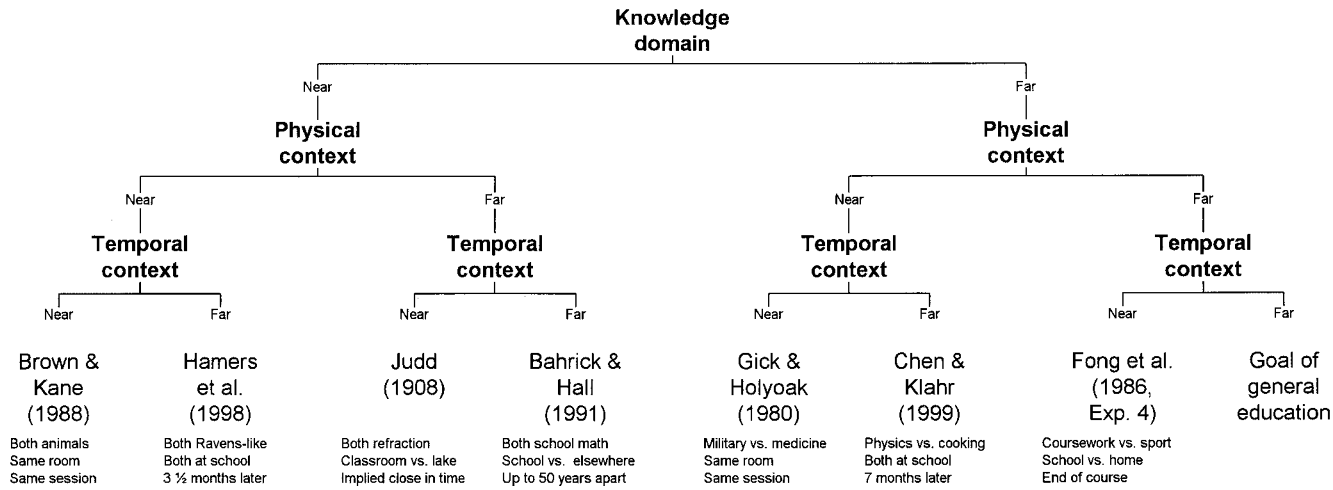

Figure 6. The knowledge domain, physical, and temporal dimensions of context and examples of transfer studies. Exp. = experiment.

studies are not all fully comparable; some differ, at least in part, on several of the other dimensions of context.) The fact that most of these studies found transfer under some circumstances suggests that one might tentatively conclude that transfer to a far domain (with near physical and temporal contexts) can be achieved with some regularity.

However, although this is an encouraging step, can we go on to conclude from this that the educational enterprise described earlier in this review is justified? Probably not, as such studies represent just one of the transfer challenges facing schools. School-taught skills must also transfer to contexts that are far on physical and temporal dimensions and in many cases on some of the other dimensions as well. Thus, such studies are informative but clearly do not yet provide a definitive answer to the question posed at the beginning of this article and at the beginning of the last century, namely whether a general education yields reasoning skills that transfer to the students' future worlds of work and family. For this, the field requires studies that are aimed at far transfer on at least all

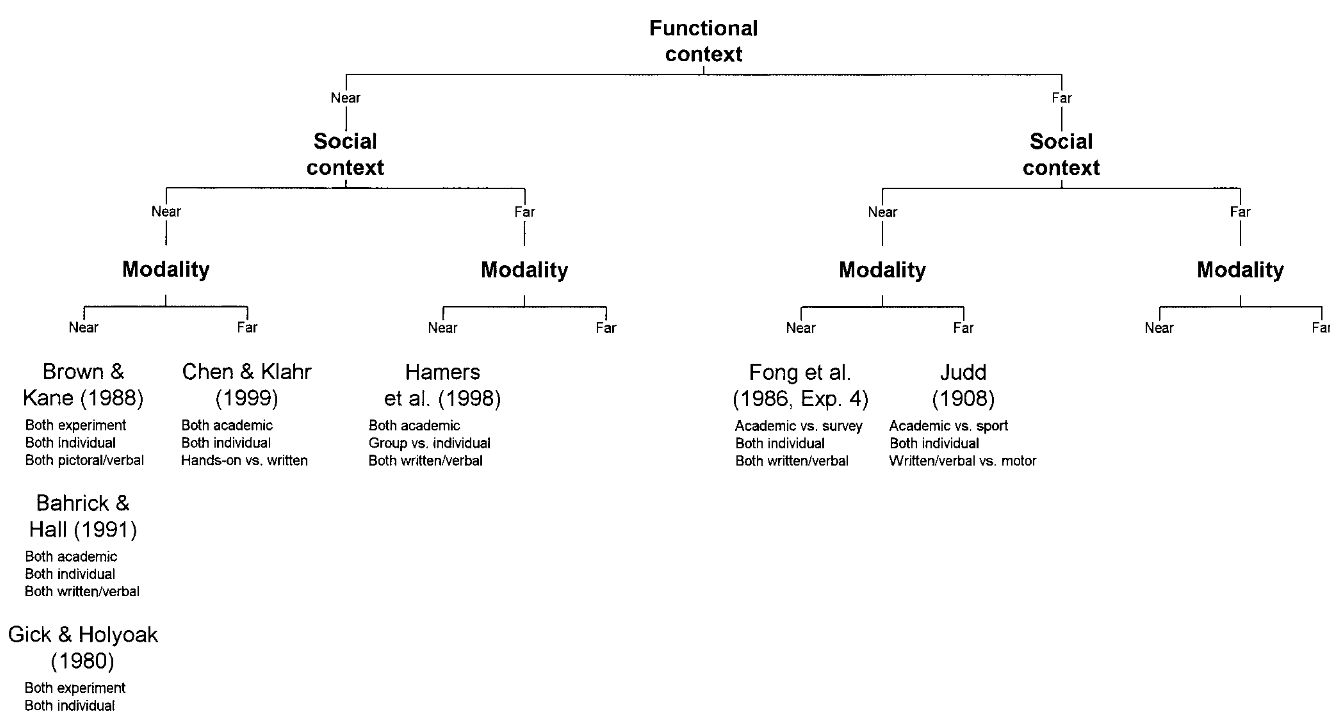

Figure 7. The functional, social, and modality dimensions of context and examples of transfer studies. Exp. = experiment.

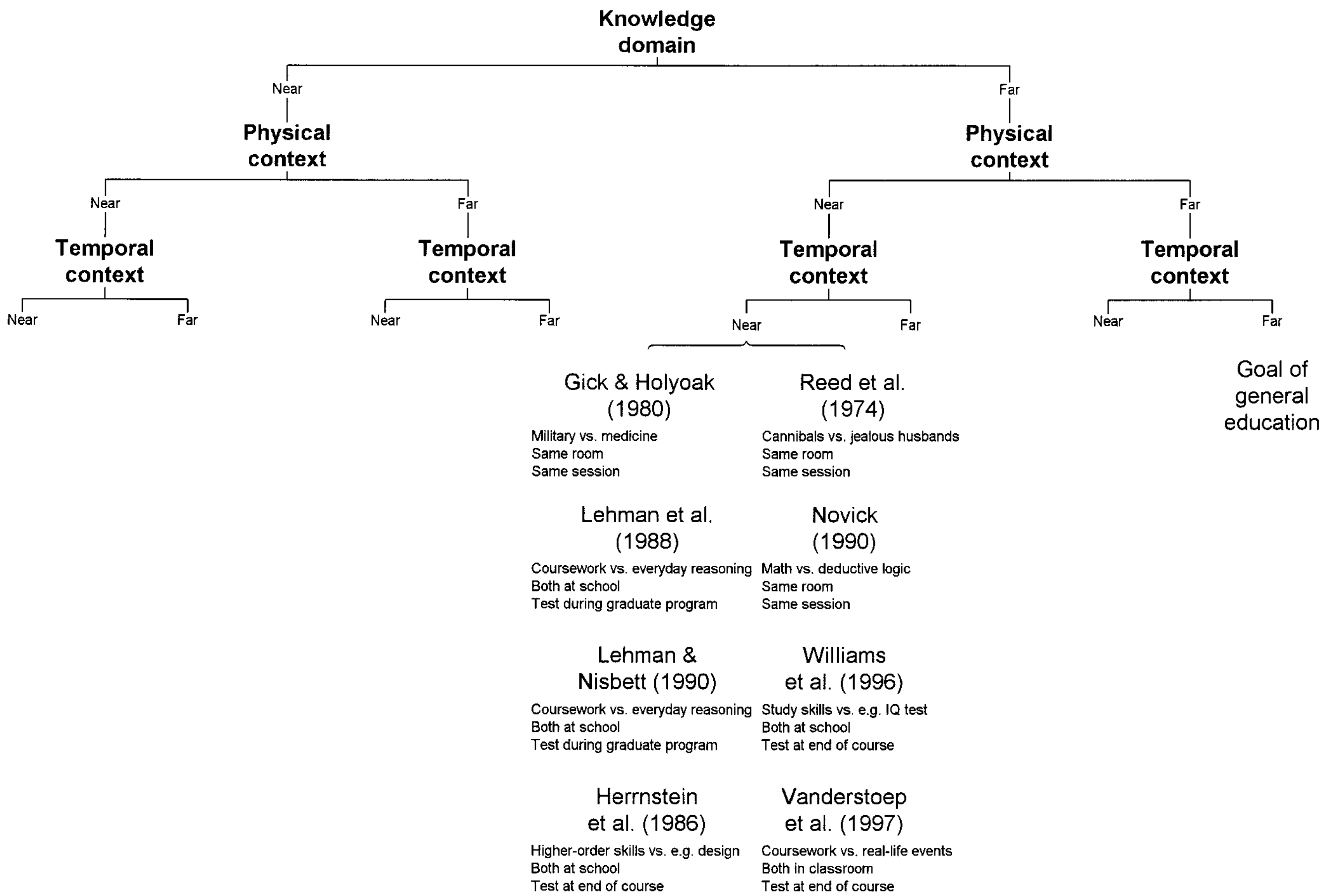

Figure 8. The knowledge domain, physical, and temporal dimensions of context with studies involving transfer to a distant domain, but with near physical and near temporal contexts.

three of the dimensions in Figure 8. In fact, most justifications of the educational enterprise would also require transfer to be far on the functional dimension, something arguably missing from all the studies in Figures 8 and 9, except perhaps the Novick (1990) study. Schools presumably hope their students would be able to apply their learning when they are not in academic mode.

As mentioned above, we did not find any evidence of successful transfer that was far on all three of the first three dimensions. However, as can be seen, limited evidence does presently exist for transfer that is far on two of these dimensions, although the studies portrayed here are virtually all that we could find among the well-known or seminal works. This evidence includes three situations in which transfer was successful across two far-context dimensions. First, it demonstrates that schools, or at least universities, may be able to claim to teach lasting domain-specific skills that transfer beyond the institution's physical context (far physical context, far temporal context). Evidence for this comes from Bahrick and Hall's (1991) study of adult memory of high school mathematics learning some 30 years later, which showed successful transfer for those participants who had gone on to study college-level math. Second, there is evidence for lasting transfer of knowledge to a new domain, albeit in a similar physical context. Evidence for this comes from Chen and Klahr's (1999) demonstration of transfer to a far domain of knowledge that lasted 7 months (far knowledge domain,[11] far temporal context). Third, Fong et al.'s (1986, Experiment 4) research offered evidence that skills can transfer to a far domain in a far physical context, albeit in a near temporal context, but only for some of the questions asked.

Thus, if one were to conduct a follow-up study of any of three sorts, either a test of Bahrick and Hall's (1991) participants in a different domain, a long-term follow-up of Fong et al.'s (1986) participants, or a comparison test of Chen and Klahr's (1999) students in a different physical context, it would represent the first demonstration of far transfer on these three key educationally relevant dimensions—provided it was a successful demonstration. Then, the further complication of functional context could be added to the mix (already included in Fong et al.) for a more complete test of the benefits of education for later life applications (assuming transfer across social and modality boundaries is not an issue).

[11] This is assuming that the significant result found in the remote transfer test was not driven solely by the near domain questions included in the test but by the far domain questions as well.

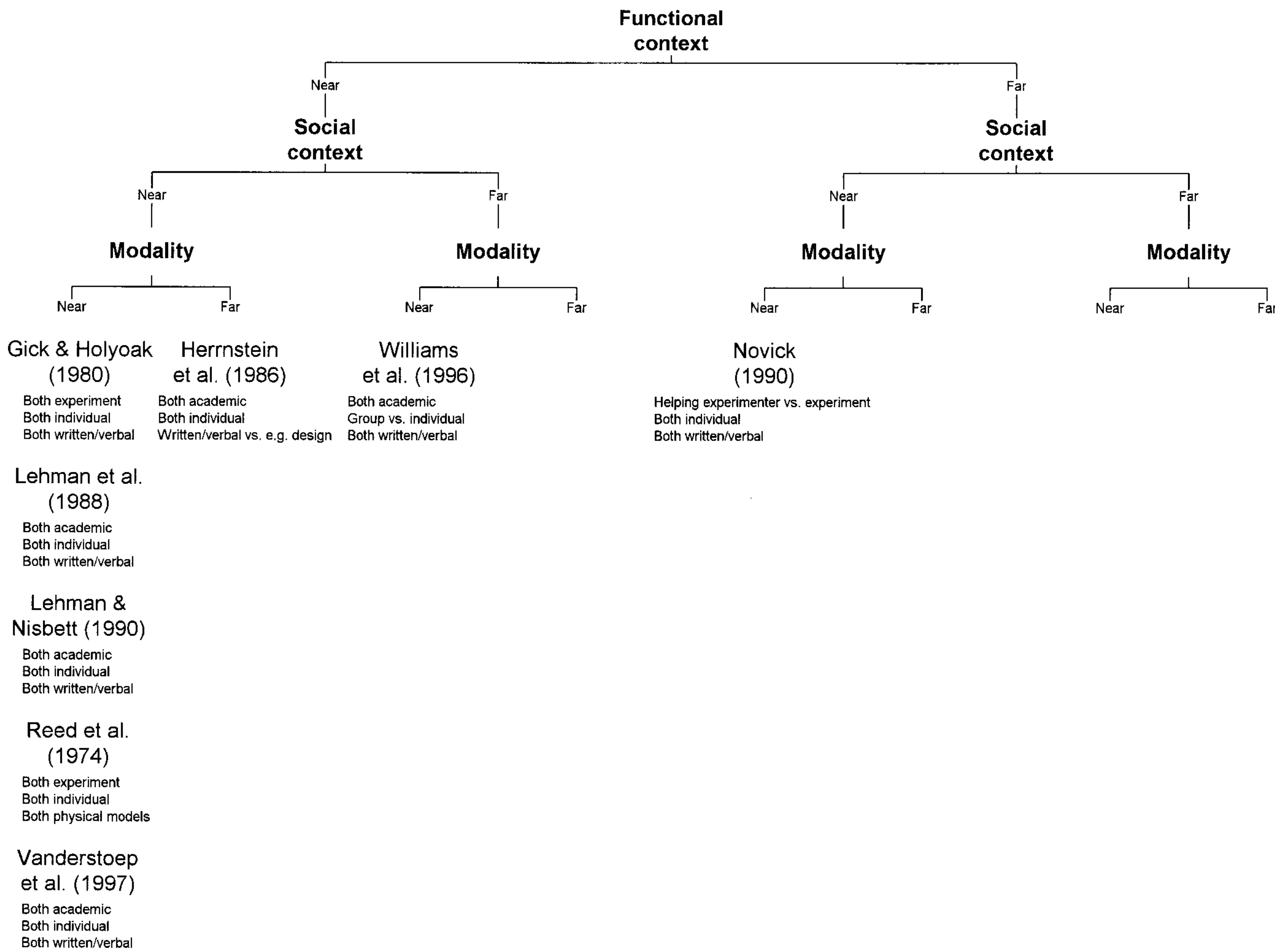

Figure 9. The functional, social, and modality dimensions of context with studies involving transfer to a distant domain, but with near physical and near temporal contexts.

Note that studies potentially could be designed that would fail to show transfer, even in each of these cells in which at least one example of successful transfer has been found. However, assuming the knowledge to be transferred had been adequately learned in the first place, it should be possible to differentiate those studies from these, either in terms of different aspects of context, in terms of aspects of content, or in terms of individual characteristics such as domain knowledge. If it can be shown that transfer can occur to a particular cell, optimal training parameters can then be systematically explored to understand how best to ensure the most effective transfer. These can then be compared with the characteristics of existing and proposed educational programs.

Finally, we do not doubt that the foregoing analysis is too simplistic, as it reflects only a sample of studies and, even for these, the analysis requires dichotomizing the continua of each of the dimensions and ignores other differences between the details of the procedures and materials used. Notwithstanding this caveat, however, this analysis amply demonstrates the value of such a taxonomic approach to systematically evaluating disparate research findings and claims as well as provides an illustration of the complexity of the issues with which a more rigorous analysis would have to cope. Making sense of the body of research is a challenge when differences between studies are often confounded and many of the important cells are empty and, as yet, unexplored.

Implications

At the beginning of this article we suggested that there were both practical and theoretical reasons why the question of whether and when transfer occurs is important to resolve, a view also expressed by Singley and Anderson in their 1989 book on transfer. We now revisit each of these topics.

Practical Issues

As we have shown, none of the studies to date have tested for transfer that would be classified as far along *all* of the dimensions described, and indeed none tested transfer that was far on even a majority of the dimensions. Whether a study addresses a form of transfer that is “far enough” depends on the nature of the generalizations to which the research is to be applied and the kinds of socially valued performance they involve. Justifications for the value of general education may require transfer that is far on many dimensions, whereas specialized graduate-level academic training might be justified by successful transfer to a relatively near domain, as more graduate students are presumably intending to apply their skills in the same discipline. Indeed, Dunbar (1995, 1997) found that what he called *near transfer* (e.g., within the field of biology, from clam genes to malaria genes) yielded many fruitful insights in research group meetings in university microbiology labs. In this situation, *far transfer* (from outside biology) was merely used to explain but not to create insights. (On an interesting note, Halpern et al., 1990, showed that the near–far nature of the analogy interacted with the type of dependent measure used. When the dependent measure was the spontaneous use of an analogy to solve a problem, near domain analogies were preferable; however, when the measure involved comprehending text, far analogies were better.)

Regarding the content of the skills to be transferred, it may not be defensible to generalize from research involving transfer of specific facts and procedures to applications involving transfer of general principles, or vice versa. This is because there is some suggestion in the literature reviewed here (e.g., Brown, 1989; Schliemann & Nunes, 1990) that general heuristics and principles may transfer more readily than more specific learning. Another reason this may be particularly problematic is because the nature of the memory demands during transfer (e.g., spontaneous recall vs. prompted execution) may differentially influence transfer success rates, thus interacting with type of learned skill. That is, in a transfer situation in which both skills are adequately learned, recall of specific facts and procedures may require prompting, whereas general skills may be spontaneously recalled. This is because the cues that would retrieve the specific facts and procedures could be more situation specific.

Regarding the context of transfer, the individual dimensions of context need to be explored separately and in combination to fully assess the generalizability of research findings. More studies that systematically explore certain dimensions of transfer context while holding others constant, similar to those by Gick and Holyoak (1980), Lehman and Nisbett and colleagues (Fong et al., 1986; Lehman et al., 1988; Lehman & Nisbett, 1990), Herrnstein et al. (1986), and Chen and Klahr (1999), would be useful in disentangling the many dimensions highlighted by this taxonomy.

Furthermore, as investments in education and training are made for many different purposes, the particular purposes need to be taken into account when designing research to test the success of these investments. If the goal is to encourage transfer from school-based lessons to nonacademic situations in the workplace years later, then something akin to this context must be explored in transfer research if it is to be applicable to the goal in question. As we have seen, most transfer studies conducted to date do not do this; therefore, the extent to which schools and universities deliver this sort of transferable benefit is, as yet, largely unknown.

This highlights one of the particular difficulties in conducting meaningful transfer research; if context matters, then the context in which the research is conducted may affect results. Thus, generalizing from abstracted, lab-based experimental work is problematic, unless the context to be generalized to is that of abstracted, lab-based tasks. In other areas of psychological research, it may be possible to reduce the phenomenon under investigation to its essentials to permit perfectly controlled, lab-based experiments. However, when the context in which the experiment is conducted becomes one of the variables under investigation, as here, its potential effects cannot be ignored. Thus, ecological validity of experiments becomes a more troublesome question. Given the difficulties in conducting longitudinal research, compromises will obviously have to be made on some dimensions. However, more systematic exploration of some of the other dimensions highlighted by this taxonomy may be practical and, we believe, worthwhile if we are to better understand the nature and contextual determinants of transfer and, hence, its applicability.

After reviewing attempts to teach transferable thinking skills, Resnick (1987) concluded, “Clearly, a most important challenge facing the movement for increasing higher order skill learning in the schools is the development of appropriate evaluation strategies” (pp. 33–34). A necessary first step toward rigorously evaluating such programs is an organized taxonomy of the dimensions of transfer. We view the taxonomy provided in this article as a first step toward imposing order on a chaotic literature. Future researchers will undoubtedly be able to refine this taxonomy. Until they do, however, it is premature to declare that transfer to novel situations is as ubiquitous as educational policymakers appear to presume—or as rare as some scholars have asserted: “Novel insights as cases of transfer are probably rarer than volcanic eruptions and large earthquakes” (Detterman, 1993, p. 2). Instead, transfer is clearly highly contextualized; as we explore the intersections of the dimensions in our taxonomy it ought to be possible to predict when, where, and how far transfer occurs.

Theoretical Issues

We have argued that transfer is multidetermined; although various forms of transfer occur, including so-called far transfer (at least by some definitions), success depends on certain aspects of the situation, including the content to be transferred and the context to which it is transferred. This claim of content and context sensitivity accords with recent evidence reviewed elsewhere suggesting that aspects of the physical, social, and semantic contexts exert significant sources of variation on cognitive performance, including but not limited to transfer (see Ceci, 1996, for a summary of evidence). Such evidence for the context sensitivity of cognition has led to the widespread use of rubrics such as situated cognition (Lave & Wenger, 1991; Rogoff, 1990), ecologically dependent nature of cognition (Ceci, 1996), and situated generalization (D. Carraher & Schliemann, 1998). Thus, the taxonomy presented here is in some sense a hypothesis about factors determining not only transfer but also cognition more broadly.

However, even if successful transfer depends on contextual dimensions, this does not rule out the existence of general cognitive skills that underpin the act of transfer, a point noted by Larkin and her colleagues (e.g., Mayer, Larkin, & Kadane, 1984) in the context of applying mathematical learning to solving new problems. Encoding, representing, retrieving, mapping, and extending knowledge are general processes that inhere in a very broad range of cognitive performances, not just in transfer tasks. Thus, the application of general cognitive skills may be involved in transfer, but their successful application may be moderated by myriad contextual factors. This suggests that individual differences in the ability to successfully transfer emanate from two potential sources, a participant's familiarity with the relevant contextual factors (e.g., the domain in question) and a participant's underlying cognitive skill involved in encoding, representing, retrieving, mapping, and transferring prior learning. Therefore, transfer may be influenced by individual differences in general cognitive capacities and other aspects of general intelligence that operate independently of the influence of context, as well as by aspects of the context and content. Theorists commonly assume that transfer and the skills that go into it are central to human intelligence (see Holyoak, 1984, p. 200; Sternberg, 1985, for complete exposition). Indeed, Goswami (1991), in a review of research and theories concerning the development of analogical reasoning in children, hypothesized that one of the key engines of developmental change in analogical reasoning ability may lie in the development of meta-cognitive skills that promote deep, structural encoding.

Implications of Transfer Success

For any given cell of the taxonomy there are three possible outcomes for a given group of participants: (a) A skill successfully learned in one context does not transfer to a new context, (b) a learned skill does transfer but only if participants have ample domain knowledge, and (c) a learned skill transfers irrespective of the participant's level of domain knowledge.

If any cells of the taxonomy are characterized by Outcome (a), the implication is that theories of learning transfer and expertise must be at least somewhat context dependent. Alternately, if any cells of the taxonomy are characterized by either Outcome (b) or (c), then such theories must include skills or rules that are broader than, or go beyond, the original context of learning. This is because, although the process of transfer may perhaps be helped by the presence of existing background knowledge in the new domain, the training that produced the transfer effect did not occur in that domain.

Finally, it is conceivable, though perhaps unlikely, that all cells could be found to be characterized by the same outcome. If they are all like (a), then theories of learning, transfer, and expertise may only require context-dependent skills and rules, whereas if all cells of the taxonomy are like (c), theories may only need to postulate context-independent abilities. The latter is akin to the strong problem-solving methods formulated by Newell and Simon (1972). If all cells are characterized by (b), then theories must be based on context-independent skills and rules, which include placeholders for domain-specific knowledge, so that their application requires some interaction with domain.

Note that, because context can be defined with greater or lesser specificity, the same breadth of application of learning can be deemed either a cross-context transfer or a context-specific application of learning, depending on whether a narrow or a broad definition of context is used. Consequently, the skill assumed to underlie the learning can be labeled *context dependent* or *context independent*, again depending on the breadth of the definition of *context*. Thus, these theoretical labels are only meaningful if used in conjunction with a description of contextual categories built on a taxonomy such as this. The taxonomy thus becomes a part of the theory.

Implications of Group Differences

Furthermore, the outcome in each cell could differ as a function of other characteristics of the participant group besides possible group differences in domain knowledge—for example, high versus low IQ, novelty seeking versus conservative, intrinsically versus extrinsically motivated. Thus, theories of learning and transfer could be different for different participant groups. For instance, it is conceivable that high-IQ individuals might use context-independent skills and rules but low IQ-individuals rely on context-specific knowledge and rules. As the same skills emphasized by many programs designed to train transfer are at the core of what is termed *general intelligence* in some theories (Ceci, 1996; Sternberg, 1982, 1985), we might expect variations in general intelligence to be intimately linked to success at transfer. The research seems to support this view.

Studies dating back to the first half of the 20th century showed that high-ability participants transferred faster and more accurately than lower ability ones. For example, Kuenne (1946) showed that transfer occurred as a function of mental age in a group of 40-to-80-month-olds. These findings were consistent with the belief that transfer depended on a set of underlying cognitive skills that were pervasive in the sense that they facilitated not only transfer performance but also performance on a wide range of tasks, including those tapped by IQ tests.

These ideas reflect aspects of theories of intelligence and transfer that date back a century. For example, Spearman (1904) suggested that analogical reasoning tasks depend on inductive reasoning and draw heavily on both the education of relations and correlates. Thus, transfer from one form of reasoning to an analogous one can be thought of as a manifestation of underlying general intelligence. This is not to claim that the content, context, and breadth of the transfer domain are irrelevant, but simply to claim that successful transfer may be differentiated by individual differences in underlying eductive capacities and other aspects of general intelligence.

Other theories of intelligence place greater emphasis on the nature of the domain involved in the education of relations (Ceci, 1996; Sternberg, 1982, 1985). Some intelligent systems are organized in such a manner that their dimensions are integrated and easily connected in transfer tasks (see Halford's, 1999, description of how neural nets are differentiated into larger and larger numbers of dimensions that can be instantiated to reason about complex domains). According to such views, transfer of a reasoning process occurs when the organism possesses the appropriate eductive skills and has an appropriately organized knowledge base on which to apply these skills (Outcome [b], above). For example, a child may transfer general metaphorical reasoning skills taught in one domain ("The hungry reader devoured the book") but fail to do so in another that involves similar complexity ("The thirsty car drank the petrol") as a result of lack of integrated knowledge (Keil, 1981, 1984).

Thus, individual differences in transfer according to these theories could be the result of differences in the organization of knowledge rather than differences in meta-cognition or basic eductive processes. Regardless of which view one adopts, the research conducted during the past century demonstrates enormous individual differences that are most likely independent of differences that we have attributed to systematic variations along the content and context dimensions described herein. A complete theory of transfer must acknowledge such sources of variation, even if their elucidation is beyond the aim of the theory.

Concluding Comments

To end at the beginning, after a century of intense research activity on the topic of transfer, scholars are perhaps in no greater agreement than they were at its inception. There are still those who claim that transfer is exceedingly rare (e.g., Detterman, 1993; Schooler, 1989), those who claim that it is much more prevalent (e.g., Halpern, 1998), and those who openly express the view that the situation is unresolveable and that no consensus is likely in the near future. Consider the following opinions:

> The issue of the transferability of thinking and learning skills . . . is still open. (Resnick, 1987, p. 19)

> The picture is still rather unclear. What was presumed to be a basic and ubiquitous process of learning has been illusive. . . . This eventually leads us to question whether transfer is at all salvageable as an explanation. (D. Carraher & Schliemann, 1998, p. 6)

In this review, we have tried to make the case that transfer is indeed a salvageable concept and that instances of far transfer, although not frequent, are documentable and may even be predictable once the relevant dimensions are specified. The taxonomy presented here describes the dimensions against which transfer of a learned skill may be assessed. To be sure, future research is needed to flesh out many of the empty cells in this taxonomy, and certainly the sort of mathematical optimization needed to assure that the framework is actually tapping the underlying taxon network will entail lots of psychometric work on the measurement properties of each dimension, the localization of optimal cut points on each dimension, "an estimate of the validities those cuts achieve" (Meehl, 1995, p. 274), and so on. Such taxonomic formalism requires not only the use of heretofore unavailable psychometrics but also the realization that the ceteris paribus clause in the philosophy of science (i.e., all other things being equal) is seldom applicable in the transfer domain because other things such as background knowledge are rarely equal. For this reason, we have used the term *framework* to describe our taxonomy rather than theory, and we admit freely to the lack of sharp edges that our framework generates in making quantitative predictions.

As we noted at the outset of this review, our concern has been principally with the end points of the transfer process—does the organism go from new learning to its transfer and, if so, when? Only once the end points of a transition are established—from the state of the organism at the time of the new learning to the end state when it can be transferred to new situations—can the transition process itself be understood. Consequently, the proposed taxonomy was designed to address the end points of transfer rather than the components of the transition process itself. The latter requires a consideration of numerous issues about the representation and deployment of knowledge: whether exemplar-specific details are discarded from the representation once an underlying schema is induced (Reeves & Weisberg, 1994); whether so-called reception tasks that provide the information rather than require its production by participants implicitly cue the encoding of superficial features (Blanchette & Dunbar, 2000); under what conditions and for what dependent measures a far analogy that shares only underlying structural features will be superior to a near analogy that shares both deep and surface features (Halpern et al., 1990); and how participants represent problems, what details are stored, and how the relations between surface elements are comprehended (Reeves & Weisberg, 1994).

To date, accounts of the transition process itself have been most successful when they have modeled the underlying processes of very near transfer tasks, for example, Singley and Anderson's (1989) study of the abstract production rules involved in transfer from using a line editor to using a full screen editor. However, this work makes no contact with the far transfer dimensions discussed in this review. Likewise, other equally elegant and abstract analyses of the components underlying the transfer process have been conducted under fairly stripped-down versions of the real world, with little consideration for whether transfer across far dimensions even occurs. Extending this work to the far-transfer dimensions we discussed awaits further research. It is obvious that a complete theory of transfer will require accounts of both the end points of transfer as well as the underlying components involved in the transition. We hope we have provided a step toward the former in this review.

Once the underlying transfer process and its end points are established, researchers would have a descriptively adequate taxonomy that is capable of providing practical consequences (when transfer is likely to occur) but also informing theory-driven models of learning, intelligence, and performance. For example, in the bioecological theory of intelligence (Ceci, 1996), transfer across domains is an integral aspect of intellectual growth, but the precise manner in which this occurs is unknown and only hinted at in the theory, for example, by speculating that it is through such elementary processes as "noticing," "detecting similarities," and "mapping these detections on to other domains" that metaphorical reasoning becomes generalized, leading to intellectual change.

When we began this line of research we had one goal in mind, namely, the development of a framework that made sense of the conflicting claims about transfer. To the best of our knowledge the corpus of scientific studies from the prior century can be fitted into this framework. We leave to readers to decide if educational programs inculcate generalizable skills or more constrained forms of transfer and if the assumptions that appear to be implicitly made by educational policy makers are supported by the research, past and future.

References

Adey, P., & Shayer, M. (1993). An exploration of long-term far-transfer effects following an extended intervention program in the high school science curriculum. *Cognition and Instruction, 11,* 1–29.

Bahrick, H. P., & Hall, L. K. (1991). Lifetime maintenance of high school mathematics content. *Journal of Experimental Psychology: General, 120,* 20–33.

Barnett, S. M., & Koslowski, B. (in press). Adaptive expertise: Effects of type of experience and the level of theoretical understanding it generates. *Thinking and Reasoning.*

Bassok, M., & Holyoak, K. J. (1989). Interdomain transfer between isomorphic topics in algebra and physics. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 153–166.

Benbow, C. P., & Lubinski, D. (Eds.). (1996). *Intellectual talent: Psychometric and social issues.* Baltimore: Johns Hopkins University Press.

Bielaczyc, K., Pirolli, P. L., & Brown, A. L. (1995). Training in self-explanation and self-regulation strategies: Investigating the effects of knowledge acquisition activities on problem solving. *Cognition and Instruction, 13,* 221–252.

Blanchette, I., & Dunbar, K. (2000). How analogies are generated: The roles of structural and superficial similarity. *Memory and Cognition, 28,* 108–124.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). *How people learn: Brain, mind, experience, and school.* Washington, DC: National Academy Press.

Brown, A. L. (1989). Analogical learning and transfer: What develops? In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 369–412). New York: Cambridge University Press.

Brown, A. L., & Campione, J. C. (1990). Communities of learning and thinking, or a context by any other name. In D. Kuhn (Ed.), *Developmental perspectives on teaching and learning thinking skills: Vol. 21. Contributions to human development* (pp. 108–126). Basel, Switzerland: Karger.

Brown, A. L., & Kane, M. J. (1988). Preschool children can learn to transfer: Learning to learn and learning from example. *Cognitive Psychology, 20,* 493–523.

Brown, A. L., Kane, M. J., & Echols, C. H. (1986). Young children's mental models determine analogical transfer across problems with a common goal structure. *Cognitive Development, 1,* 103–121.

Brown, A. L., Kane, M. J., & Long, C. (1989). Analogical transfer in young children: Analogies as tools for communication and exposition. *Applied Cognitive Psychology, 3,* 275–293.

Campione, J. C., Brown, A. L., & Bryant, N. (1985). Individual differences in learning and memory. In R. J. Sternberg (Ed.), *Human abilities* (pp. 103–126). San Francisco: Freeman.

Carraher, D., & Schliemann, A. D. (1998, April). *The transfer dilemma.* Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Carraher, T. N., Carraher, D., & Schliemann, A. D. (1985). Mathematics in the streets and in the schools. *British Journal of Developmental Psychology, 3,* 21–29.

Catrambone, R., & Holyoak, K. J. (1989). Overcoming contextual limitations on problem-solving transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 1147–1156.

Ceci, S. J. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology, 27,* 703–722.

Ceci, S. J. (1996). *On intelligence: A bioecological treatise on intellectual development.* Cambridge, MA: Harvard University Press.

Ceci, S. J., & Bronfenbrenner, U. (1985). "Don't forget to take the cupcakes out of the oven": Prospective memory, strategic time-monitoring, and context. *Child Development, 56,* 152–164.

Ceci, S. J., & Ruiz, A. (1993). The role of context in everyday cognition. In M. Rabinowitz (Ed.), *Applied cognition* (pp. 164–183). Hillsdale, NJ: Erlbaum.

Ceci, S. J., & Williams, W. M. (1997). Schooling, intelligence, and income. *American Psychologist, 52,* 1051–1058.

Chen, Z. (1996). Children's analogical problem solving: The effects of superficial, structural, and procedural similarity. *Journal of Experimental Child Psychology, 62,* 410–431.

Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development, 70,* 1098–1120.

Chi, M. T. H., Glaser, R., & Farr, M. J. (1988). *The nature of expertise.* Hillsdale, NJ: Erlbaum.

Cummins Dellarosa, D. (1992). Role of analogical reasoning in the induction of problem categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 1103–1124.

Detterman, D. K. (1993). The case for the prosecution: Transfer as an epiphenomenon. In D. K. Detterman & R. J. Sternberg (Eds.), *Transfer on trial: Intelligence, cognition, and instruction* (pp. 1–24). Norwood, NJ: Ablex.

Druckman, D., & Bjork, R. A. (1994). *Learning, remembering, believing: Enhancing human performance.* Washington, DC: National Academy Press.

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

Dunbar, K. (1997). How scientists think: On-line creativity and conceptual change in science. In T. B. Ward, S. M. Smith, & J. Vaid (Eds.), *Creative thought: An investigation of conceptual structures and processes* (pp. 461–493). Washington, DC: American Psychological Association.

Duncker, K. (1945). On problem solving. *Psychological Monographs, 58*(5, Serial No. 113).

Ericsson, K. A., & Smith, J. (1991). *Toward a general theory of expertise: Prospects and limits.* New York: Cambridge University Press.

Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). The effects of statistical training on thinking about everyday problems. *Cognitive Psychology, 18,* 253–292.

Forbus, K. D., Gentner, D., & Law, K. (1995). MAC/FAC: A model of similarity-based retrieval. *Cognitive Science, 19,* 141–205.

Gabbert, B., Johnson, D. W., & Johnson, R. T. (1986). Cooperative learning, group-to-individual transfer, process gain, and the acquisition of cognitive reasoning strategies. *Journal of Psychology, 120,* 265–278.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology, 12,* 306–355.

Goswami, U. (1991). Analogical reasoning: What develops? A review of research and theory. *Child Development, 62,* 1–22.

Gould, S. J. (1981). *The mismeasure of man.* New York: Norton.

Halford, G. (1999). The development of intelligence includes capacity to process relations of greater complexity. In M. Anderson (Ed.), *The development of intelligence* (pp. 193–213). Hove, England: Psychology Press.

Halpern, D. F. (1998). Teaching critical thinking for transfer across domains. *American Psychologist, 53,* 449–455.

Halpern, D. F., Hansen, C., & Riefer, D. (1990). Analogies as an aid to understanding and memory. *Journal of Educational Psychology, 82,* 298–305.

Hamers, J. H. M., de Koning, E., & Sijtsma, K. (1998). Inductive reasoning in third grade: Intervention promises and constraints. *Contemporary Educational Psychology, 23,* 132–148.

Hatano, G., & Inagaki, K. (1984). Two courses of expertise. *Research and Clinical Center for Child Development, 83* [Annual report], 27–36.

Hendrickson, G., & Schroeder, W. H. (1941). Transfer of training in learning to hit a submerged target. *Journal of Educational Psychology, 32,* 205–213.

Herrnstein, R. J., Nickerson, R. S., de Sanchez, M., & Swets, J. A. (1986). Teaching thinking skills. *American Psychologist, 41,* 1279–1289.

Holyoak, K. J. (1984). Analogical thinking and human intelligence. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 2, pp. 199–223). Hillsdale, NJ: Erlbaum.

Humphreys, L. G. (1951). Transfer of training in general education. *Journal of General Education, 5,* 210–216.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence.* New York: Basic Books.

Judd, C. H. (1908). The relation of special training to general intelligence. *Educational Review, 36,* 28–42.

Keil, F. (1981). Constraints on knowledge and cognitive development. *Psychological Review, 88,* 197–227.

Keil, F. (1984). Mechanisms in cognitive development and the structure of knowledge. In R. J. Sternberg (Ed.), *Mechanisms in cognitive development* (pp. 81–100). New York: Freeman.

Klausmeier, H. J. (1961). *Educational psychology. Learning and human abilities.* New York: Harper.
Koslowski, B., Barnett, S. M., Thompson, S., Masnick, A., Rosenblum, T., Swiderick, M. (2000). *Formal models.* Unpublished manuscript.
Kosonen, P., & Winne, P. H. (1995). Effects of teaching statistical laws on reasoning about everyday problems. *Journal of Educational Psychology, 87,* 33–46.
Kuenne, M. R. (1946). Experimental investigation of the relation of language to transposition behavior in young children. *Journal of Experimental Psychology, 36,* 471–490.
Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation.* New York: Cambridge University Press.
Lehman, D. R., Lempert, R. O., & Nisbett, R. E. (1988). The effects of graduate training on reasoning: Formal discipline and thinking about everyday-life events. *American Psychologist, 43,* 431–442.
Lehman, D. R., & Nisbett, R. E. (1990). A longitudinal study of the effects of undergraduate training on reasoning. *Developmental Psychology, 26,* 952–960.
Luchins, A. S. (1942). Mechanization in problem solving. *Psychological Monographs, 54*(6, Serial No. 95).
Mayer, R. E., Larkin, J. H., & Kadane, J. (1984). A cognitive analysis of mathematic problem-solving ability. In R. J. Sternberg (Ed.) *Advances in the psychology of human intelligence* (Vol. 2, pp. 231–273). Hillsdale, NJ: Erlbaum.
Meehl, P. E. (1995). Bootstrap taxonometrics: Solving the classification problem in psychopathology. *American Psychologist, 50,* 266–275.
Meudell, P. R., Hitch, G. J., & Boyle, M. M. (1995). Collaboration in recall: Do pairs of people cross-cue each other to produce new memories? *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 1,* 141–152.
Meudell, P. R., Hitch, G. J., & Kirby, P. (1992). Are two heads better than one? Experimental investigations of the social facilitation of memory. *Applied Cognitive Psychology, 6,* 525–543.
Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92,* 289–316.
Needham, D. R., & Begg, I. M. (1991). Problem-oriented training promotes spontaneous analogical transfer: Memory-oriented training promotes memory for training. *Memory and Cognition, 19,* 543–557.
Newell, A., & Simon, H. A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice Hall.
Nisbett, R. E., Fong, G. T., Lehman, D. R., & Cheng, P. W. (1987, October 30). Teaching reasoning. *Science, 238,* 625–631.
Novick, L. R. (1990). Representational transfer in problem solving. *Psychological Science, 1,* 128–132.
Perkins, D. N., & Grotzer, T. A. (1997). Teaching intelligence. *American Psychologist, 52,* 1125–1133.
Piaget, J. (1952). *The origins of intelligence in children.* New York: International Universities Press.
Piaget, J., & Inhelder, B. (1969). *The psychology of the child.* New York: Basic Books.
Raven, J. L. (1958). *Standard progressive matrices.* London: Lewis.
Reder, L. M., & Klatzky, R. L. (1994). Transfer: Training for performance. In D. Druckman & R. A. Bjork (Eds.), *Learning, remembering, believing: Enhancing human performance* (pp. 25–56). Washington, DC: National Academy Press.
Reed, S. K., Ernst, G. W., & Banerji, R. (1974). The role of analogy in transfer between similar problem states. *Cognitive Psychology, 6,* 436–450.
Reed, S. K., & Saavedra, N. C. (1986). A comparison of computation, discovery, and graph procedures for improving students' conception of average speed. *Cognition and Instruction, 3,* 31–62.
Reeves, L., & Weisberg, R. W. (1994). The role of content and abstract information in analogical transfer. *Psychological Bulletin, 115,* 381–400.
Resnick, L. B. (1987). *Education and learning to think.* Washington, DC: National Academy Press.
Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context.* Oxford, England: Oxford University Press.
Ross, B. H. (1987). This is like that: The use of earlier problems and the separation of similarity effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 629–639.
Rovee-Collier, C. (1993). The capacity for long-term memory in infancy. *Current Directions in Psychological Science, 2,* 130–135.
Schliemann, A. D., & Magalhães, V. P. (1990). Proportional reasoning: From shopping to kitchens, laboratories, and, hopefully, schools. In G. Booker, P. Cobb, & T. Mendicuti (Eds.), *Proceedings of the XIV PME Conference* (Vol. 3, pp. 67–73). Oaxtepec, Mexico: International Group for the Psychology of Mathematics Education.
Schliemann, A. D., & Nunes, T. (1990). A situated schema of proportionality. *British Journal of Developmental Psychology, 8,* 259–268.
Schooler, C. (1989). Social structural effects and experimental situations. In K. W. Schaie & C. Schooler (Eds.), *Social structure and aging: Psychological processes* (pp. 1–21). Hillsdale, NJ: Erlbaum.
Sewell, W. H., Hauser, R. M., & Featherman, D. L. (1976). *Schooling and achievement in American society.* New York: Academic Press.
Singley, K., & Anderson, J. R. (1989). *The transfer of cognitive skill.* Cambridge, MA: Harvard University Press.
Slavin, R. E. (1983). When does cooperative learning increase student achievement? *Psychological Bulletin, 94,* 429–445.
Slavin, R. E. (1984). Students motivating students to excel: Cooperative incentives, cooperative tasks, and student achievement. *Elementary School Journal, 85,* 53–63.
Spearman, C. (1904). General intelligence objectively determined and measured. *American Journal of Psychology, 15,* 206–221.
Spencer, R. M., & Weisberg, R. W. (1986). Context-dependent effects on analogical transfer. *Memory and Cognition, 14,* 442–449.
Starch, D. (1910). A demonstration of the trial and error method of learning. *Psychological Bulletin, 7,* 20–23.
Sternberg, R. J. (1982). Reasoning, problem-solving, and intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 225–307). New York: Cambridge University Press.
Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence.* New York: Cambridge University Press.
Sternberg, R. J. (1988). Applying cognitive theory to the testing and teaching of intelligence. *Applied Cognitive Psychology, 2,* 231–255.
Sternberg, R. J. (1989). Domain-generality versus domain-specificity: The life and impending death of a false dichotomy. *Merrill Palmer Quarterly, 35,* 115–130.
Sternberg, R. J., Ferrari, M., Clinkenbeard, P. M., & Grigorenko, E. (1996). Identification, instruction, and assessment of gifted children: A construction validation study of a triarchic model. *Gifted Child Quarterly, 40,* 129–137.
Thorndike, E. L. (1906). *Principles of teaching.* New York: Seiler.
Thorndike, E. L., & Woodworth, R. S. (1901a). The influence of improvement in one mental function upon the efficiency of other functions: (I). *Psychological Review, 8,* 247–261.
Thorndike, E. L., & Woodworth, R. S. (1901b). The influence of improvement in one mental function upon the efficiency of other functions: II. The estimation of magnitudes. *Psychological Review, 8,* 384–395.
Thorndike, E. L., & Woodworth, R. S. (1901c). The influence of improvement in one mental function upon the efficiency of other functions: III. Functions involving attention, observation and discrimination. *Psychological Review, 8,* 553–564.
Tudge, J., & Winterhoff, P. (1993). Can young children benefit from collaborative problem solving? Tracing the effects of partner competence and feedback. *Social Development, 2,* 242–259.
Tudge, J. R. H., Winterhoff, P. A., & Hogan, D. M. (1996). The cognitive

consequences of collaborative problem solving with and without feedback. *Child Development, 67,* 2892–2909.

Tulving, E. E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80,* 352–373.

Tversky, A., & Kahneman, D. (1974, September 27). Judgements under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131.

VanderStoep, S. W., & Shaughnessy, J. J. (1997). Taking a course in research methods improves reasoning about real-life events. *Teaching of Psychology, 24,* 122–124.

Van de Vivjer, F J. R., & Hutschemaekers, G. J. M. (1990). *The investigation of culture.* Tilburg, the Netherlands: Tilburg University Press.

Walberg, H. J. (1982). A psychological theory of educational productivity. In F. H. Farley & N. J. Gordon (Eds.), *Psychology and education: The state of the union* (pp. 81–108). Berkeley, CA: McCutchan.

Wiley, F. E. (1976). Another hour, another day: Quantity of schooling, a potent path for policy. In W. H. Sewell, R. M. Hauser, & D. L. Featherman (Eds.), *Schooling and achievement in American society* (pp. 225–266). New York: Academic Press.

Williams, W. M., Blythe, T., White, N., Li, J., Sternberg, R. J., & Gardner, H. (1996). *Practical intelligence for school.* New York: Harper Collins.

Winne, P. H. (1995). Inherent details in self-regulated learning. *Educational Psychologist, 30,* 173–187.

Woodworth, R. S., & Schlosberg, H. (1954). *Experimental psychology.* New York: Holt, Rinehart & Winston.

Received January 2, 2001
Revision received October 17, 2001
Accepted October 23, 2001 ■