Sameness and Difference in Transfer

Ference Marton
Department of Education and Educational Research
Göteborg University, Sweden

Discussions about transfer have mainly dealt with how people manage to do something in a situation thanks to having done something similar in a previous situation. From an educational point of view, however, it appears more fruitful to consider the case when the learner, having learned to do something in 1 situation, might be able to do something different in other situations, thanks to perceived differences (and similarities) between situations. The case is made for widening the focus of attention to how situations are related through differences (and similarities).

The debate on transfer has occupied educational psychologists for more than a century. After more than a decade of reduced interest, the topic seems to be attracting more attention again. The National Science Foundation has supported two symposia on transfer within little more than 2 years. Mestre (2005) reported on the first one, and this article originates from the second (described by Lobato, 2004). Transfer is about how what is learned in one situation affects or influences what the learner is capable of doing in another situation. As this is basically the main question of all education, we can probably agree that it should not be neglected for long.

Lave’s (1988) foundational criticism of “the culture of transfer experiments” (p. 34) seems a suitable point of departure. She refers to Thorndike’s (1913) and Judd’s (1908) work as the dual historical origins of the research on transfer.

THE CULTURE OF TRANSFER EXPERIMENTS

Thorndike (1913)

According to Thorndike (1913), the human mind makes “… particular reactions to particular situations” (p. 249), hence learning something specific cannot be ex-
Thorndike studied the effect of learning one thing on dealing with something else that is apparently very similar. He showed that transfer between two apparently similar tasks can be negligible and drew the conclusion that transfer between two tasks that are more different (e.g., knowing Latin and being smart in general) cannot be larger than transfer between two tasks that are more similar to each other (as in the experiments he carried out).

An example of such an experiment involved training learners to judge the size of rectangles that had areas between 10 cm² and 100 cm² with feedback after each trial (Situation A). Both before and after the training, learners were asked to judge the size of rectangles with areas of 140 cm² to 200 cm² (Situation B₁) and 200 cm² to 300 cm² (Situation B₂). The aim was to find out whether there was transfer (spread over) from the training of judging the size of the rectangles with areas of 10 cm² to 100 cm². The results showed that the improvement in judging the size of rectangles different from those used in training was moderate and decreased with increasing difference (no improvement in judging the size of rectangles with areas of 200 cm² to 300 cm²). Incidentally, these results were from 4 participants only and varied widely across participants (Thorndike & Woodworth, 1901).

Thorndike’s (1913) interpretation was that learning Task A influences the learner’s capability of performing Task B to the extent that there are identical elements between the two tasks: “By identical elements are meant mental processes which have the same cell action in the brain as their physical correlate” (p. 359). For example, learning to recognize black color and learning to recognize a square facilitates the learning of recognizing a black square. Similarly, there is transfer from training in addition to carrying out multiplication, because the former is a part of the latter.

Judd (1908)

As is evident from the previous paragraph’s quote from Thorndike’s early work, he talked about identical features of stimuli and identical features of responses in two different situations; the notion of “identical elements” referring to the former (“mental processes” in the previous quote denote “effective stimuli”). Judd (1908) went a step further and assumed that the relationship between two situations, A and B, was not only a function of the similarities and differences between A and B as seen by the researcher but was also a function of the way in which A was dealt with by the learners. In a well-known experiment, a group of children practiced throwing darts at an underwater target located 12 in under the water surface (Situation A). Another group did exactly the same, but they received an explanation of the discrepancy between the target’s apparent location in the water and its actual location, in terms of general principles of the refraction of light. The difference between the two treatments was supposed to induce differences in how Situation A was dealt with by the two groups of learners, which in turn was hypothesized to in-
fluence how they dealt with Situation B—the task of trying to hit a target located 4 in under the water surface with darts. The two groups performed equally well in Situation A but the principle-of-refraction group clearly outperformed the other group in Situation B. Judd’s explanation was that the principle-of-refraction group learned to handle Situation A by means of using a principle that was also applicable to Situation B, whereas the other group learned to adjust to the specific features of Situation A (which were not generalizable to Situation B).

If Thorndike (1913; Thorndike & Woodworth, 1901) represented the behaviorist paradigm—according to which learning is conceived as the constitution of bonds between stimuli (features of the environment) and responses (reactions of the learner)—Judd (1908) was in a sense a forerunner of the cognitivist paradigm, according to which learning involves the constitution of more and more powerful representations of the world around us. Many more recent studies of transfer resemble Judd’s study, insofar as they focus on the effect of dealing with one instantiation of a general principle on the capability of dealing with another instantiation of the same principle (“same” as perceived by the experimenter).

Gick and Holyoak (1980)

Lave examined four studies (Gentner & Gentner, 1983; Gick & Holyoak, 1980; Hayes & Simon, 1977; Reed, Ernst, & Banerji, 1974) carried out within the cognitivist paradigm. From a cognitivist perspective, learning to solve one problem may enhance the solving of another problem depending on similarities between how the two problems are mentally represented.

One of these studies was Gick and Holyoak’s (1980) experiment, in which they investigated to what extent and under what conditions learners make use of analogies to connect different but structurally identical problems (as perceived by the researcher). For example, they utilized Duncker’s (1945) well-known radiation problem: How can we destroy a tumor in the stomach by radiation without damaging healthy tissue? The target solution is the use of rays converging on the tumor and thus producing maximal intensity of radiation there, whereas the radiation would be less intense and therefore hopefully less harmful to the tissue surrounding the tumor. If the learners first deal with another problem that can be solved in a structurally identical way (namely by means of simultaneously converging paths)—and deal with it successfully—would their experience of solving the first problem in that way enhance the likelihood of their solving the second problem in the same way? The similar problem utilized in the study involved how to capture a fortress with mined surroundings. By dividing the attacking force into small groups and sending them along different paths that converge on the fortress, the load on the surrounding terrain would not be sufficient to activate the mines, but the concentration of military force would be sufficient to seize the fortress.
All of the studies scrutinized by Lave (1988) raised basically the same question and got basically the same answer: Spontaneous transfer is negligible. Solving the first problem might boost the likelihood of solving the second problem in the same way, but only as long as the similarity between the two problems, the shared principle or structure, is explicitly pointed out to the learners.

LAVE’S CRITIQUE OF THE CULTURE OF TRANSFER EXPERIMENTS

According to Lave (1988), transfer research builds on the metaphor of knowledge as a set of tools stored in the memory of the learner for use in different situations. Sometimes there are different specific tools for different specific tasks (as in the case of Thorndike, 1913), sometimes there are a few multi-purpose tools, each one to be used for dealing with a large number of tasks (as in the case of Judd, 1908).

The different approaches discussed belong to what Lave (1988) called the functionalist view, according to which the prototypical mechanism of knowing is developing explicit knowledge that can be transferred and applied to a number of situations. According to the alternative practice view, the prototypical mechanism of knowing is people participating in specific practices and developing knowledge specific to those practices in the very settings where they are located. The community of practice decides what counts as knowledge and what is excluded. Lave questioned the functionalist view in favor of the practice view.

Let us now take a closer look at four of the questions raised by Lave (1988) in her critique of the culture of transfer experiments. I have chosen these four questions from the point of view of the logic of my argument and not because they necessarily represent what Lave herself would have considered the most important of the questions she raised.

What Is Learned?

The researcher’s view of transfer is a function of his or her views of learning: It is, of course, what people learn that they transfer, and how the question “What is learned?” is answered constitutes the fundamental difference between different theories of learning. Whereas early behaviorists (like Thorndike) saw learning as a forming (occasionally weakening) of stimulus-response associations, cognitivists (like Gick & Holyoak) put the emphasis on cognitive structures and mental representations, the acquisition of which enable people to do things that they have not been able to do previously. But the acquisition of specific skills and knowledge does not take place in a vacuum, Lave argued. Whatever people do, they do it in some situation. The situations are parts of certain practices that people participate in. The most important
thing that people learn is to participate in those practices. The specific skills and knowledge are byproducts or component parts of the participation.

Lave’s (1988) critique of the culture of transfer had much to do with the question “What is learned?.” She challenged the view that people typically learn to master decontextualized tasks and that what they learn is transferred to other decontextualized tasks. To the extent that people manage to establish relationships (of similarity) between different practices or different settings in which they participate, they can make use of their experience of one practice when they participate in another practice or another setting. This is probably as close as Lave got to making use of the concept of transfer.

Who Defines the Relations Between Situations?

In transfer experiments the researcher defines the tasks and the expected outcomes, and transfer is concluded when what the learners do matches the researcher’s expectations (Lave, 1988, p. 37). But there might be (and there certainly are) relations between situations created by the learners themselves that might be quite different from the specific relation defined by the experimenter. Lave’s alternative to the culture of transfer experiments was to “focus attention on questions about how people establish relations of similarity between the problems they encounter in different settings” (p. 44).

How Many Situations Are Involved?

Lave (1988) pointed to the “… crucial differences between two-problem transfer in an experiment and uses of knowledge in the varied arenas of the lived-in world” (p. 40) and to “… the importance of repeated occasions (rather than unfamiliar situations) as effective sites of shaping knowledge and its uses” (p. 44). My interpretation of Lave’s argument is: If there are any relations constituted between two situations, it is highly unlikely that there would be no relations constituted to other situations. Hence, the relation between any two situations is contingent upon the relation between each of the two situations and other situations that are relevant for what is to be learned as seen from the learner’s point of view. This does not imply that the relations between all situations that the learners have ever experienced have to be taken into consideration. It implies, however, that rather than focusing on relations between two isolated situations, we should focus on relations between sets of situations that have certain relevant aspects in common.

Where Does Transfer Happen?

As mentioned previously, Lave (1988) argued that knowledge is locally constituted or situated. In the history of transfer there has always been one situation (Sit-
uation A) in which the learner learns something and another situation (Situation B) in which what was learned in the first situation is applied—more or less unproblematically. This means that the knowledge used in the second situation exists prior to its application. According to Lave, however, the knowledge applied in Situation B does not exist prior to its application but is constituted when it is applied, even if the learner always draws on previous experiences as resources. From an enactivist perspective, knowing and acting are one and the same. Knowledge is not something independent of action (Maturana & Varela, 1980).

BEYOND LAVE’S CRITIQUE

Several novel formulations of the problem of transfer have been presented since Lave published her critique in 1988 (see, for instance, Beach, 1999; Bransford & Schwartz, 1999; Carraher, Nemirovsky, & Schliemann, 1995; Detterman & Sternberg, 1993; Greeno, Smith, & Moore, 1993; Lobato, 2003; Mestre, 2005; Packer, 2001; Perkins & Salomon, 1992; Tuomi-Gröhn & Engeström, 2003). I discuss two of them here in some detail. Again, the choice is made from the point of view of the logic of the argument inherent in this article. The authors of the two studies discussed below are in agreement with Lave’s view of transfer as I tried to capture it in terms of the four questions. However, the view of transfer inherent in the line of reasoning in this article is in agreement with these two studies except on a single point. In this article, I make the case for expanding the conceptualization of transfer to include also the important process of discerning differences.

Greeno et al. (1993)

Greeno et al. (1993) developed a situated view of transfer: “The question of transfer … is to understand how learning to participate in an activity in one situation can influence (positively or negatively) one’s ability to participate in another activity in a different situation” (p. 100). So, how can this happen? The Gibsonian concept of affordance is central in Greeno et al.’s reasoning. It refers to features of the environment that make particular activities possible. For example, chairs (but also stools) afford sitting, small stones (but also apples) afford throwing, and so forth. Chairs afford sitting for those who can sit on them, and small stones afford throwing for those who have hands to grasp them and arms to throw with. Affordance is thus a nondualistic concept—it presupposes both an actor and an acted upon: Certain things afford certain activities for certain organisms, but other objects do not. Affordances thus have to be attended to and perceived as affordances, and the organism has to be attuned to them: “For an activity learned in one situation to transfer to another situation … the
second situation has to afford that activity and the agent has to perceive the affordance” (p. 102).

One of the examples Greeno et al. (1993) cited to illustrate their view of transfer was Judd’s (1908) experiment, mentioned previously, or rather a replication of it by Hendrickson and Schroeder who studied the effects of three conditions (instructionA + practice; instructionB + practice; and only practice) on transfer from one situation to another. In the first situation the children participating in the experiment tried to hit a target immersed in water with an air gun. In the second situation the task was the same, but the depth of the water was just one third of the depth of the water in the first case. According to Greeno et al., the perception of invariant affordances in the two cases was necessary for transfer, and for this to happen children had to learn to adjust to the refraction of light in water. The immersed target’s apparent location on the bottom of the container differed from its actual location: It seemed farther away than it was. There were two ways (at least) in which the shooter could adjust for the apparent displacement of the target. First, he might focus on the angle of the shooting (how much steeper he should angle the gun). The angular adjustment was invariant over the two situations (because the ratio between the apparent horizontal displacement of the target on the bottom of the container and the apparent vertical displacement of the bottom of the container was invariant). So if the child had learned to pick up the angular affordance of the first situation, he would have been able to handle the second situation, thanks to what he had learned in the first situation.

Also in the second case, the children had to learn to make a readjustment in their shooting in the first situation to the refraction of light. They discovered that if they aimed at the target where it seemed to be, they would hit too far away. Hence, they had to aim at a point closer to themselves. So here the adjustment was made in terms of what they aimed at, rather than in terms of how they angled the gun. But this affordance was not invariant across the two situations because deeper water meant greater apparent horizontal displacement (of the target lying on the bottom of the container). From such a point of view, these were two different situations that were related to each other through a constraint (namely deeper water ⇒ greater distance). The original adjustment of what was aimed at had to be adjusted or transformed. The invariance that was necessary for transfer (according to Greeno et al.) was achieved through the transformation—the learner became attuned to the constraint.

Greeno et al. presented a view of transfer as a relation between two situations, or rather, between the learner’s relation to one situation and the learner’s relation to another situation. Transfer is thus constituted by the person in the situation and the situation for the person. In the example they cited—Hendrickson and Schroeder’s study—there was only one kind of relationship examined, although this relationship could be constituted in different ways, the relationship that is conducive for solving the problem, as seen by the researcher.
Lobato and Siebert (2002)

Lobato (2003, in press) has drawn attention to actor-oriented transfer. She investigated generalizations that learners make across situations. By abandoning the experimenter-defined normative stance, the space of generalization potentially increases from one single “guess-what-is-in-my-mind” researcher expectation to multiple, perhaps infinitely many possibilities.

In Lobato’s (2003) case the learners and not the researcher defined the criteria of transfer. Instead of looking for whether a certain kind of transfer defined by the researcher in advance has taken place, Lobato explored what kind of transfer took place thanks to the similarities that learners have established between two or more situations.

Lobato and Siebert (2002) illustrated this view of transfer by means of a case study of an eighth-grade student participating in a 10-day (3 hr/day) teaching experiment aimed at boosting quantitative reasoning. Interviews were conducted with the student on the first, fifth, and last days of the teaching experiment. The researchers were looking for possible instances of transfer between the events during the teaching experiment and the student’s problem-solving activities during the last interview.

The specific aspect of the chosen student’s reasoning was about measuring the slope of a line as a ratio between height and length (e.g., of a wheelchair ramp). A main difficulty for the student in Interview 1 seemed to be that he could not distinguish between slope and height. This amounted, for instance, to thinking that if one walked upwards on a hill along a straight line, the path would become steeper and steeper the higher one get. For this reason he had great difficulty with a problem about how to make a wheelchair ramp higher while keeping the slope invariant.

During the last interview he seemed to have a sudden insight and solved the problem of how to make a 15-ft long and 2-ft high wheelchair ramp 3-ft high instead of 2-ft high without changing its steepness. Instead of simply applying the steepness formula (height over length), which had been taught in his regular math class prior to the teaching experiment, the student solved the problem, after quite a bit of struggle, in his own way. He found out how much length went with 1 ft of height and thus formed the ratio 15:2 = 7.5:1. Then, in order to increase the height with one unit (1 ft), he had to increase the length with one unit (7.5 ft), too. The length of the ramp thus had to be increased to 22.5 ft. In the traditional sense there was no transfer taking place (he had not used the slope formula). But with Lobato’s (2003) idea as the point of departure, Lobato and Siebert (2002) chose another option for investigating possible occurrence of transfer. They asked the question: Was the student’s sudden insight a spontaneous event without any connection to what had taken place during the teaching experiment, or can we find anything that has possibly foreshadowed and laid ground for the sudden insight? The authors meant that it was in fact reasonable to point to an event during the teaching experiment that could possibly be linked to the sudden insight as an instance of transfer between the two situations.
Lobato and Siebert (2002) had actually found such an event in Session 8, when the student had engaged in an activity called “same speed” in which Simcalc Mathworlds software (1996) was used. There was an animated clown walking across the computer screen at a constant speed, covering 10 cm in 4 s. The task was to let a frog walk at the same speed covering other distances and times of the student’s choice. The case study student developed a line of reasoning that 20 cm in 8 s or 30 cm in 12 s would do because these alternatives could be decomposed into 2 and 3 units of 10 cm/4 s. As the speed was the same as the clown’s in every unit part, it must have been the same for the composite distances as well. This means that whereas in the wheelchair ramp problem steepness was conserved through the 7.5:1 length-to-height ratio and greater length was constructed by adding such unit ratios, in the clown–frog problem speed was conserved through forming the 10:4 distance-to-time ratio and greater distance was constructed by adding such unit ratios.1

Nobody can argue with great certainty that the learner actually made use of what he had learned in the clown–frog problem when he was dealing with the wheelchair problem. Lobato and Siebert (2002) simply argued that it was probable that this was the case. Above all, they postulated that if it had been the case, then it was an example of transfer between situations thanks to the learner’s own construction of similarities between the two situations when there was no transfer in the traditional sense. They wanted to demonstrate also that the relation between the two situations should not be described in terms of structural similarities and superficial differences. In both problems the student had made fundamental conceptual advances in order to deal with the specific details of the problems.

THE DOCTRINE OF SAMENESS

Different views of transfer differ as far as the four questions (“What is learned?,” “Who defines the relations between situations?,” “How many situations are involved?,” and “Where does transfer happen?”) are concerned. Yet what they all have in common is that transfer is about people being able to do similar things in different situations because of similarities between those situations.2 It is exactly this belief in

---

1The student actually did more than this in both cases. The simplification in my presentation is due to space restrictions.

2Beach (1999), however, argued for a view of transfer as consequential transitions (i.e., as the transformation, construction, and reconstruction of knowledge, identities, and skills in new situations). By doing so, he emphasized discontinuities as much as continuities. Transfer in this sense is facilitated by expanding the boundaries of conditions for learning. Beach made a distinction between “examining transfer between tasks” on the one hand and “studying transfer between larger socially organized situations” on the other (p. 111). Obviously, I am doing the former, Beach is doing the latter. Tuomi-Gröhn and Engeström (2003) presented a developmental view of transfer, similar to Beach’s. The difference in “grain size” between their discussion of transfer and mine is even more pronounced, as are their concluding words: “… transfer is about changing communities rather than individuals” (p. 35).
the defining role of sameness that I want to question in this article. There cannot be any transfer without sameness, of course. But—and this is my point—there cannot be any transfer without difference, either. One is just as fundamental as the other, and realizing this fact should considerably improve our understanding of transfer.

I will now make some further comments on two of the studies mentioned previously in order to demonstrate that by including a discussion of discerning differences in the interpretations, we might gain a better understanding of what is happening. As mentioned previously, the reason for discussing these studies (Greeno et al., 1993; Lobato & Siebert, 2002) is that the view of transfer expressed in them is very similar to my own, except concerning the role of similarity and difference in transfer. In these two cases I can thus focus on what is different (between their views and my view). In the section “Six Arguments for Considering Sameness and Difference in Accounts of Transfer Instead of Sameness Alone”, I offer some positive evidence by discussing studies that show how differences contribute to bringing about transfer.

Let us thus return to Greeno et al.’s (1993) reinterpretation of the Hendrickson and Schroeder study, according to which children could draw on the first situation (target immersed in 6 in of water) when dealing with the second situation (target immersed in 2 in of water) in either of two ways. First, they may have become attuned to the invariance of the angular rotations in the two cases (the gun had to be angled in the same way). But in order to perceive this invariance (sameness), they had to perceive two kinds of differences (between apparent horizontal location and actual horizontal location, on the one hand, and between apparent depth and actual depth on the other). So the participants may have perceived the invariance because they perceived the two differences (displacements) and the proportionality of the two displacements. Invariance was thus perceived thanks to the perception of variation (differences).

The other way of bringing transfer about between the two situations was by becoming attuned to the constraint deeper water $\Rightarrow$ bigger displacement. One can say that through the transformation (by becoming attuned to the constraint) the learner can establish the similarity between the two situations. But what does it take to get attuned to this constraint? I can reasonably propose that it amounts to noticing that as one thing varies (e.g., the depth of water) another thing varies (size of displacement) and that the two things vary together. So once one sees the differences (variation), one sees the similarity between the two situations. Furthermore, for the specific solution to the first problem to be a resource to draw upon, the solution has to be distinguished from the specific problem. This can only happen if there are at least two different instances of the same solution. This first occurs when the learner approaches the second problem (unless she has had some previous experience of the same kind of problem).

Let us also have another look at Lobato and Siebert’s (2002) study. They described an instance of possible actor-oriented transfer. The key to transfer is supposed to be the learner’s constitution of similarity between the two tasks. In both cases the student formed a ratio that preserved an attribute (speed and steepness,
respectively), and this assumed construction of similarity (sameness) between the two tasks was seen as an instance of transfer from an actor-oriented transfer perspective. The first ratio was between distance and time and the second between length and height. So what does it take to form a ratio and perceive invariance—or “construct invariance,” as Lobato and Siebert would say—in what it denotes (speed and steepness)? In order to constitute speed as an invariant ratio, the learner has to differentiate between distance and time and see the two vary together. And in order to constitute steepness as an invariant ratio, the learner has to differentiate between height and length and see the two vary together. The struggle to do so was very obvious, especially in the last interview. The experienced invariance thus originates from experienced differences and variation.

The argument that in order to draw upon the earlier problem as a resource for dealing with the second problem the problem solution must somehow be discerned from the specifics of the problem also applies in this case. And this also first happens here when the second problem is dealt with, and it happens because the two problems are different.

TOWARD A MORE INCLUSIVE DEFINITION OF TRANSFER

Let us now briefly return to the point of departure of this article. From different theoretical perspectives, transfer is depicted in different ways. But what is it that is depicted in different ways? Can we distinguish between what is depicted and how it is depicted? In order to do so we need to juxtapose the different views and separate what is common for all views from what is specific for each one. There are differences among different theoretical perspectives regarding how what is learned is characterized, how situations are delimited, what kind of relations are considered, how they are constituted, and, above all, what is seen as similar to what. When the concept of transfer was first introduced, let’s say by Thorndike, transfer and Thorndike’s (1913) particular view of transfer (as a function of identical elements) were identical. Thorndike’s view of what was learned in one situation (stimulus-response associations) and what was common between two situations (identical elements) was challenged subsequently by Judd (1908). So at that point finding a definition that covered both meant relaxing constraints: No characterization of what is learned in one situation and what is common between two situations could be parts of a definition of transfer covering both views. Cognitivists (e.g., Gick & Holyoak, 1980) introduced other views of the nature of what is learned and what is similar between situations. Lave (1988) and researchers after her (e.g., Greeno et al., 1983; Lobato, 2003) questioned even how situations are delimited, how many situations are under consideration, who is defining similarities between situations, and where the learning is taking place.
If we want to make a distinction between “transfer” and all of the different “views of transfer,” we have to consider the specific features of the different views as constraints on the definition and then relax them. If we look at all of the views of transfer discussed in the first part of this article and try to find a definition that covers all of them, it would amount to something like the following: “Transfer refers to the relations between what people learn and can do in different situations thanks to similarities between those situations.” However, the main point of this article is that the kind of relations between situations that we have in mind are constituted as much thanks to (perceived) differences as thanks to (perceived) similarities. What matters is the structure of what is attended to, and this can only change if there are discernible differences as well as recognizable similarities. To the extent the arguments in this article are convincing we should relax the constraint of similarity as well and simply settle for the definition of transfer as “relations between what people learn and can do in different situations.” Within this more inclusive definition of transfer there are a number of competing or complementary ways of characterizing it—one of which is to be found in this article, emphasizing the perceptual aspects of learning and the central role of the discernment of differences in transfer.

Transfer and Learning

How the relation between learning and transfer is conceptualized is, of course, closely related to how transfer and learning are conceptualized. For one thing, it has been argued that transfer is a redundant concept. Smedslund (1953) claimed, for instance, that the question of transfer and the question of “What is learned?” are indistinguishable. He said that if we want to find out what is learned, we must introduce changes between the conditions of learning and the conditions of assessment. For instance, if learners have practiced addition tasks and we give them the same tasks again, we cannot distinguish between the case that they have learned the tasks by rote from the case that they have learned to add. If we change the numbers we can say that they have learned to add, but we cannot distinguish between the case that they have learned to add, on the one hand, and the case that they have learned part–whole relations between numbers on the other. If we introduce missing addend problems, for instance, that distinction can be made. Now, if we want to find out what kind of transfer can take place from the original learning experiences, we must do exactly the same thing: systematically manipulate the differences between the learning situation and the situation of application. The notion of transfer is thus redundant.

On the grounds of a similar argument (saying that there is always some difference between the situation in which learning is taking place and the situation in which it is displayed later), Perkins and Salomon (1992) pointed to the difficulty in distinguishing transfer from learning. I should note, however, that Smedslund’s (1953) line of reasoning rests on the (theoretical) assumption that transfer is en-
tirely a function of what is learned in one situation (and is unproblematically used in another situation). This is actually one of the main issues in Lave’s (1988) critique of transfer research, as should be obvious from what was said before.

If I follow the definition of transfer that I have suggested in this article, then I can simply say that when we study learning we usually study what is happening under what conditions with what effect in one delimited situation or set of situations. When we study transfer we study what is happening under what conditions with what effect in two or more situations or delimited sets of situations.

**PERCEPTUAL ASPECTS OF LEARNING**

Partly by making use of a Gibsonian framework, both Greeno et al. (1993) and Lobato and Siebert (2002) emphasized the perceptual aspects of learning. Such a focus does not contradict, but instead complements, the activity aspect of learning that is more prominent in practice-oriented theories, notably activity theory (cf., Tuomi-Gröhn & Engeström, 2003). Perceptual aspects of learning refer to how people perceive the world in relation to how they act. Marton and Booth (1997) gave an account of differences in how people see the world around them in terms of what aspects are discerned and focused on simultaneously when they observe certain phenomena.

As the very term *situated learning* suggests, the notion of situation is very important in practice-oriented theories of learning. Situation, or context, refers to what surrounds the learning event; that is, to the socially constructed life-world in which a particular instance of learning occurs. Context in this sense is characterized by embeddedness. Take any component part of language, a word for instance. Any word uttered has a context of other words: a sentence in which it is used, a line of argument in which the sentence is a part, and so forth. This context is organized in linear temporality: One thing follows the other. But we can add a spatial dimension: There is a social situation in which someone is saying something to someone with some purpose in mind. This is the kind of context dealt with by practice-oriented theorizing on transfer.

But there is context in another sense as well. As de Saussure (1983) pointed out, every word has the context of every other word that could have been uttered but was not. A word derives its meaning not only from what it refers to, or what precedes and what follows it, but also from what it is an alternative to. The meaning of *brown*, for instance, derives from the fact that it is not *red*, *blue*, or *yellow*. The meaning of a question derives partly from the fact that it is not a statement, an imperative, or an exclamation. Language is a system, and when we are using its component parts the system is present in our awareness and generates meaning. In this sense, meaning derives from the distinctions we are making. This is not only true about language. Goodwin gave some excellent illustrations of *professional seeing*—how systems of distinctions are used for describing certain kind of phenomena such as dirt by archeologists and the reaction of an alleged criminal by police.
Such systems of distinctions define professional seeing; they are what make an anthropologist, or a policeman, a professional (Goodwin, 1994).

I can, however, extend this line of reasoning beyond language and institutionalized forms of seeing the world. We all learn to see the world—at home, in school, and in the workplace. We learn to notice differences and to make distinctions. We see everything against the background of our experience. We see someone as tall because we have seen people of different heights. We experience wine as fruity because we have had wine before that was not fruity. Marton and Tsui (2004) described and illustrated a framework for the study of learning in which learning to see certain things in certain ways is due to the patterns of variation and invariance that we have experienced.

Differences appear on two levels in this article. First, learning is a function of perceived differences (and similarities) in the learning situation. Second, transfer is seen as a function of perceived differences (and similarities) in two or more situations and between those situations.

In the following section of this article, I build upon a particular emphasis on perceptual aspects of learning (in a broad sense) in order to present six arguments for the consideration of both sameness and difference in accounts of transfer rather than focusing on sameness alone. I accomplish this goal by doing the following:

1. Illustrating—through reasoning based on the role of discernment in perceptual learning—why the classical idea of transfer, as something being extracted from one situation and carried over to another situation, is untenable.
2. Using Bransford and Schwartz’s (1999) “preparation for future learning” (p. 92) approach to transfer in order to illustrate the necessity of differences in transfer.
3. Showing how learning one thing makes the learning of another thing possible because these two things are different.
4. Illustrating how one experience modifies another (prospectively or retrospectively) due to differences.
5. Illustrating that preparation for transfer to an unknown situation requires variation.
6. Illustrating that transfer effects may increase with time, experience, and differences.

SIX ARGUMENTS FOR CONSIDERING SAMENESS AND DIFFERENCE IN ACCOUNTS OF TRANSFER INSTEAD OF SAMENESS ALONE

Can a Principle be Discerned From One Example?

Let us first consider Judd’s (1908) original experiment from the point of a more recent transfer study by Gick and Holyoak (1980). As mentioned previously, in the
Gick and Holyoak study the learners first mastered a certain problem (Situation A) that the researcher saw as an instantiation of a principle. Then the participants were presented with another problem (Situation B), which was also perceived (by the researcher) as an instantiation of the same principle. The assumption was that in the case of successful transfer, the learner discerned the general principle and its specific instantiation in Situation A and applied it to Situation B. Specifically, participants in the Gick and Holyoak study were supposed to discern from the fortress problem the general principle that simultaneously converging paths yield increasing intensity and apply that principle to the radiation problem.

In order to grasp a general principle by handling an instance of that principle, the general principle must be discerned by separating it from the specific instance in which it is embedded. But how can such a separation be brought about? This is actually the classical problem of general qualities and specific instances. The former are what dwell in Plato’s world of pure ideas, whereas the latter are what populates our world. How can we tell the two apart? For example, how can we distinguish “threeness” from three apples, the idea of serendipity from a serendipitous event, or as in Gick and Holyoak’s (1980) case the idea of simultaneously converging paths from destroying a tumor in the stomach by radiation?

As I have argued, in relation to the Greeno et al. (1993) and Lobato and Siebert (2002) studies, discernment or separation can hardly come about by focusing on one instance only, in which the general and the specific are completely intertwined; the former implicit, the latter explicit. If there are two sufficiently different instances of the same principle, what is common (the principle) may possibly be discerned from what is different (the instances). The more different cases available to the learner, the greater the likelihood that the principle will be discerned as the primary, or only, thing common to all, because of the likelihood of ruling out what is different (Reeves & Weisberg, 1994). The likelihood of distinguishing the principle and its specific instantiation can also be enhanced by drawing learners’ attention to the commonalities between the different instances.

The main point is that in order to discern the general principle to be used in the second problem, empirically at least two examples are needed. The traditional idea of transfer—learning something in Situation A (discerning a general principle) and using it in Situation B—is logically untenable. As all the transfer experiments quoted by Lave (1988) demonstrated, the first instance in which learners discern a principle may actually occur when they are dealing with Situation B rather than Situation A. From Lave’s point of view, this can be seen as an example of local construction of the solution (i.e., the solution does not exist prior to the problem being solved). From the point of view of my own line of reasoning, this is an example of the necessity of variation. Without different instances (at least two) the learner is most unlikely to become aware of the general principle. In fact, in a follow-up of their study, Gick and Holyoak (1983) showed that although learners had little success in abstracting (or separating) generalized solutions from single spe-
pecific instances, they frequently managed to do so when dealing with two different instances. One could say, of course, that several examples make it possible for the learner to discern the general principle because the examples all embody the same general principle. It is true, provided the examples are different. Using the same example twice or several times instead would not do.

Preparation for Future Learning

As stated, this article is intended to challenge the doctrine of sameness, which is the idea that when learners profit in a new situation from what they have learned in an earlier situation, they do so because they make use of the same capability in relation to the same features of the two situations. I am going to use a recent study as the basis for my second argument. This study was chosen because it differs in important respects from the studies discussed so far. The study in question belongs to a set of studies oriented toward the preparation for future learning (e.g., Bransford & Schwartz, 1999; Schwartz & Bransford, 1988). In these studies the object of research is the effect of learning in one situation on learning in a second situation and, through that, on achievement in a transfer situation. Transfer has also previously been studied in terms of the effects of learning in one situation on learning in another situation, but the preparation for future learning studies have unique features that are essential to the point I want to make.

The point of departure in these studies is the standard transfer paradigm aimed at comparing the transfer effect of two different conditions for learning. In a traditional transfer approach, two comparable groups of students try to learn what is nominally the same thing under two different conditions, and they are tested for what they have learned under the same condition novel to both. To the extent that there are differences, they are in transfer from the two different conditions for learning to the same condition for appraisal.

Now, it might be the case that no differences are found between the two conditions. One of the conditions might, however, be better at preparing the learners for future learning, even if it does not yield superior achievement directly in a standard transfer design study. In order to investigate this, Bransford and Schwartz (1999) suggested the use of what Schwartz and Marting (2004) called “the double transfer” design (p. 147). Students were assigned one of two instructional treatments (Situations A1 and A2). Half of the students from both treatments were given access to a common learning resource (Situation B), such as a lecture or a sample worked-out problem, followed by a request to solve a transfer problem (Situation C). The other half from both treatments were asked to solve the transfer problem (Situation C) directly without access to the learning resource. The researchers called this a double transfer paradigm because students needed to “transfer in” what they had learned from the instructional treatment to learn from the resource, and they needed to “transfer out” what they had learned from the resource to solve the target transfer
problem. To the extent that there were differences, they were between the two conditions (Situations A₁ and A₂) as far as their direct effect on the transfer task (Situation C) and their effect on learning in Situation B are concerned. The differences might thus have originated from two kinds of transfer effects. What follows is an example of how such a design has been used in a specific study.

Schwartz and Martin (2004). In this study, the object of learning (what was to be learned) was comparing (high) scores on different scales from different distributions, operationalized in terms of the target task given after instructions. The following task was utilized: “Which of two students, who were in different biology classes and took different tests, did better on their respective test?” Such problems are commonly solved in statistics by using standardized scores (i.e., by dividing the difference between the actual score and the mean for the distribution that the score belongs to by the standard deviation for the same distribution). In order to grasp the rationale underlying this transformation, the ninth grade students participating in the study had to notice two critical respects in which distributions might differ from each other, namely central tendency and variability. (There are similarities with the children in Judd’s (1908) experiment, who had to notice the angle of refraction and the depth of water as two critical differences between the two tasks they were dealing with.)

After having dealt with statistics for 2 weeks (all in the same way), the students were presented scenarios, such as the following: “Who broke the world record by the most impressive amount—John in high jump or Mike in javelin throw?” In one of the instructional treatments (Situation A₁), the students were asked to invent their own way of solving the problem given raw data for the distributions of the best results during the year. In the other instructional treatment (Situation A₂), the students were shown a graphical method for standardizing the scores by using histograms and a method for comparing the standardized scores subsequently. They kept practicing this method under the supervision of the teacher. Half of both treatment groups also received a learning resource—two worked-out examples (Situation B). The learning resource was intended to provide the students with a tool for dealing with the target transfer problem. The students were instructed how to compute and compare standardized scores (e.g., “Is Betty better at assists or steals?”). Finally, all of the students received the target transfer problem, namely to compare scores on two different biology tests (Situation C). In this case not all the raw data were provided, but only the means and standard deviations for the two tests along with raw scores for the two students to compare.

There was no significant difference between the two instructional treatments for the students who did not receive the learning resource. However, the additional component of the worked-out examples in Situation B yielded a striking difference, namely a strong effect for the “invent a solution strategy” group (Situation A₁) and little or no effect on learning for the group that was simply shown a graphical solution (Situation A₂).
The experiment shows that although the students’ attempt to invent a solution for the first problem (Situation A1) did not have any advantage for solving the target problem (Situation C) as compared to telling the students a way of solving it (Situation A2), the attempt had a definite advantage when it came to learning to solve the target problem (Situation B). So the differential effect on the transfer problem was not due to the first part of the experiment (the instructional treatment) nor to the second part (exposure to a learning resource). The effect was due to the relationship between the first and second parts. The effect of the first part was thus contingent on the second part. If we denote the three parts A (by collapsing Situations A1 and A2 into each other), B, and C, we can conclude that no differential effect of A on C and no differential effect of B on C can be identified. The effect is contingent on what other tasks the learners may encounter. So the effect of a task on another task can only be determined specifically in the context of a given set of tasks. In accordance with Lave’s (1988) argument for considering sets of situations, instead of only two situations at a time, changing the set may imply that the effect of a given task on another task will change as well.

But what is the nature of the relationship between Situations A and B? None of the students in the invention group came up with the canonical solution of the problem (standardized scores), whereas students in the tell-and-practice condition group were shown something close to that. Yet, there was a greater positive effect of A1 and B on C as compared to the effect of A2 and B on C. An important element in the invention approach was the use of contrasting cases. As mentioned above, the students were supposed to compare raw scores and distributions to which those scores belonged. By doing so the students tried to determine how the distributions differed from each other, and they then noticed critical differences between them, even if they could not capture those features and their relations to each other explicitly in technical language. If the differences between cases provided the students with one kind of contrast, juxtaposing their own vague ideas about central tendency and variability with the standard statistical versions of the same phenomena provided them with another contrast.

As has been the case with many studies in the preparation for future learning paradigm, Schwartz and Martin (2004) made use of differences (contrasts) on two levels: between the conditions inventing (Situation A1) and being told (Situation B), on the one hand, and within condition A1 (using contrasting cases in inventing), on the other. By encountering differences between cases, the learners noticed differences that were critical for distinguishing between them.

This line of reasoning leads to the conclusion that what was learned in this case was the increased sensitivity to perceive certain situations in certain ways. That is, students developed a capability to discern aspects of the situation as critical or relevant and take them into consideration at the same time. Additionally, they began to recognize relationships between those aspects as relevant. This implies that the acts of discernment (that which is learned) and what is discerned in the situation
(the features of the situation toward which the acts are directed) are not separate. One cannot discern without something being discernable, nor can anything be discerned without an act of discernment. Accordingly, learning about central tendency or variability, for instance, amounts to learning to discern central tendency and variability in the distributions one encounters in the future.

Why Understanding of One Theory is Necessary for the Understanding of Another Theory

I can go further and argue that it is actually impossible to grasp anything without having experienced an alternative option. It is, for instance, difficult, if not impossible, to understand a theory without having come across an alternative theory of the same phenomenon. How could one otherwise distinguish between the theory and the phenomenon? But this means at the same time that it is impossible to understand the first theory of a phenomenon that we encounter. Arguments for having history of science as part of the science curriculum are in line with this claim:

Thus, by contrast (and paradoxically for many), learning about the Aristotelian paradigm of force-motion relationship fortifies understanding of its Newtonian counterpart, the Cartesian interpretation of weight helps students to understand Newtonian gravitation, the idea of absolute space–time reveals the meaning of the relativistic conception, the geocentric world system facilitates understanding of the heliocentric model … . (Tseitlin & Galili, 2004, p. 21)

Aristotle and Newton. Let us examine briefly the first of these examples. According to Newton’s first law of motion, a body remains at rest or continues to move with the same velocity unless a force acts upon it. This can be illustrated by a spaceship moving from the earth to Venus with all of its engines switched off. If this were all we focused on, Newton’s law would not even seem a law, but rather a straightforward empirical generalization, much like the observation that things fall when we drop them. In order to perceive the significance of Newton’s formulation, we have to step outside it. In fact, looking at what is happening on the earth instead of looking at what happens in space would do. If one tries to switch off the car engine while traveling at 50 km/hr and then declutches, she will notice that the car will not continue to travel at 50 km/hr. It slows down and soon comes to a halt. As no force appears to act on the car, we may conclude that Newton must be wrong. But then someone might remind us that air resistance is one force that acts on the car running forward with the engine switched off, and another is friction in the axle and on the road. It is these forces that make the car stop. If there were no air resistance and no friction, the car would continue to move forward at 50 km/hr. There is always air resistance, and there is always friction in the world in which we live. By juxtaposing Newton’s first law with our everyday experience, we can conclude that far from making a rather trivial
observation, Newton saw and formulated a principle that cannot be seen directly at all on earth. Our everyday experience is more in line with the Aristotelian idea that in order to make something move we need a force. Now, this principle probably seems straightforward to most of us, yet it contradicts the Newtonian formulation. How is this? What Aristotle says is that if a body is at rest, it takes a force to make it move. This is true. But the principle is overgeneralized if we believe that if something is moving there must be a force that keeps it moving.

The difference between Aristotle and Newton is that the former tries to explain the difference between rest and movement, whereas the latter tries to explain differences in velocity, where rest can be seen as a special case of zero velocity. Change in velocity is called acceleration (or deceleration). Change in velocity, for example between rest and movement, requires a force. Movement with constant velocity (including rest) does not need any explanation according to Newton. We can see the difference between Aristotle and Newton in terms of two different ways of making distinctions. The former makes a distinction between rest and movement, whereas the latter makes a distinction between constant velocity and changing velocity. I can draw two conclusions (at least) from this. First, the Newtonian way of making distinctions between bodies in movement is more powerful than the Aristotelian way (according to the study of physics). Second, the Newtonian way of making the distinction becomes visible when it is juxtaposed with the Aristotelian.

This is the tentative answer to why understanding one theory is necessary for the understanding of another theory. But does it really work like this in practice? Galili and Hazan (2000) provided empirical evidence that it does indeed. They compared the understanding of optics by students engaged in an experimental year-long historically oriented physics course (the target group) with that by students who participated in a comparable conventional course in physics (the comparison group). Many scientific ideas were appropriated by virtually all of the students in the target group. By way of contrast, the students in the comparison group showed very little understanding. The main mechanism behind this difference in achievement was the frequent use of contrasts between different ideas of the same thing in the experimental treatment. For example, in the comparison group, the view of light rays was reified as “what light comprises.” In contrast, in the target group, light rays were conceived as an auxiliary tool in modern science, as a result of exploring contrasting ontological claims of historical models.

My interpretation is thus that the important function of the introduction of alternative conceptions; of historical origin; or of the students’ own naïve views, judged to be wrong by current science; is that those alternative conceptions make the conceptions judged right by current science visible. Without a contrast, all students can do is learn the “whole,” often by rote, and this does not prepare them for handling novel problems in powerful ways in the future. For example, the notion of an auxiliary tool does not have meaning without access to something that is not an auxiliary tool, such as the ontologically much “heavier” reified view of “light ray.”
Learning about the antique view makes it possible for the learner to understand the modern view, not because the two are similar, but because they are different. Although the two views are views of the same thing, the positive (transfer) effect does not derive from perceiving what is same, but from discerning what is different. My suggestion is that learning about the antique view and then learning about the modern view is more conducive to learning than learning about the modern view twice. This should be true, at least if learning is measured in terms of the learner’s discernment of critical differences between different instances of the phenomenon in question (e.g., force and motion or light rays).

Perceptual Learning

Perceptual learning as an academic specialization was established by Eleanor J. Gibson. In 1955 she published a highly influential article in *Psychological Review*, “Perceptual Learning: Differentiation or Enrichment?,” coauthored by her husband, James J. Gibson. The authors argued that there are two schools regarding perceptual learning. According to the enrichment school, we receive scarce, impoverished information from the environment, which must be added to and enriched. According to the differentiation school, we receive so much information from the world that we have to differentiate or select. Consequently, learning to perceive amounts to learning to find the differences that are most critical in relation to our goals. J. J. Gibson and Gibson (1955) demonstrated that learning to know something (in the sense of being able to recognize it) is a matter of learning how it differs from other things (i.e., becoming able to discern the respects in which it differs from other things).

The ecological approach to perceptual learning championed by Eleanor Gibson belongs clearly to the differentiation school. In her book with Anne Dick, she included the Gestaltists in the same camp, whereas the behaviorists and cognitivists were listed under enrichment theories (E. J. Gibson & Dick, 2000).

Learning to differentiate means making finer and finer discriminations. This differentiation amounts to becoming attuned to distinguishing features—or critical differences—that can be used for making distinctions. These distinctive features are simply dimensions in which things vary, and they can be used for telling apart instances from noninstances. Thus, perceptual learning amounts to discerning distinctive features or critical dimensions of variation. If I now apply the terminology of transfer to perceptual learning, it means that in Situation A the learner learns to differentiate between instances or noninstances, whereas in Situation B she encounters something that is either an instance or a noninstance. To the extent that the learner is better at correctly differentiating between instances and noninstances in Situation B as a result of having participated in Situation A, then there will be a positive impact from Situation A on Situation B. If we agree with the Gibsonian view of perceptual learning, then performance in Situation B is a function of the
differences within Situation A, as well as the differences between Situations A and B. Eleanor Gibson (1969) cited a transfer study carried out by Vurpillot et al. (1966) in which it was demonstrated that perceptual learning involves learning and transferring distinctive features that separate instances from noninstances (as opposed to learning and transferring features that the instances have in common).

This is true, of course, only if the learners are given opportunities for making such discriminations. An observation from Pavlov is relevant here. He found that reflexes that were conditioned to a certain stimulus were generalized to other similar stimuli (due to sameness). The next question was, then, how can differentiation be learned? Pavlov believed that there were two options: One could either repeatedly reinforce the conditioned stimulus, or one could introduce an unreinforced contrast along with the reinforced conditioned stimulus. He found the first method did not work. There was no differentiation even though the stimulus was repeated with reinforcement more than 1,000 times. On the other hand, even a single application of unreinforced contrast could lead to rapid differentiation (Pavlov, 1927, p. 111, as cited by E. J. Gibson, 1969, p. 100).

The chapter in E. J. Gibson’s (1969) book on differentiation theory begins with a quote from the 19th-century British philosopher and theologian, James Martineau, that explains well the differentiating force of contrasts:

When a red ivory ball, seen for the first time, has been withdrawn, it will leave a mental representation of itself, in which all that simultaneously gave us will indistinguishably co-exist. Let a white ball succeed to it; now, and not before, will an attribute detach itself, and the color, by force of contrast, be shaken out into the foreground. Let the white ball be replaced by an egg, and this new difference will bring the form into notice from its slumber, and thus that which began by being simply an object cut out from the surrounding scene becomes for us first a red object, than a red round object, and so on. (Martineau, Essays Philosophical and Theological, as quoted by E. J. Gibson, 1969, p. 95)

This is an interesting quote in more than one respect. First, it describes what we might call retrospective transfer: how the image of an object is affected by experiences following the birth of the image. Second, we can look at the effects in the reverse direction, namely how the perception of objects is affected by previous experiences. The observer may not have separated out and noticed the form of the white egg without having seen the white ball just before, nor may she have separated out and noticed the color of the white ball without having seen the red ball just before.

Motor Learning

In 1975, Schmidt published his article “A Schema Theory of Discrete Motor Skill Learning,” which has become highly influential in the field of motor learning. One
of the best known implications of the theory is the variability of practice hypothesis, which states that varied practice (“Situation A” in transfer terminology) is likely to enhance performance in a new situation (Situation B). What is interesting is that the relationship is not considered as a function of sameness between two situations, but as a function of differences across situations. The hypothesis received a great deal of support from empirical tests during the years following its publication (Shapiro & Schmidt, 1982; Shea & Wulf, 2005; Sherwood & Lee, 2003).

Let us look at a typical study carried out by Moxley (1979). Eighty children, aged 6 to 8 years old, participated in the experiment and were distributed across two conditions. The task was to try to hit a target (a carpet) with a shuttlecock. In one of the conditions the children threw 20 times from five different angles, each from the same distance, sitting on the floor with their feet pointing toward the opposite wall. In the other condition the children had to throw shuttlecocks 100 times from the same position. Afterwards both groups tried to hit the target with shuttlecocks from a new position. The group subjected to the varied condition succeeded best with the criterion task.

Kerr and Booth (1978) carried out a similar study, in which the task was again to hit an object, but the critical difference between attempts was the distance to the target. The task was to hit a target with miniature beanbags. Thirty-six children with an average age of 8.3 years and 28 children with an average age of 12.5 years participated in the experiment. Both age groups were randomly assigned to specificity and schema groups. All children were tested in the beginning and at the end of the experiment, the younger group throwing from 3 ft and the older from 4 ft. The deviation from the target was measured after each throw and the difference between the average error at the beginning and at the end of the experiment was the measure of the effect of the practice between the two tests.

The practice for the specificity group consisted of a great number of throws, all from the criterion distance of 3 ft for the younger children and 4 ft for the older children. Children in the schema group practiced at a variety of distances, none of which was the criterion distance. It is interesting to note that the children in the schema group outperformed the children in the specificity group. Therefore, practicing something other than what was tested was more effective than practicing exactly what was tested.

Although the variability of practice hypothesis received much empirical support, Schmidt’s (1975) theory has been seriously questioned (see Shea & Wulf, 2005; Sherwood & Lee, 2003). I would also like to interpret the variability of practice hypothesis in a way that differs from Schmidt’s original formulation. He believed that varied practice enhances future performance with novel tasks thanks to the contribution of varied practice to a more robust schema governing behavior. My interpretation is that variation is necessary for discernment, and that the learners must discern critical differences between situations in order to be able to adjust to new situations.
Making Systematic Use of Variation to Enhance Learning

Understanding the relationship between what students learn in the classroom and what is happening in the classroom remains a central question in educational research. This question has typically been addressed by relating learning outcomes to the ways in which learning is organized (individualized instructions, whole-class teaching, project learning, peer learning, reciprocal teaching, and so on). Describing differences in how learning is organized does not necessarily say anything about differences in the content of learning or about differences in what learners attend to. In order to identify the latter, we have to compare different classrooms in which the same content is addressed. When we do so, we might notice that when specific content is dealt with, some of its aspects are varied, whereas others remain invariant. When looking at the graphical representations of linear equations of the form \( y = ax + b \), for instance, the teacher might keep \( b \) invariant (e.g., \( b = 0 \)) and substitute different values for \( a \) (e.g., \( a = 1, \frac{1}{2}, 2, 3 \ldots \)), thus focusing attention on the slope of the line. In a similar manner, the teacher might keep the slope invariant and vary the \( b \) value, thus focusing attention on parallel lines and varying intercepts. If only the slope is varied, the students might learn one thing. If only the \( y \)-intercept is varied, the students might learn another thing. If both parameters are varied, the students might learn a third thing. If both sources of variation are mixed in a random fashion, students might learn differently again. If there is one line in a coordinate system on the screen that the students can change by plugging different values into the equation \( y = ax + b \), the outcome may again be different.

Differences in what varies and what is invariant make different things possible to learn. Consequently, the pattern of variation and invariance when dealing with some particular content is highly correlated with what the students learn about that content. For example, Marton and Pang (2006) found that in five comparable high-achieving secondary school classes, taught by five highly experienced teachers, the proportion of students who mastered a difficult economic concept varied between 6% and 97% depending on the different patterns of variation and invariance and on how systematically those patterns were used in the classrooms. In the next two sections I explore the effect of the systematic use of variation in classrooms by considering several examples.

Learning Cantonese words. In two comparable second grade classes in Hong Kong, the students were supposed to learn seven Chinese words in the context of a short story (Chik & Lo, 2004). Their success was measured by means of a “cloze-test,” a short novel text with a number of words deleted, leaving gaps between the words. The students were asked to fill in those gaps with the appropriate words. It was possible to fill the gaps by using the seven words that had been studied. In Class A, all 30 students completed the task correctly, whereas in Class B only 9 of 31 students did so. When comparing how the same content was handled
in the two classes, it was found that in Class A different aspects of each word (form, meaning, pronunciation) were dealt with in conjunction with one other (i.e., these aspects were kept together for each word). By way of contrast, in Class B each aspect was dealt with separately (i.e., words were grouped according to each aspect). This meant that in Class A the integrity of the Chinese words was maintained (their related aspects were kept together), whereas in Class B the aspects were kept separate and the words appeared once for each aspect.

The same content was thus dealt with in both classes, but differently. In Class A, for each word the aspects varied. In contrast, in Class B, for each aspect the words varied. Thus, the differences had to do with how the content was structured, what varied primarily and what varied secondarily, and hence what learner attention was attracted to.

In many of the previous examples presented in this article, I have argued for the importance of contrast and of variation in critical aspects. In the study of Cantonese words, however, it was important that a number of critical aspects of words (form, meaning, and pronunciation) varied together (from word to word). As the words were basically known as spoken words to the students already, in order to link the spoken words to the written forms (characters) and to the elaborated meanings, it was advantageous for the different aspects to vary together. This shows that there are no patterns of variation or invariance that are superior for learning or for transfer in general. The point I want to make is that patterns of variation and invariance are critically important for learning and transfer, but the specific pattern depends on the particular object of learning and on the group of learners participating in the learning experience.

**Learning Cantonese tones.** I now extend the discussion to consider an instructional approach in which the pattern of variation and invariance is theoretically based. Ki and Marton (2003) investigated the difficult process of learning Cantonese tones by adult foreigners. Cantonese is a complex dialect of Chinese, and spoken Chinese words are distinguished by both sound and tone. This means that two or more words pronounced with the same sound can have (and usually do have) two or more different meanings based on distinctions in tone or pitch. There are six different tones in Cantonese: high level, high rising, mid level, low level, low falling, and low rising. One might believe that Cantonese speakers are more sensitive to tonal differences, but this is not the case. Stagray and Downs (1993) compared native speakers of English and Cantonese and concluded that the former
actually made finer discriminations in pair-wise comparisons of singular tones. Thus there must be another explanation for foreigners’ difficulties with Cantonese.

Differences in pitch (of which tones make up a special case) are also important in nontonal languages, but in these they are used to make distinctions at the sentence level (e.g., between interrogative and imperative modes), not to indicate differences in meaning at the word level. Consequently, when a speaker of an atonal language hears a tonal language in which every word has both a sound and a tone component, she automatically tries to distinguish between meanings in terms of differences in sound alone. By doing so she remains incapable of making essential distinctions in perception or in production between words that have the same sound but different tone components. When variation in sound is suppressed and variation in tone and meaning is afforded, the learner must, however, use tonal variation for making distinctions between meanings.

Ki and Marton (2003) found that by inviting nonnative learners to learn sets of Cantonese words that have the same sound but different tone components, they learned to attend to tonal differences in order to make distinctions between meanings on the word level. This represented a fundamental switch in learners’ attentional field. Not only did the learners get better at distinguishing between meanings by means of tones as regards the words in the exercise, but they also got better at learning to distinguish between meanings by means of tones when variation between sounds was not suppressed. In a follow-up study, Ki and Marton (2005) demonstrated that by first separating variation in tone and meaning and keeping sound invariant, then separating variation in sound and meaning and keeping tone invariant, and finally letting all three dimensions vary simultaneously, learners can make great progress toward handling the complexities of everyday use of Cantonese.

**Learning to spell the tj (C) sound in Swedish.** Another study showed that what distinctions the learners can make—and have to make—is highly important for learning. A group of Swedish teachers, working with a researcher, designed a lesson that they thought would contribute to improving students’ spelling of words with a tj sound (Holmqvist, Gustavsson, & Wernberg, 2005; in press). There are different ways to spell this sound that conform to a rules-governed system based on consonant–vowel combinations with certain exceptions. One of the teachers carried out the lesson designed by the group. Afterwards the group redesigned the lesson, and it was carried out by a second teacher with another group of students. The group revised the lesson again, and it was carried out by a third teacher with a third group of students. The students participating in these three lessons came from different classes, and they were kept together for one lesson only, after which they returned to their own classes. All three groups took the same spelling test immediately after the lesson, again after 4 weeks, and once again after another 8 weeks. One of the groups lagged behind the other two groups initially but surpassed both groups some time after the experiment. This group improved on
each test, whereas the other two groups remained at the same level or performed slightly weaker. The researchers investigated how the instructional treatments contributed to the differences across groups.

The two groups that succeeded less well in the long run had been given a worksheet in which all of the tj sounds were removed in a text, leaving gaps in their places. The students were asked to fill in the missing letters. The same text had been used with the group that kept improving its results after the experiment. However, in this case, the text was unmodified and the task was to underline all of the tj sounds. In the former case the students did not have to differentiate between the tj sound and other sounds. Actually they did not even need to read any other words than those with missing letters. They had only to differentiate between the different spellings of the tj sound. Under the second condition, however, the students had to differentiate between what signals a tj sound (should be underlined) and what signals a sj (S) sound—which in Swedish is very close to the tj sound. Thus, the students had to differentiate between different spellings of the same sound and different spellings of another sound.

To explain why one group improved more than the two other groups after the learning occasion, the researchers conjectured that the students in the more successful group learned to pay attention to the differences between the tj sound and the sj sound (and other sounds), as well as to different spellings of the tj sound as compared to different spellings of the sj sound. The students became sensitized to a wider range of differences than the students in the other groups. Hence they had more opportunities for practicing what they had learned whenever they read or wrote something in Swedish subsequent to their participation in the experiment. This condition was simply a better preparation for future learning, in the sense used by Bransford and Schwartz (1999).

**Providing for individual differences.** The last study described in the previous section was a blend of Japanese “lesson study” and design-based research (Holmqvist et al., in press). In such a blended approach, called learning study, an object of learning is chosen by a group of teachers assisted by a researcher. The students’ prior understanding of the object of learning is revealed by means of a pretest. One or more lessons are designed to help students develop an understanding of the chosen object of learning by building upon students’ prior understandings. The lesson (or sequence of lessons) is carried out by one of the teachers. Students are assessed after the lesson, and the results are discussed in relation to what has taken place in the classroom. As a result, the lesson is revised and taught by another teacher. One study might comprise one to four such iterative cycles.

Twenty-seven classes participated in a research and development project aimed at providing for individual differences through learning studies in Hong Kong. Special attention was paid to low-achieving students. Systematic use of variation was adopted as the main approach to the enhancement of learning in the project (Lo, Pong, & Chik, 2005). Seventeen lessons were designed, with a number being revised
once or twice, for a total of 27 lessons. In all of the lessons, variation was used in systematic ways. In 25 of the 27 cases, students performed significantly better after the lesson than they had before. Furthermore, this improvement was significantly higher among initially low-achieving students than among high-achieving students, even when correction was made for ceiling effects. Transfer to the annual attainment test was found, especially for the low-achieving students. Finally, researchers observed that high-achieving students developed insights that went well beyond the object of learning and hence were not measured by the assessment tools.

**Historical examples of the systematic use of variation.** Although the studies in this section are not transfer studies, they are relevant to the discussion because they demonstrate the historical roots of the systematic use of variation and invariance in instruction. In China there is a well-known approach to teaching mathematics called *teaching with variation* (Gu, Huang, & Marton, 2005). Kangshen, Crossley, and Lun (1999) documented the use of this method dating back nearly 2000 years. Gu (1991) nicely captured its appearance in modern practice by demonstrating that more successful mathematics teachers use the teaching with variation approach more often than do less successful teachers. One example of the systematic use of variation can be found in the use of nonstandard figures in order to investigate aspects of geometric figures that students falsely assume to be critical. Polygons are, for instance, always drawn in textbooks with a horizontal baseline. Rotating them can help students realize that angular orientation is not a critical feature of polygons. Additionally, presenting noninstances of a concept can help students become aware of a critical feature. For instance, by presenting a line perpendicular to a radius and intersecting the circle in two points, the students’ attention is drawn to the fact that a tangent of the circle has only one point in common with the circle.

Mason and Watson (2005) analyzed textbooks in mathematics used in England from the 16th century on. They demonstrated that in many cases the authors made systematic use of variation and invariance by building sequences of tasks that bring out critical differences between them.

A final example of the systematic use of variation and invariance can be seen in Maria Montessori’s pedagogy. She was inspired by two French physicians, Jean Marc Gaspard Itard (1775–1838) and Eduard Seguin (1812–1880). The former became famous for his treatment of the Wild Boy from Aveyron, an 11- or 12-year-old boy who had grown up in the forest without any contact with human beings. An important principle used by Itard was to try to teach the boy to make distinctions by means of contrasts (e.g., learning to tell apart a round figure from a flat figure by touch, red paper from blue paper by vision, and bitter from sweet by taste). Itard also tried to link the distinctions between sensory impressions to distinctions between corresponding words by means of simultaneous variation in sensory impressions and words (see, for example, Itard, 1801/1962; Lane, 1977). I can compare Itard’s approach with the more common approach of introducing one
concept or word at a time and pointing to different instances of that concept (e.g., looking at a red rose and saying “red,” followed by pointing to a red apple and saying “red,” and so on). The difference between Itard’s approach and the more common approach is the difference between the use of contrasts and generalization.

Seguin, who was Itard’s follower, worked with retarded children. A central element in his work was the training of the senses by means of the physiological method. Seguin wanted to help children develop finer and finer discriminations with all their senses. He did so by producing sensations that were at first far apart (e.g., quiet sound/loud sound; light touch/heavy touch; dim light/bright light; and so on), and then later producing sensations that were closer and closer to each other (Seguin, 1907). Seguin was considered to be uniquely successful in developing the retarded children’s capabilities to deal with the world around them (Marton & Signert, 2005).

The training of the senses was a central element in the pedagogy that Montessori developed for children without specific handicaps (as described by Signert, 2005). Montessori believed that it was very important to isolate different sense dimensions and to let the children focus all of their attention on one specific dimension at a time. Consider the following example, aimed at letting children discern different dimensions of geometric figures. Four sets of materials are utilized:

Set 1. Cylinders varying both in diameter and height. The tallest cylinder has the largest diameter and the shortest cylinder has the smallest diameter.
Set 2. Cylinders varying both in diameter and height, as in Set 1, but the tallest cylinder has the smallest diameter and the shortest cylinder has the largest diameter.
Set 3. Cylinders varying in diameter only. They all have the same height. Each cylinder is 0.5 cm larger or smaller than the next one.
Set 4. Cylinders varying in height by 0.5 cm per step, all having the same diameter.

Noticing systematic differences of this kind and solving problems by noticing such differences frequently captures young children’s intense and uncharacteristically sustained attention, as illustrated by Montessori’s own examples (see, for instance, Montessori, 1917/1965, p. 67).

IMPLICATIONS

Transfer as a Function of Differences (and Sameness)

The point of departure for this article is the view that what the learner learns to do in a certain situation may enable her to do the same thing in another situation to the extent that the second situation resembles (or is perceived to resemble) the first sit-
uation. I set out to challenge this view, arguing that not only sameness, similarity, or identity might connect situations to each other, but small differences between them might connect them as well. My interest has been how what the learner learns in some situations might enable her to do something different in other situations thanks to perceived differences (and similarities) between situations. I agree with the above-mentioned four points made by Lave (1988) in her critique of the culture of transfer experiments, that (a) answers to the question “What is learned?” should not be taken for granted; (b) any generalizations made by the learner contribute to connecting situations; (c) we should focus on relations between sets of situations rather than on only two situations at a time; and (d) connecting situations is done in all of the situations involved as opposed to “what is learned” being applied more or less automatically to a single situation. The main focus of this article is, however, a point not raised by Lave, namely the role of differences in connecting situations.

In the first example in the section on motor learning, we considered a study in which children hit a target with a shuttlecock from different angles. In order for children to become aware of the dimension of angle variation, children must be able to discern it in the very throwing in order to be able to adjust to changes in it. It is only by encountering differences, changes, variation, that they become able to discern it in the throwing. So in order to be able to handle a new angle in Situation B without starting from scratch, they must have encountered throwing from another angle before, in Situation A, or they must have encountered throwing from different angles before. In order to be able to discern a certain dimension of variation, however, not only must the learner experience variation in that dimension, but she must also experience invariance in other dimensions. Nothing can be discerned when everything varies, just as nothing can be discerned when nothing varies. In many of the examples presented in this article, it is by enabling the learner to discern a critical dimension of variation that one situation is linked to another. This was certainly the case in Schwartz and Martin’s (2004) study, in which learners in one condition became better than learners in the other condition at discerning the dimension of central tendency and the dimension of variability.

There are values in the dimensions of variation. For example, every specific angle is a value in the dimension of angles, and every mean for every specific distribution is a value in the dimension of central tendency. Contrasts between values enable learners to discern values. One can never discern just one value. It is the difference that is discerned, and to experience a difference takes at least two values. Through contrast, the learner can discern values as well as the dimension of variation in which they are values. This is what Watson and Mason (2005) referred to as the range of permissible change in a dimension of possible variation.

Let us take another look at the Cantonese tone example. As a foreigner in China, when I hear a word (“A”), I cannot discern its two vocal aspects—the sound and the tone—unless the tone differs markedly from the intonation pattern that I am used to. This is Situation A in the traditional transfer perspective, and it is followed
by another word (“B”), which has the same sound as A but a different tone (Situation B). Then all of a sudden I can hear the tone in B and retrospectively also in A (thanks to the contrast). Without having heard A, I would not have been able to discern the tone in B. We could thus call it transfer (according to the above definition of transfer as the effect of what someone learns in one situation on what she can do in another situation). But I am not doing the same thing when hearing B as I did when hearing A (cf., the definition of transfer as the extent to which someone can do the same thing that she learned to do in one situation, also in another situation). When hearing A, I could not differentiate its tone from its sound; yet I can discern the tone in B after hearing A. Although the sameness of the sounds across the two words was a necessary condition for discerning the tone, it was the difference—and not the sameness—that was attended to, discerned, and transferred. Neither repeating Word A nor repeating Word B would have allowed for that.

We find the same phenomenon in the other examples presented in this article. The students in Schwartz and Martin’s (2004) study benefited from the learning resource and succeeded in the invention activity because the two arrangements for learning were different (but similar in terms of content). Students may have grasped some of Newton’s ideas thanks to having read about Aristotle’s different ideas (about the same thing). The form of an egg is discerned because it differs from the previously encountered ball (with which it shares the color), according to Martineau’s observation of the differentiating force of contrasts (as cited in E. J. Gibson, 1969). In each of these cases, discernment is made possible because some things are the same whereas others are different.

Furthermore, in order to discern a dimension of variation, it must also vary, while other dimensions remain invariant. In the Cantonese tone example, we might not only discern two different tones, but we might also become aware of the fact that there is such a thing in Cantonese words as tones. That is, we might be able to discover tones as a dimension of variation. As a result, we could then distinguish that dimension from other dimensions, such as sound. In order to be able to speak and understand Cantonese, we must be able to discern the different aspects of words when they all vary. After all, this is what Cantonese speakers unconsciously do.

So far I have only dealt with the simplest relationships between discernment and differences or variation in this section. I have talked about the discernment of dimensions of variation or separation, and I have talked about the discernment of values, thanks to contrasts (Marton & Tsui, 2004). Perception of one thing affects the perception of another thing. But even these simple relationships are sufficient, I believe, to illustrate that what people learn to do in one situation might enable them to do something different in another situation, thanks to differences (and similarities) they perceive between or within the two situations.

In the description of Itard’s (1801/1962) approach to teaching the Wild Boy from Aveyron color names, I mentioned an example of the most common use of variation in educational contexts, namely generalization (Marton & Tsui, 2004). It
refers to the introduction of a concept or principle followed by a great number of examples of the concept or principle in question. Dienes (1963) developed an approach to the teaching of mathematics based to a large extent on this specific pattern of variation and invariance. Indeed, invariance in the midst of change is a central theme of modern mathematics (Mason & Johnston-Wilder, 2004).

When learning to deal with novel situations, people have to learn—as a rule—to take different features into consideration simultaneously. As an example, consider the situation of children learning Chinese words. In one of the classes the teacher created conditions for linking the three different aspects of Chinese words—character, pronunciation, and meaning. This happens when the different aspects vary together, resulting in what Marton and Tsui (2004) refer to as fusion. Another example of fusion is Itard’s (1801/1962) attempt with the Wild Boy from Aveyron to link differences in sense impressions to differences in words, denoting them by simultaneous variation in both.

In the introduction to this article I expressed agreement with Lave’s (1988) critique of selecting two situations and looking at the relationship between them in the researcher’s narrow perspective only. The interesting question is how we manage to discern critical aspects of the widely varying concrete and complex situations of everyday life. In this article I have only tried to illustrate some fairly straightforward relationships between discernment, on the one hand, and variation or invariance on the other. Far more complex relationships between discernment and variation or invariance are elaborated upon by Marton and Tsui (2004).

Pedagogical Aspects of Transfer

For much of the 20th century, transfer was a well-researched topic in educational psychology. More recently, transfer is a much-criticized concept. Why does transfer continue to attract so much attention and interest? The reason is, I believe, that the question of how what we learn in one situation effects our capabilities for handling novel situations is, as Lave (1988) rightly pointed out, absolutely central to the educational enterprise. As I argued above, the question of transfer is closely linked to the question of what is learned. If we were to cast the transfer issue in a pedagogical form, we would not only ask “What is learned?” and “What is transferred?,” but also “What should be learned?” and “What should be transferred?.” We would then ask how we can possibly increase the likelihood of learners handling novel, unknown situations in the future in powerful ways. Bransford and Schwartz’s (1999) preparation for future learning approach points to such a reformulation of question of transfer. The general answer is that by learning (now) how to learn (in the future), learners will be better able to cope with novel situations.

Generative Learning

The notion of transfer as a function of sameness derives from the deeply seated view of learning as being based on repetition and habituation. However, the funda-
mentally fascinating question, both from the point of basic research on learning and from the point of schooling, is how learning one thing now can prepare us to learn something else in the future. Were we to widen the scope of the concept of transfer, such an effect could be called transfer: One learns something in some situations, and then one becomes better at learning something else in other situations. Let us call it generative learning in accordance with Mestre’s (2002) conclusion from the first National Science Foundation Transfer Conference that generativity should be a main topic for future research on transfer.

One of the main goals of schooling is to enable students to handle novel situations in powerful ways. How can generative learning be instrumental to that? There is no action without perception. We have to perceive, discern, and distinguish in order to make sense of the world around us and to deal with it in powerful ways. We have to construct a personal narrative informed by the narratives of others. It could be the case that different people typically understand the same situation in the same way but handle it in different ways. That is to say, their perceptions may be the same, but they do different things in relation to the same perceived reality. However, a great deal of research shows that differences in expertise are reflected in differences in how situations are perceived or understood. For example, a chessboard is perceived differently by a master chess player than by a novice (de Groot, 1965). Similarly, physics problems look different to physicists than to physics students (Chi, Feltovich, & Glaser, 1981). Even cars look different to more and less successful engineers (Sandberg, 1994). Differences in more and less powerful ways of handling simple arithmetic tasks originate from differences in how the tasks are perceived and understood (Neuman, 1987). Differences in how good university students are at writing essays are closely linked to how they view the task of writing an essay (Hounsell, 2005). Differences in comprehending text are intimately related to differences in how the same reading task appears to different readers (Marton & Säljö, 2005).

Looking at the chessboard and determining the meaning of what we see amounts to learning about its momentary status. Looking at a physics problem and ascertaining what kind of problem it is amounts to learning about that specific problem. Examining a patient as a doctor amounts to learning about that specific patient. Grasping what someone says or understanding a text in a certain way amounts to learning about what someone says or what a text is meant to tell us. In each case we remain in touch with the surrounding world by continuously keeping ourselves informed. This is perception, but it is learning as well. Perceptual learning is then a kind of learning to learn—learning how to learn from and about chessboard positions, physics problems, patients, what we hear, and what we read. Perceptual learning is generative (Holmqvist et al., 2005).

According to the line of reasoning followed in this article, perceptual learning amounts to learning to make distinctions. In order to be able to notice differences, the learner has to have encountered differences. In order to see how something dif-
fers, the learner must previously have seen something that it differs from. Seeing one thing affects how the learner sees another thing subsequently—not because of the sameness of the two, but because of both similarities and differences.

ACKNOWLEDGMENTS

First, I thank Joanne Lobato who organized the National Science Foundation “International Working Conference: Addressing the Transfer Dilemma” (REC–0450208) in San Diego, January 13–16, 2005 from which this article originates. We were actually six people determined to challenge “the doctrine of sameness”: King Beach, Yrjö Engeström, John Mason, Ricardo Nemirovsky, Nephi Thompson, and myself. The joint article still remains to be written. When writing this version of mine, I immensely profited from our initial discussions and from the comments I had received. John Mason, Yrjö Engeström, and Thabit Al-Murani have read an earlier version of this article, and they contributed to its betterment with comments, suggestions, and support. After submitting the article, I received constructive critique and highly helpful suggestions from James G. Greeno, two anonymous referees, and the editors (Janet Kolodner & Joanne Lobato). John Mason read and commented on this latest version of the article, too. Lisbeth Söderberg helped to bring the whole thing together. I am deeply moved by the generosity of all involved.

The Swedish Research Council has financially supported the writing of this article, and I am very grateful for that.

REFERENCES


Lobato, J. (2004). *An international working conference: Addressing the transfer dilemma*. Proposal funded by the National Science Foundation.


