*Instructional Science* **25:** 307–346, 1997. 


# An extended dual search space model of scientific discovery learning

WOUTER R. VAN JOOLINGEN* & TON DE JONG
*University of Twente (*Correspondence address: Wouter van Joolingen, University of Twente, Faculty of Educational Science and Technology, P.O. Box 217, 7500 AE ENSCHEDE, The Netherlands; Phone: +31 53 489 36 37; Fax: +31 53 489 28 95; E-mail: joolingen@edte.utwente.nl*

**Abstract.** This article describes a theory of scientific discovery learning which is an extension of Klahr and Dunbar's model of Scientific Discovery as Dual Search (SDDS) model. We present a model capable of describing and understanding scientific discovery learning in complex domains in terms of the SDDS framework. The concepts of hypothesis space and experiment space, central to SDDS, are elaborated and used as a representation of the learner's knowledge. Also, we introduce a taxonomy of search operations in hypothesis space which allows us to describe in detail the processes of discovery. Our ideas are tested against data of subjects who comment on the discovery processes of a simulated learner. It is found that the conditions for performance a search operation in hypothesis space include both sufficient knowledge of the search operation itself and reasons for choosing a specific search operation. Furthermore, a number of constraints on the search in hypothesis space is discussed: domain specific and generic prior knowledge, learning goals, and personality factors. We conclude with some recommendations for the design of discovery-based learning environments.

**Key words:** discovery learning, scientific discovery, problem solving, simulation-based learning

## Introduction

Discovery learning has been the subject of study in cognitive and instructional science since the work by Bruner (1961). Over the years, however, discovery learning lost a significant part of its appeal for instruction, following criticisms that pointed to the inefficiency and lack of effects of (pure) discovery learning (e.g., Ausubel, Novak & Hanesian, 1978). Recently, discovery learning is getting renewed attention. The first reason for this is the availability of powerful educational simulation environments. Discovery learning is easily facilitated in these simulation learning environments (Alessi & Trollip, 1985; de Jong, 1991; de Jong & van Joolingen, in preparation), because they allow the learner to actively engage in a scientific discovery process by doing experiments, stating hypotheses and testing these and, more importantly, because the more powerful of these environments offer the learner additional support for the discovery process (de Jong et al., 1994; de Jong & van Joolingen, 1995; Reimann, 1989; Shute & Glaser, 1990). A second reason for this renewed attention is that learning by discovery is taking a pivotal position within new approaches to learning and instruction such as constructivism (Duffy & Jonassen, 1991; Jonassen, 1991) and cognitive apprenticeship (Brown, Collins & Duguid, 1989; Collins, 1988; Collins, Brown & Newman, 1989). For a successful (re-)introduction of (simulation-based) discovery environments in instruction a detailed theory of discovery learning is necessary. In the present paper we present such a detailed theory, as an extension of one of the currently prevailing theories (SDDS from Klahr & Dunbar, 1988), and we evaluate our theory against data from subjects operating in a simulation-based discovery environment.

*Scientific discovery learning*

In discovery learning, the main task of the learner is to find the properties of a given domain. These properties are not given directly, but have to be inferred or induced from other data. Originally, discovery learning was studied in the context of concept discovery (Bruner, 1961), but research has gradually evolved towards more complex situations where one can speak of *scientific discovery*. Klahr & Dunbar (1988) list the main differences between discovery learning and scientific discovery learning, among those the necessity of designing experiments in scientific discovery. This paper focuses on scientific discovery learning, however, in the remainder of this article we will also use discovery learning as a short term. A natural start to describing discovery learning is by looking at research on the process of scientific discovery. An important difference between scientific discovery learning and scientific discovery, however, is that in scientific discovery learning the environment in which the discoveries are made is usually chosen or designed specifically for the purpose of discovery learning. This does not mean that the discovery process itself is radically different, but usually, the representation of and access to data in the environment will be such that discovery is facilitated.

An influential class of theories about scientific discovery has been developed by Simon and co-workers (e.g., Kulkarni & Simon, 1988; Qin & Simon, 1990; Simon & Lea, 1974; see also Greeno & Simon, 1984 for an overview). The basic approach taken in this work is to describe scientific discovery as a problem solving process in the tradition of Newell & Simon (1972). For describing scientific discovery two problem spaces are postulated: a *rule space*, consisting of all rules, possibly describing a domain, and an *instance space*, containing data from the domain itself. Hypotheses about the domain are stated by searching the rule space and they are tested against instance space. This marks the difference with 'ordinary' problem solving in which only one problem space of partial or candidate solutions is present and the evaluation of a certain solution is either defined by the problem description itself or some external criterion, or is evident from the solution trace through problem space.

The ideas of Simon and co-workers were tested in a number of studies. The discovery processes of people were compared with the output of a computer program simulating the discovery. Information on the discovery processes of people were obtained from experimental subjects (Qin & Simon, 1990) or from historical accounts of scientific discoveries (Kulkarni & Simon, 1988). The computer programs simulating discovery were based on the General Rule Inducer (Simon & Lea, 1974). For example, in Qin & Simon (1990), subjects were presented decontextualized data about planetary motion and asked to find a rule describing this data, while thinking aloud. Subjects able to find the correct rule, which is Kepler's third law, displayed behavior equivalent to that of the BACON program (Langley, Simon, Bradshaw & Zytkov, 1987).

The ideas expressed by Simon were elaborated by Klahr & Dunbar (1988; Dunbar & Klahr, 1989) in a model called "scientific discovery as dual search" (SDDS). SDDS is proposed to be "a general model of scientific reasoning, that can be applied to any context in which hypotheses are proposed and data is collected" (Klahr & Dunbar, 1988, p. 32). The basic assumption of SDDS is too that scientific reasoning requires search in two distinct but related search spaces, now called *hypothesis space* and *experiment space*. SDDS distinguishes three basic processes in scientific discovery: search hypothesis space, test hypothesis and evaluate evidence. The first process generates a fully specified hypothesis, the second generates a prediction and collects evidence which can be evaluated in the third component. Each process is decomposed into a number of subprocesses including searches in experiment space, running experiments and using prior knowledge, including general manipulative knowledge like "vary one thing at a time" (Lavoie & Good, 1988; Tschirgi, 1980). Klahr & Dunbar (1988) have tested their ideas in the context of the field of operation of a programmable robot called BigTrak. Subjects had to discover a rule describing the function of one of the buttons on this robot. Hypothesis space consisted of possible rules that described the function of the unknown button, experiment space consisted of all possible programs and the resulting behaviors of BigTrak.

According to Klahr & Dunbar (1988) their theory deviates from Simon's approach in two ways. First, they state that the concept of a rule space is too limited for describing discovery processes in semantically rich domains. Second, in the studies described by Klahr and Dunbar, experiment space is more elaborate than a list of data and learners have to construct and perform experiments themselves, rather than just use data that are presented, like in the Qin & Simon (1990) study.

The goal of the current article is to apply the SDDS framework to an even more complex domain than the one used by Klahr & Dunbar (1988). In order to do so, it appeared that the level of description of SDDS is not sufficient to describe discovery processes detailed enough to yield predictions of learner behaviour.

**An extended model of discovery learning**

The model we present here is an extension of the SDDS model by Klahr & Dunbar (1988), and is meant to allow for a more detailed description of learner behaviour in complex domains. The main ingredients of this model are a detailed elaboration of the structure of hypothesis space and experiment space, as well as mechanisms to describe the search in these spaces: search operations in hypothesis space and a representation of learners' knowledge states during discovery.

The need for the first extension can be illustrated by discovery environments used in instruction. In the case of the Klahr & Dunbar (1988) study, the domain was rather simple. Only one rule had to be found and it was clear beforehand which kind of rule would be sufficient for describing a domain. In more realistic discovery environments the domain is often more complex than this. Examples of these environments are computer simulations like SOPHIE (Brown, Burton & deKleer, 1982), STEAMER (Hollan, Hutchins & Weizman, 1984), QUEST (White & Frederiksen, 1990), ELAB (Böcker, Herczeg & Herczeg, 1989), Mach III (Kurland & Tenney, 1988), Smithtown (Shute, Glaser & Raghavan, 1989), Voltaville (Schauble, Glaser, Raghavan & Reiner, 1991), and Refract (Reimann, 1989). In these simulations, the number of variables is typically large, implying that often more than one relation is necessary for describing the domain.

In the case that there is a substantial number of variables present in a domain, finding relations between variables may not be the only part of the discovery task. Also identifying or creating variables or variable classes may constitute a significant task. For example, in Smithtown, a simulation-based discovery environment about economics (Shute & Glaser, 1990; Shute et al., 1989), one of the tasks of the learner is to decide upon which of the economic variables are related to each other. In contrast, in Voltaville (Glaser, Schauble, Raghavan & Zeitz, 1992), a discovery environment on elementary electronics, all variables interact with all others, and learners have to discover the relations between the given variables. Glaser et al. (1992) characterize the discovery tasks in Smithtown and Voltaville as being of a correlational nature and of a classical rule discovery nature respectively. The Voltaville and Smithtown examples mark two cases with respect to the variable structure of a domain. In both cases all variables are given, and in the first case, all variables are equally important, whereas in the second they are not. A third case, where not all variables are given but have to be constructed by the learner, can also be distinguished.[1]

A second characteristic of complex domains is that relations may be formulated at different levels of *precision* (see van Joolingen, 1995; van Joolingen & de Jong, 1993; Plötzner & Spada, 1992), meaning that the evaluation of a hypothesized rule or set of rules is dependent on the precision needed.

These two aspects of complex domains, the large number of variables and relations and different levels of precision, make that we need a refined vocabulary for describing these structures.

The second extension of SDDS is needed because SDDS does not provide us with tools allowing detailed descriptions of search operations in hypothesis space. For instance, SDDS provides no description of different types of search operations in hypothesis or experiment space. Such descriptions are needed if we want a better understanding of difficulties learners have with discovery in specific domains. Therefore, we need a classification of search behavior in hypothesis and experiment space, which describes to a sufficient level of detail the changes in a learner's knowledge about the domain. This is especially important if we aim at designing supportive environments for discovery learning which must keep track of the behavior of the learner.

### *Hypothesis space*

Hypotheses about a domain take the form of a statement that a certain *relation* holds between two or more *variables* (van Joolingen & de Jong, 1991; Reimann, 1989). This implies that hypothesis space is spanned by two subspaces, a variable space and a relation space.

Variables can be directly observable in the domain, but learners can also make statements about more general concepts. An example can be found in physics of multiple particle systems. In this domain statements can be made about the position and velocity of an individual particle, e.g. about their size or relation to a specific other particle. Also, more general statements can be made about the position and velocity of *all* particles in the system, like how they depend on a force field. The center of mass may be introduced and we can also state hypotheses about its position and velocity. Moreover, statements may be made about position and velocity in general, like "the derivative of position is the velocity".

This example makes clear that hypotheses can be ordered according to a level of *generality*. For describing these different levels we introduce a

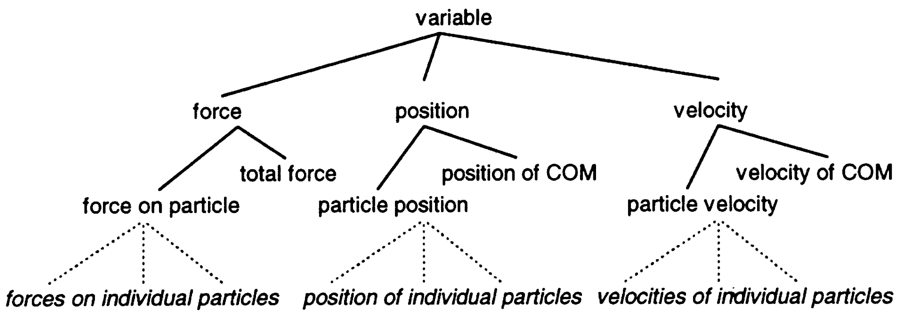


*Figure 1.* A part of the variable hierarchy for the multi-particle system that is used as an example in the text. In this figure, COM stands for Center of Mass, the weighed average of the positions of all particles.

generality hierarchy for variables. Variables that are higher in this hierarchy are more general than those which are placed lower. Hypotheses stated for general variables also apply for their children in the hierarchy. For instance, if a hypotheses is stated for the relation between position and velocity in general, then this also applies to the position and velocity of individual particles. In Figure 1, the variable hierarchy for the example just presented is depicted. For a more extensive argument on the structure of variable space, see van Joolingen (1995).

The variable hierarchy is introduced in order to describe the different levels of generality. However, this hierarchy can also play a role in distinguishing different types of variable structures, as was discussed above. In a domain in which all variables interact with all others as in Voltaville (Glaser et al., 1992) one expects that all variables have equal relative importance in understanding system behavior. If, moreover, the number of variables is not very large, classification of variables into a generality hierarchy is not meaningful, resulting in a ‘flat’ structure of variable space. Conversely, in a domain where the main task is to discover which of the given variables contribute to the behavior of the system, like Smithtown (Shute et al., 1991), a generality hierarchy can be a useful classification instrument. This is illustrated by a study by Simmons & Lunetta (1993) who found that successful subjects in a discovery task use more higher order concepts than unsuccessful subjects.

For relation space, different levels of *precision* exist. For example, for the relation between the position and velocity, one can state a fairly imprecise statement like: “if the velocity is positive, the position will increase”, but also a more precise statement like: “the derivative of the position is the velocity”. The difference between these two statements is the number of possible outcomes of experiments that would contradict the statement. No experiment in which the direction of change is consistent would contradict the first statement. The second statement would be contradicted by many more outcomes of experiments. A main distinction on the precision dimension is that between *qualitative* and *quantitative* relations. Qualitative statements about a domain are less precise than quantitative ones, but may be useful for understanding a domain. Plötzner, Spada, Stumpf & Opwis (1992; Plötzner & Spada, 1992) also introduce an ordering in precision for relations. They distinguish three levels: a qualitative relational, quantitative relational and quantitative numerical. Following this distinction in precision, we can present the concept of a relation hierarchy. Figure 2 presents an example of such a relation hierarchy. Child relations are more precise than their parents, and the ordering is such that if a relation is true between two variables then also its parent relations are true for these variables.

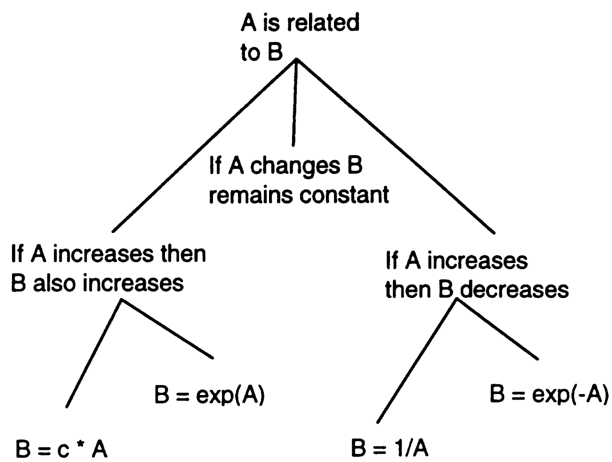


*Figure 2.* Example of a relation hierarchy.

The hierarchical structure of hypothesis space introduced here is well in line with the work by Collins & Michalski (1989), who assume that ". . . a large part of human knowledge is represented in 'dynamic hierarchies', that are always being updated, modified or extended" (p. 8). We adopt a similar view by representing hypothesis space as a set of hierarchical structures, type-hierarchies in terms of Collins & Michalski (1989). However, as it is presented here, these hierarchies are static, describing the whole of hypothesis space. The dynamic aspect of the search in hypothesis space will be discussed below.

*Experiment space*

Whereas the basic elements of hypothesis space are variables and relations, experiment space consists of *value-tuples*, sets of value assignments to variables. For example a specific experiment in the domain of particle physics can be described by [time=0; position of particle$_1$=10; velocity of particle$_1$=5]. Variables in these tuples are instances of variables in hypothesis space. Values in these assignments may be numeric or qualitative. The values in a value-tuple may be set by the learner or be generated by the simulation. An *experiment design* is defined by a value-tuple for which only the values manipulated by the learner are set, the remaining values can be retrieved by performing the experiment.

*Searching dual search space*

Searching dual search space implies searching both experiment and hypothesis space. Searching experiment space has two main components. First, the learner has to decide which variables to manipulate, i.e. of which variables the value will be changed. Also the learner has to decide which variables will contain the output of the experiment. Then, the learner should determine how to manipulate the variables, i.e. decide which value to assign to the variable. After running the experiment, the output variables have assumed values, and a new value-tuple is complete.

The rest of this section is dedicated to describing search in hypothesis space. Searching hypothesis space is a repeated process of generating hypotheses, by applying search operations, starting from existing hypotheses and assessing the merits of the resulting hypotheses to determine if further search is necessary and in which direction the search should continue.

In general, the target of discovery is not to find a single relation describing the whole domain but a number of relations between variables in the domain. Therefore, the goal state of hypothesis space search is a set of hypotheses, rather than a single one. During search, a learner maintains a set of candidate hypotheses. In constructing such a hypothesis set, learners need to search both variable space, i.e. identify variables to state hypotheses about, and relation space, selecting a relation to hold between two or more variables. The fact that multiple rules may be discovered requires that assessment of hypotheses applies to the complete set of hypotheses maintained by the learner, rather than to single hypotheses only. Based on related studies in the field of qualitative reasoning (Borbrow, 1984; Sime & Leitch, 1992; Spada, Stumpf & Opwis, 1989), we identify a number of aspects on which hypothesis sets can be assessed:

- correctness,

- precision,
- scope,
- range.

These aspects together determine the quality of a set of hypotheses in terms of its predictive power, being the number and quality of the predictions that can be generated using the hypotheses in the set and the number of situations about which predictions can be generated.

The *correctness* of a hypothesis set is determined by the predictions that can be generated using the hypotheses within the set. If instances in experiment space contradict the prediction, the hypothesis set is incorrect.

The *precision* of a set of hypotheses is defined by the precision of the relations therein, as described above.

The *scope* of a hypothesis set is the number of variables about which predictions can be generated. The goal of a full discovery will be to identify all relevant variables and find relations capable of predicting the behavior of the values of all these variables. Scope is related to the concept of generality of a hypothesis. A more general hypothesis has a larger scope.

The *range* determines for which parts of experiment space valid predictions can be made. The range determines the values of the variables for which the hypothesis is assumed to be true, whereas the scope defines for which variables the hypothesis is valid. So, a hypothesis on position in general has a larger scope than a hypothesis on the position of a specific particle. A hypothesis that is valid only for positive velocities, has a smaller range than a hypothesis valid for all, positive and negative, velocities.

The assessment of a hypothesis set may motivate the need for generating new hypotheses. This can be done by applying search operations in hypothesis space. Using the structures of variable and relation space that were presented in the previous sections, a list of possible search operations can be derived. In making this list comprehensive, we should consider that a learner needs to maintain multiple hypotheses in order to arrive at a set of rules. Therefore, also search operations that affect the set of hypotheses, e.g., by adding or deleting hypotheses will be considered. This means that there are three main categories of possible search operations in hypothesis space:

- search operations in variable space;
- search operations in relation space;
- search operations that change the hypothesis set as a whole.

Some search operations apply to an individual hypothesis, i.e. replace one hypothesis in the hypothesis set by another. Other search operations act on the set as a whole in the sense that they add or remove hypotheses. Below an inventory of all possible search operations is given.

Variable space search operations can be of the following kinds:

- *Generalization of a hypothesis.* Generalization takes place by choosing variables which are higher in the variable hierarchy. For example, the hypothesis “the derivative of the particle position is the particle velocity” can be generalized to “the derivative of position is velocity”, implying that the relation is now assumed to hold for all instances of the named variables. This operation extends the scope of the hypothesis;
- *Specialization of a hypothesis.* This is done by choosing variables lower in the variable hierarchy and as such it is the reverse of generalization;
- *Change of variable in a hypothesis.* A hypothesis can also be changed by replacing a variable by another one which is neither its ancestor nor its descendant. For example, the hypothesis: “the derivative of the particle position is the particle velocity” can be changed in “the derivative of the position of the Center of Mass is the velocity of the Center of Mass”.

Relation space search operations can be of the following kinds:

- *Specification of a hypothesis* by choosing a more precise relation, for example going from a monotonic to a linear relation: “if A increases then B increases” becomes “if A doubles then B doubles”;
- *Abstraction of a hypothesis.* The reverse of the above: moving from a precise relation to one that is less precise;
- *Adding a characteristic* by specifying a second relation on the same variables which holds concurrently, e.g., when a monotonic relation (“if A increases, B also increases”) already has been specified, a second relation can be added to the hypothesis set, which specifies that the relation is asymptotic (“if A keeps increasing, B goes to a constant value”) as well, yielding that the resulting relation will be a monotonic increasing asymptotic function;
- *Deleting a characteristic* is deleting a relation from the hypothesis set under the condition that at least one relation between the same variables remains in the set. This is the opposite of the previous search operation;
- *Specification of parameters* in a hypothesis. Some relations take one or more extra parameters, e.g., the value of the constant value in the asymptotic relation in the example above. This relation can be further specified by specifying: “if A keeps increasing B approaches 1”;
- *Restriction of a hypothesis.* The range of a relation may be restricted by adding a condition, or by further constraining an existing condition. For example: “the volume of a constant quantity of water decreases with decreasing temperature” may be restricted to: “the volume of a constant quantity of water decreases with decreasing temperature as long as the temperature is above 4 °C”;
- *Expansion of a hypothesis.* The opposite of the above: removing or changing a condition such that the range of the relation increases;

- *Change of relation in a hypothesis.* A choice for a relation which is neither a more precise nor a less precise version of the old relation.

Possible operations on the hypothesis set are:

- *Adding a hypothesis to the hypothesis set.* A new hypothesis about variables not investigated before, can be added to the hypothesis set. This is different from adding a characteristic to a hypothesis, because the new hypothesis applies to variables for which no hypothesis was stated before, with the effect that the scope of the hypothesis set increases, which is not the case for adding characteristics;
- *Removing a hypothesis from the hypothesis set.* This can occur, for instance, because a hypothesis is judged to be false or irrelevant;
- *Splitting a hypothesis.* A hypothesis can be split in two, by introducing an intermediate variable. For example, the relations "If A increases then C also increases" ($M^+$(A,C)) can be split in $M^+$(A,B) and $M^+$(B,C). Splitting a hypothesis is used to emphasize the role of an intermediate variable in a process. The result of splitting is adding two relations to the hypothesis set;
- *Combination of hypotheses.* The reverse of the above: from relations between A and B and between B and C, a relation between A and C may be inferred, and added to the hypothesis set. This search operation, as well as the previous one represents the possibility of using logical inferences in searching hypothesis space;

The changes to the hypothesis set enumerated above provide a classification of search operations in hypothesis space. Search operations of one of the types listed have consequences for the predictive power and conceptual complexity of the hypothesis set. In Table 1 this has been elaborated for the four aspects of hypothesis sets mentioned at the start of this section. The cells of this table are filled with indications of consequences of hypothesis space search operations for the various aspects of predictive power of the learners' hypothesis set. In this table "$+$" means an increment, "$-$" decrement, "o" means no change, and "$\pm$" means a possible change in any direction, "o/$+$" and "o/$-$" mean that, depending on circumstances there may be a change in the direction indicated.

Some examples of entries in Table 1 are:

- if the relation: "the derivative of the position of particle 1 is the velocity of particle 1" is generalised to "the derivative of the position is the velocity", then the scope will increase, but the correctness *may* decrease, because the more general relation may include cases in which the relation is incorrect (which is not the case in this specific example);
- if the relation "if A doubles, then B doubles" is abstracted to "if A increases then B increases", then the precision decreases by definition.

*Table 1*. Overview of possible search operations in hypothesis space and their consequences for the hypothesis set.

| | Correctness | Precision | Scope | Range |
|---|---|---|---|---|
| Variable space search operations | | | | |
| Generalization | o/− | o | + | o |
| Specialization | o/+ | o | − | o |
| Change of variable | ± | o | ± | ± |
| Relation space search operations | | | | |
| Specification | o/− | + | o | o |
| Abstraction | o/+ | − | o | o |
| Specification of parameters | o/− | + | o | o |
| Addition of characteristics | o/− | o/+ | o | o |
| Deletion of characteristics | o/+ | o/− | o | o |
| Restriction | o/+ | o | o | − |
| Expansion | o/− | o | o | + |
| Change of relation | ± | ± | o | ± |
| Hypothesis set operations | | | | |
| Addition of hypothesis | o/− | o | + | o |
| Deletion of hypothesis | o/+ | o | − | o |
| Splitting of hypothesis | o | o | o/+ | o |
| Combination of hypotheses | o | o | o/− | o |

However, the correctness may increase, since the latter statement may be true, even if the first is not;

- if a new hypothesis is added to the hypothesis set, then its scope will increase. The correctness may decrease, because the new hypothesis may be incorrect.

In this manner it is possible to check the entries for the whole of Table 1. It is important to note that an increment in precision, scope or range, may mean that the hypothesis set is no longer correct: these operations increase the information content of the hypothesis set, and therefore may add incorrect information. The classification of search operations presented here allows to describe in detail the search processes in hypothesis space.

### *Decomposing hypothesis space as knowledge representation*

During the discovery process, the domain knowledge of the learner changes. A complete description of discovery learning should therefore represent this changing knowledge. Part of the learner's domain knowledge is represented by the current hypotheses set, but there is more to say about the learner's knowledge. First, there is the learner's knowledge about the existence of variables and relations, which may exist independent of the actual statement of hypotheses using these variables or relations. This knowledge defines the part of hypothesis space that the learner can search directly. Furthermore, also knowledge about hypotheses that have been *rejected* is useful. Knowledge about what is *not* true can be a good lead in the search for what *is* true.

In our approach the learner's domain knowledge will be represented by a configuration of subspaces of hypothesis space. In order to distinguish the complete hypothesis space from its subspaces we call it:

- The *universal hypothesis space*, containing all possible hypotheses about a certain domain, independent of their truth value, plausibility, learner's judgment or whatever attribute can be found.

Two subspaces represent the learner's knowledge about the domain:

- The *learner hypothesis space*, spanned by the variables and relations the learner knows of, still independent of the learner's judgment. This is the space that the learner is able to search directly. In order to go outside this space, the learner must acquire knowledge about new relations or variables;
- The *effective learner search space*, the space of hypotheses that the learner decides to be worthwhile for testing. This is a subspace of the learner hypothesis space, since learners may decide not to explore specific parts of their learner hypothesis space. In a study by Schauble, Klopfer, & Raghavan (1991) it was for example found that subjects stopped working when they thought they had be experimenting enough. In terms of Klahr, Fay & Dunbar (1993), the effective learner search space is the set of *plausible* hypotheses. The learner tests the hypotheses in this space, and, after testing, marks them as being true, false, or unknown.

The aim of discovery is to bring the learner's knowledge close to the 'true' description of the domain. Here we distinguish:

- The *space of true hypotheses*, containing all true hypotheses that describe the domain;
- The *target conceptual model*, a subset of the space of true hypotheses. From the hypotheses in this model, all hypotheses in the space of true hypotheses can be derived. At the end of a successful discovery process, we expect the learner to have found a set of relations equivalent to the target conceptual model, in the sense that they imply the same set of true hypotheses.

In Figure 3 various spaces are depicted. For reasons of clarity, the space of true hypotheses is not drawn in this figure. The configuration depicted in just an example configuration. For example, the target conceptual model

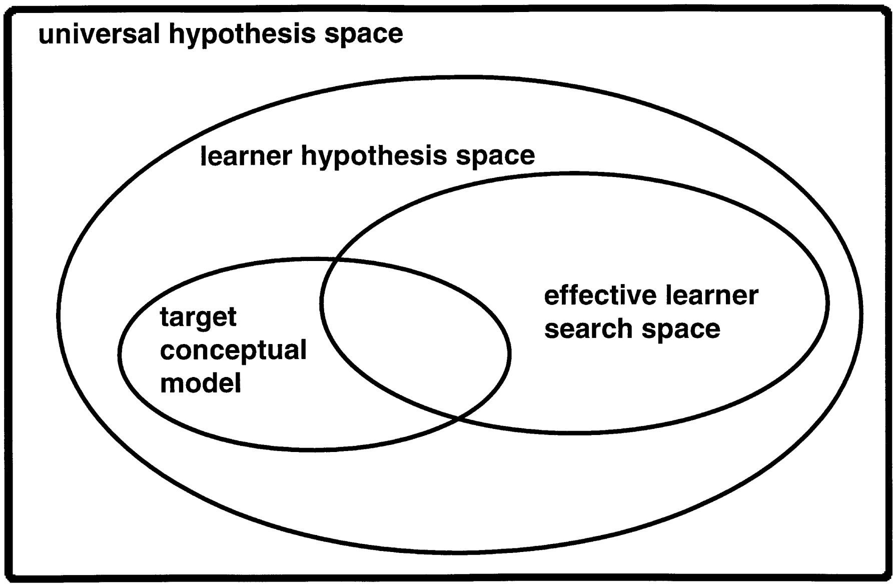


*Figure 3.* The different regions in hypothesis space representing the knowledge of the learner about the domain.

is not necessarily a subspace of the learner hypothesis space. Moreover, the configuration is not static, but normally will change during a session. The effective learner search space, for instance, can be enlarged, by adding new hypotheses.

In the next section, an empirical study is presented investigating subjects' performance of selected search operations in hypothesis space. This study intended to find evidence for the occurrence of search processes in hypothesis space and to obtain insight in the conditions under which people apply these search operations.

## Empirical study

The study presented here was designed to obtain more detailed insight in the search processes of subjects through hypothesis space and the selection of nodes in experiment space while discovering a complex domain. The study focused on three types of search operations in hypothesis space: specification, restriction, and generalization. This selection was made because these three search operations are of central importance in searching hypothesis space, and because in earlier studies (van Joolingen & de Jong, 1991; 1993) it was found that subjects often were unsuccessful in finding sufficiently precise and general relations. A possible explanation for this finding is that subjects lack the possibility to apply one or more of these search operations. Two reasons may underlie this problem: either the subjects know how to perform the search operation, but see no reason to do so, or they just do not know the operation at all. Therefore, the central question of this study is: “Do subjects have the search operations of specification, restriction and generalization available and, if so, under which conditions are these operations applied?” Furthermore, the relation between experiment space and hypothesis space was investigated, by studying the reasons for selecting experiments. According to the model presented above, the difficulty of selecting variables for experimentation is dependent of the generality of the hypothesis under investigation: the more general a variable in hypothesis space, the more interpretation steps it takes to instantiate it with a variable in experiment space.

Subjects in our study were confronted with a computer simulation of a complex domain: error analysis in chemical titration experiments. In the method section below this domain will be described in more detail. In the discovery process we specifically examined what we have called *choice moments*, which are the moments on which a learner decides whether to perform an experiment or to state an hypothesis, and, more importantly, on the specific experiment or hypothesis to be performed or stated. In terms of our theoretical model, at the choice moments a learner decides on the execution of a hypothesis or experiment space search operation. At each choice moment the interesting data are the search operation chosen by the subject, the actual performance of the search operation, and the reasons for making the choice.

The choice moments that subjects encountered were experimentally controlled by introducing a simulated learner of which subjects could observe the actions in the simulation. This simulated learner arrives at several choice moments, at which one or more of the search operations that we are interested in can be appropriate. The subjects were asked for comments on these choice moments: just before the simulated learner performed a search operation they were asked what they would do in the situation they were observing; just after the choice moment they were asked to comment on the search operation performed by the simulated learner. Especially, they were asked if they recognized the operation and if they could imagine performing such a search operation themselves. If subjects stated that they could imagine performing the search operation themselves in the presented situation we say that they *approved* of this search operation. The first reason for controlling the choice moments and not allowing for free discovery is to confront all subjects with the same choice moments which makes a comparison between subjects possible. A second reason is that we also wanted to confront subjects with choice moments that they would possibly not encounter in a free discovery situation

*Table 2.* Overview of anticipated subject responses on choice moments. Indicated are the responses before (chosen) and after (recognized and approved) the subject is confronted with a search operation performed by simulated learner, depending on the subject's knowledge of and reason for performing a search operation.

| | Subject responses | | |
|---|---|---|---|
| Knowledge of search operation | Chosen before confrontation | Recognized | Approved |
| Not known | No | No | – |
| Known, no reason | No | Yes | Yes or no |
| Known, positive reason | Yes | Yes | Yes |
| Known, negative reason | No | Yes | No |

and also we wanted to obtain information about the subjects' attitude to search operations they would themselves not choose spontaneously.

In this experimental session, the subjects themselves were *not* engaged in discovery learning themselves, they were asked to comment on the discovery process of the simulated learner. Such a setting may of course be used as a learning method, but this was not our primary goal. Our main interest was to confront all our subjects with the same set of choice moments.

We can now anticipate the responses of subjects to the several types of search operations depending on their knowledge. A search operation that a subject does not know of will not be chosen by this subject before he or she sees the performance of a search operation by the simulated learner. Also, this operation will not be recognized as a valid operation after the subject has seen the simulated learner performing the operation. A known operation for which a subject sees no specific reason will also not be chosen spontaneously, but it will be recognized and possibly, but not necessarily, approved of. A known operation for which a subject sees a negative reason will not be chosen before the subject sees the simulated learner perform a search operation but it will be recognized and not approved of. Finally, if a subject sees a positive reason for performing a search operation that he or she knows of, the search operation will be chosen, recognized and approved of. The reasons given by subjects for their choices assist in the interpretation of their responses. In Table 2 an overview is given of expected subject responses.

## *Method*

### *Domain*

The domain used in the current study is error analysis in chemistry. In this domain the relations between the various kinds of error occurring in chemical experimentation are described. As a central example, we chose a titration experiment to determine the concentration of Hydrochloric Acid (HCl) in water. Such titration experiments are very common in chemistry, which allows us to emphasize the aspects of error analysis in a situation familiar to the subjects.

In a chemical experiment different types of error occur, systematic and random errors. 4SEE concentrates on random errors, which are due to the limitations of the measuring equipment used. Measuring results can be used in calculations, meaning that the errors in the measuring result propagate to the final result. Using statistics, the relations between errors, expressed as standard deviations of a distribution of measuring results, can be calculated. However, chemists, and other experimental scientists, often do not use the formulas for calculating the final error, but estimate this, using qualitative and semi-quantitative relations, like "if some quantity doubles, the relative error in its measurement divides by two", or "if two errors are combined in a calculation and differ by at least one order of magnitude, then it is safe to ignore the smaller error". The aim of the discovery task we investigated was to find a number of these relations in a simulation of a titration experiment.

*The discovery environment*

The experiment was conducted with 4SEE (Statistics Simulation System as a Supportive Exploratory Environment), a simulation environment on error analysis in chemistry. One of the features of 4SEE is a hypothesis scratchpad, a dedicated editor for stating hypotheses, similar to the hypothesis menu used in *Smithtown* (Shute et al., 1989). The scratchpad contains relations and variables, which can easily be combined to form complete hypotheses (see Figure 4). Also conditions can be added to hypotheses, corresponding to the search operation of restriction. The scratchpad stores hypotheses created in a separate list, which can be inspected at any time by the learner. Hypotheses on this list can be marked as being currently tested and can be assigned a truth value, from the set "true", "false", and "unknown".

Originally, the scratchpad was designed as a supportive instrument for discovery learning (van Joolingen & de Jong, 1991). It offers insight into variable and relation space by showing their elements. In the current study it was used to show the search trace of a simulated learner through hypothesis space. In the experiment, the main task of the subjects was not to use 4SEE as a learning environment but to observe and comment on the actions of the simulated learner. This learner[2] was created by recording a session of an interaction with 4SEE, which could be replayed at a pace of the subject's preference, controlled by the keyboard. The simulated learner received two assignments while exploring the simulation. Both of these assignments were investigation assignments (de Jong et al., 1994): "Investigate the relation

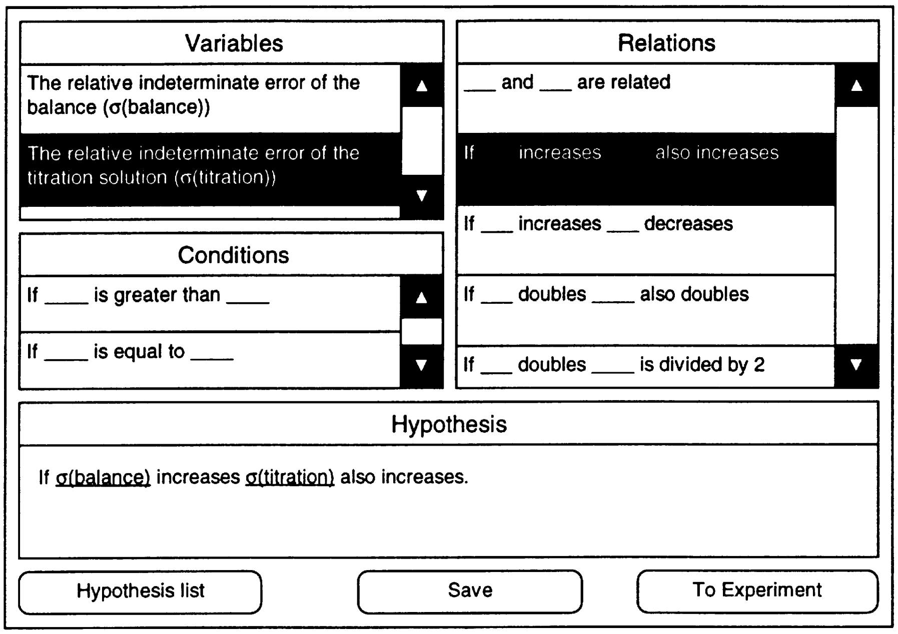


*Figure 4.* The hypothesis scratchpad as present in 4SEE. A user of this scratchpad can combine the variables and relations in various lists to form a complete hypothesis, which is shown in the bottom window.

between . . . and . . . ”, where the slots had been filled in with names of variables. The assignments were stated at different generality levels. The first assignment contained instance variables, at a low level of generality: the amount of primary standard and the total calculated error in the concentration of the titration solution, the second one involved general variables: the partial error in a calculation and the total error (see Figure 5). While working on an assignment the hypothesis scratchpad contained the variables appearing in the assignment. In the case of the second assignment both the general variables and their instances were included in the variable list on the scratchpad. In the recorded session, the simulated learner does not necessarily make the best choice for a search operation. Also, in many cases more than one search operation can be appropriate.

The simulated learner starts the first assignment by stating an (incorrect) hypothesis and performing an experiment. On the basis of the result of this experiment the first hypothesis is rejected and replaced by another, which is correct. After a second experiment this hypothesis is accepted (by marking it as “true” on the hypothesis scratchpad) and specified to a quantitative relational level (in terms of Plötzner et al., 1992). After some more experiments,

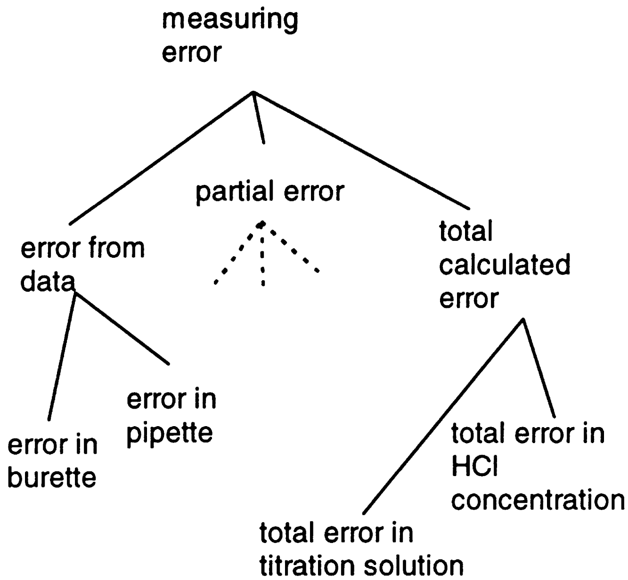


*Figure 5*. Part of the variable hierarchy of the domain used in the empmirical study. The 'error from data' occurs due to the limitations of the measuring device and is calculated from data of repeated measurements. Partial errors take part in a calculation, whereas the total calculated error is the error in the result of such a calculation.

the second hypothesis, resulting from the specification, is accepted and the first assignment is completed.

The simulated learner starts working on the second assignment by stating a hypothesis for instances of the error types involved, the error of the pipette and the error in the final result. After one experiment the hypothesis is specified to a qualitative relational level and another experiment is performed. The hypothesis is accepted and generalized to the generic variable types occurring in the assignment. Then the simulated learner performs an experiment with another instance of the partial error and finds that its results contradict the general hypothesis. Another hypothesis is stated but not tested by doing an experiment since it is in contradiction with earlier results (these results are shown to the subject). The general hypothesis is then specialized (i.e., made less general) for the instance involved and for another instance, which is also tested by performing an experiment. Then the simulated learner states a new hypothesis: a restriction of the general case, capturing the general rule, followed by a second restriction for complementary cases of the first restriction. Finally, the first of these hypotheses is further specified to a quantitative relational level.

At nineteen choice moments in the recorded session, subjects were interrogated about their thoughts. Eight of the moments of interrogation were located at a statement of a new hypothesis by the simulated learner, six were at design and performance of an experiment and five were at moments the simulated learner was about to draw a conclusion from an experiment (data analysis). Table 3 provides an overview of the choice moments for which subjects were asked to state what they would do and to judge the search operation in the recorded session. For hypotheses, the type of search operation is given in the table, for experiments and data analysis moments the generality of the last hypothesis stated (general or instances) is included. For instance, at choice moment 5, the simulated learner moved from the hypothesis: "As the amount of substance increases, the total error also increases" to: "As the amount of substance doubles, the total error also doubles"; at choice moment 8, an experiment was designed to test a hypothesis on the relation between the error in a single measurement instrument, with the total error, at choice moment 12, the same was done to test a more general hypothesis.

*Subjects*

Subjects were 22 second year students of chemistry at Eindhoven University of Technology. They had received a formal introduction in error analysis in their first year of study. They participated in this study on a voluntary basis and were paid for their participation.

*Procedure*

After a short introduction to the experiment, subjects were introduced to 4SEE. They were allowed to explore the simulation for 30 minutes, in order to familiarize themselves with the interface. During this period, they were given no specific task, except for trying out the various elements of the user interface, including the hypothesis scratchpad. After this period the simulation program was switched to replay mode and the prepared recording of the simulated learner was shown to the subject. During replay, the subject could inspect on paper the assignments given to the simulated learner and the variables and relations which were present on the hypothesis scratchpad. Before the selected choice moments in the recording, it was announced which type of action (i.e., "state a hypothesis", "perform an experiment" or "draw a conclusion") the simulated learner was going to perform, without revealing the content of this action. The type of action was given to the students in order to get a reply directed at the situation presented and because the sequence of events in the recorded protocol already pointed at the action the simulated learner was about to perform. Moreover, we wanted to be able to present on screen the information the subject needed to make a sensible choice. For example, before stating a new hypothesis, former hypotheses, if present, were

*Table 3.* The nature of the choice moments for which the subjects were asked to provide comments on the search operations performed by the simulated learner. In this table, for moments where the simulated learner states a new hypothesis, the search operation performed by the simulated learner is specified. For experiment design and data analysis choice moments, the generality of the variables in the last-stated hypothesis is indicated (general or instance). The horizontal line marks the transition to the second assignment given to the simulated learner.

| Choice moment | Statement of hypothesis | Experiment design | Data analysis |
|---|---|---|---|
| 1 | | Instances | |
| 2 | | | Instances |
| 3 | | Instances | |
| 4 | | | Instances |
| 5 | Specification to quantitative relational level | | |
| 6 | | Instances | |
| 7 | | | Instances |
| 8 | | Instances | |
| 9 | Specification to qualitative relational level | | |
| 10 | | | General |
| 11 | Generalization | | |
| 12 | | General | |
| 13 | Change relation (at general level) | | |
| 14 | Specialization (move to instance level) | | |
| 15 | | Instances | |
| 16 | | | Instances |
| 17 | Restriction | | |
| 18 | Restriction | | |
| 19 | Specification to quantitative relational level | | |

presented on screen, before data analysis, relevant data was presented. The subject was asked to say what he or she would do in this situation and why. The possibility that the subject would choose an operation of a different type than was announced for the simulated learner was explicitly included in the wording of the question. After each choice moment, when the action of the simulated learner was displayed, the subject was asked to comment on the specific operation performed by the simulated learner and to state whether he or she could imagine performing the same search operation. Subjects' answers were tape recorded.

*Data*

Data consisted of the transcribed recordings of the subjects' answers to the questions posed during the sessions. Due to a technical error, for one subject the tape recording was of such a quality that transcription was impossible. This subject was removed from the analysis. For subjects' responses before each choice moment it was assessed if a subject chose a search operation and if so, which operation was chosen. If the subject chose a search operation in hypothesis space, it was assessed whether the subject executed this operation by actually stating a new hypothesis. Subjects' responses on seeing the performance of the simulated learner were used to determine if the subject recognized the search operation and if the subject approved of the search operation, which was defined as a statement that the subject could imagine performing the same search operation him or herself. Experiments designed by subjects' were assessed on the variables mentioned and the level of precision on which the experiment was described. Data analysis by subjects was assessed on the correctness of the subjects' response.

*Results*

We start this result section by presenting a general overview of the choices and actual operations that subjects made at the eight choice moments where the simulated learner made a move in hypothesis space. Then we will zoom in at the three search operations that form the focus of this study: specification, restriction, and generalization, and we will analyze subjects' use and reasons for applying these operations, to see if subjects have these operations at their disposal and under which conditions they are applied. Finally, we will analyze subjects' moves in experiment space and examine the relation between hypothesis space and experiment space operations.

*Subjects behavior at hypothesis choice moments*

In Table 4 an overview of the search operations chosen by the subjects is given for each of the eight choice moments at which the simulated learner made a hypothesis space search operation. In this table numbers that are underlined indicate the search operation performed by the simulated learner.[3] The table shows that when subjects at these choice moments say that they want to create a new hypotheses they more often choose for change variable, specification, and restriction. Generalization and abstraction are rarely chosen, and specialization and change relation form a middle group. In the table we can also see that in 36% of the cases (no new hypothesis, no change, and not given) the subjects did not follow the search operation type ("state a hypothesis") that was announced for the simulated learner. This seems particularly the case for choice moment 5 and choice moment 19. At both of these

*Table 4.* Number of times a specific operation was chosen by subjects for the different choice moments. The search operations performed by the simulated learner at a specific choice moment are in italic type.

| | Choice moment | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Operation type chosen | 5 | 9 | 11 | 13 | 14 | 17 | 18 | 19 | Total |
| Specialization | 0 | 0 | 0 | 2 | *6* | 2 | 0 | 2 | 12 |
| Generalization | 0 | 0 | *0* | 0 | 1 | 1 | 0 | 0 | 2 |
| Change variable | 3 | 5 | 3 | 5 | 1 | 6 | 1 | 3 | 27 |
| Globalization | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Specification | *2* | *13* | 12 | 1 | 0 | 0 | 2 | *1* | 31 |
| Restriction | 0 | 0 | 0 | 1 | 6 | *1* | *9* | 3 | 20 |
| Change relation | 1 | 0 | 1 | *5* | 3 | 2 | 0 | 0 | 12 |
| No new hypothesis | 9 | 1 | 5 | 6 | 1 | 5 | 8 | 7 | 42 |
| No change | 3 | 1 | 0 | 0 | 0 | 0 | 0 | 3 | 7 |
| Not given | 1 | 1 | 0 | 1 | 3 | 4 | 1 | 2 | 13 |
| Total | 21 | 21 | 21 | 21 | 21 | 21 | 21 | 21 | 168 |

choice moments, the current hypothesis of the simulated learner was correct and at a medium level of precision. At these moments specification was one of the appropriate search operations (and in both cases indeed performed by the simulated learner). A close analysis of the reasons for choosing and performing operations is presented later on for specification, restriction, and generalization.

The choice for a specific search operation did not necessarily mean that the subject actually performed that search operation. In a number of cases subjects mentioned that they would perform a search operation without stating a resulting hypothesis. Table 5 displays the number of times that subjects actually performed the search operations they indicated for the 8 choice moments on which the simulated learner performed a search operation in hypothesis space. The moments for which subjects indicated that they did not want to state a new hypothesis (the categories: no new hypothesis and not given in Table 4) were not included in this table and the search processes that were rarely chosen (generalization and abstraction and no change (i.e. state the same hypothesis a second time) were aggregated in the category other. The actual performance of search operation by subjects is significantly related to the search operation type ($\chi^2_{(5,N=113)}$ = 24.94, $p < 0.001$).

*Table 5*. Number of times that, following the choice for a search operation, this operation was actually performed, meaning that a new hypothesis was created.

| | Operation chosen by subject | | | | | | |
|---|---|---|---|---|---|---|---|
| Operation peformed? | Specialization | Change variable | Specification | Restriction | Change relation | Other | Total |
| No | 9 | 23 | 11 | 13 | 2 | 7 | 65 |
| Yes | 3 | 4 | 20 | 7 | 10 | 4 | 48 |
| Total | 12 | 27 | 31 | 20 | 12 | 11 | 113 |

*Specification, restriction, and generalization*

Table 6 contains an overview of the number of times the subjects chose, recognized and approved of the hypothesis space search operations of specification, restriction and generalization. The number of times that an operation can be recognized and approved or disapproved of is, of course, limited by the number of occurrences of that search operation in the recorded session. A detailed overview of the results for specification, restriction and generalization will be given in the following subsections. In these sections, first the choices by subjects before they were confronted with the actual performance of the simulated learner are discussed, followed by a discussion of the subjects' comments on occurrences of the search operations of the simulated learner.

*Specification*

Out of a maximum of 168 times (21 subjects $\times$ 8 choice moments) 31 times a specification was chosen. On average, subjects chose a specification 1.48 times. By contrasting two typical choice moments we can illustrate some of the conditions that hold for the choice for specification. We compare choice moment 5 with moment 9. On moment 5, the current hypothesis is at a qualitative relational level ("if the amount of primary standard increases, the error in the titration solution decreases") on moment 9, the current hypothesis is less precise: "if the error in the burette (a measuring device) changes, so does the total error".

For choice moment 5, most subjects saw no need for further action (specification or otherwise) on the hypothesis. Only two of the 21 subjects chose a specification. Nine subjects indicated that they did not want to state a new hypothesis, four of which stated as a reason that the current hypothesis was sufficient; the other five gave no specific reason. The remaining subjects did not want to change the hypothesis; wanted to globalize it, i.e., decrease precision; or turned their attention to other variables (see Table 4). This can be illustrated by the following citations from the transcribed protocols:[4]

*Table 6.* Means and standard deviations for number of times that subjects chose, recognized and approved of specific hypothesis space search operations.

| | | | |
|---|---|---|---|
| **Specification (3 occurrences)** | | | |
| | Times chosen (max. 8) | Times recognized (max. 3) | Times approved (max. 3) |
| Mean | 1.48 | 2.95 | 2.14 |
| sd | 0.93 | 0.22 | 0.85 |
| **Restriction (2 occurrences)** | | | |
| | Times chosen (max. 8) | Times recognized (max. 2) | Times approved (max. 2) |
| Mean | 0.95 | 1.71 | 1.52 |
| sd | 0.97 | 0.46 | 0.60 |
| **Generalization (1 occurrence)** | | | |
| | Times chosen (max. 8) | Times recognized (max. 1) | Times approved (max. 1) |
| Mean | 0.10 | 0.90 | 0.85 |
| sd | 0.30 | 0.30 | 0.35 |

*I have not seen a new relation yet, what he*[5] *could look at. He has seen this relation now completely and drawn a conclusion. It seems to me that he cannot draw another conclusion about it (s7, 5b).*

*I think this relation has been demonstrated. I think he can stop (s9, 5b).*

This is different for choice moment 9, for which the current hypothesis is at a very low level of precision. At this moment thirteen subjects chose a specification, meaning that in this case subjects saw reason for performing a specification.

The fact that in 20 of 31 cases (see Table 5) in which subjects chose a specification they also perform it means that subjects are familiar with specification as a search operation.

The subjects comments on specification operations were collected at three choice moments. One (choice moment 5) was a specification from qualitative relational level to quantitative relational level; one (choice moment 9) was a specification from a very global level to qualitative relational level; finally, choice moment 19 was a specification from qualitative relational level to quantitative relational level with as an extra complicating factor that the full hypothesis was a conditional statement. Table 6 shows that subjects almost

*Table 7.* Frequency of reasons for approval or disapproval of hypothesis space search operations.

| Reason for (dis)approval | Approval of hypothesis: Approved | Disapproved | Not given | Total |
|---|---|---|---|---|
| Not enough evidence | 0 | 4 | 0 | 4 |
| False expected | 3 | 31 | 0 | 34 |
| True expected | 55 | 2 | 2 | 59 |
| Hypothesis is testable | 39 | 2 | 0 | 41 |
| Other/not given | 19 | 9 | 2 | 30 |
| Total | 116 | 48 | 4 | 168 |

always recognize specifications, and in 70% (2.14 out of 3) of the cases, subjects approve of specifications.

*Restriction*

In twenty cases subjects stated that they wanted to perform a restriction. These cases were concentrated near the second half of the second assignment where restriction is indeed appropriate. Near choice moment 14, the simulated learner has a number of cases in which sometimes an error contributes to the total error and sometimes it does not. In this situation looking for a condition to distinguish between these cases is a good thing to do. However, Table 5 shows that in thirteen of all twenty times a restriction was chosen, subjects did not actually perform this restriction. This is relatively often when compared to specification. For choice moment 14, where for the first time a substantial number of restrictions are given (six in total; see Table 4), only one subject stated a hypothesis resulting from a restriction. For choice moment 18, which was located after the first encounter with a restriction in the recorded protocol, nine times a restriction was chosen, and six of these choices were accompanied by a new hypothesis. The following citations from subject's answers illustrate the cases where subjects indicate the choice for a restriction without actually performing this operation:

> *Yes, then I would also state a new hypothesis, I think. In any case for every separate partial sigma,*[6] *er . . . , the relation with sigma total . . . So, not sigma partial in general but the separate ones, or something like that (s3, 14b).*

> *Er . . . I would also state a new hypothesis again, and hope that it is true, something like er . . . there are different partial sigma's (s16, 14b).*

These subjects indicate that they see the need to make a distinction between different partial errors but show no sign of knowing how. They do not state the idea that a conditional relation may do this.

The subjects' comments on restriction operations were collected at choice moments 17 and 18. Restrictions are recognized only slightly less frequent than specifications and also often approved of (see Table 6).

*Generalization*

The number of times subjects stated they would generalize a hypothesis was extremely low. Generalization was chosen only two times of all 168 times subjects were confronted with a hypothesis space search operation. However, the generalization of choice moment 11 was recognized by 19 of the 21 subjects, of whom 18 approved of the search operation. Subjects thus saw no reason to perform a generalization, but knew the operation when they saw one.

Related to generalization is a change in variable of the hypothesis if the new variable is a child variable of the same generic variable as the current one. In such a case the validity of hypothesis is extended to more variables of the same type, but without making a general statement. Table 4 shows that 26 times a variable change was chosen. In all of these cases, except the three for choice moment 5, subjects indicated they wanted to change variables of the same general type. In some cases the subjects also explicitly stated that they changed the variable in their hypothesis because they wanted to explore all instances of the general variable. For example:

> *The assignment is about partial sigma and total sigma, so then you could take for the pipette, or er . . . the mass balance another sigma partial and look what the influence is on sigma total (s16, 9b).*

In such a case the subject shows that he or she is aware of at least part of the variable hierarchy.

*Reasons for hypothesis space search operations*

In Table 7 an overview of the reasons of subjects to approve or disapprove of a hypothesis is given. These reasons were classified into a number of categories. The few cases in which the reasons are apparently inconsistent with the (dis)approval of a search operation, like disapproval of a hypothesis expected to be true, were actually generated by the subjects. Table 7 shows that the main reason why subjects have disapproved of hypotheses space search operations is that they expect a certain hypothesis to be false. Twenty-one of the 31 disapprovals of search operations with "false expected" were motivated by theoretical ideas, not by data observed earlier in experiments performed with the simulation. Also, the main reason for approving hypotheses is that the subject thinks the hypothesis is true. This can be an indication that in many cases subjects do not want to state a hypotheses before they are sure of its correctness. This can be a indication of what we can call a “fear of rejection” (see the conclusion section). However, in still a large number of cases, subjects approve of a hypothesis “because it is suitable to be tested”, a reason belonging to more open-minded discoverers.

*The relation between hypothesis and experiment space: design of experiments*

For the design of experiments, the relation between the level of generality of the hypothesis to be tested and the experiment designed by the subject were investigated. The experiment designed was assessed on two aspects: the variable selected to vary and the type of value change that was chosen. For the variables it was assessed whether the variable(s) selected to manipulate were appropriate to test the current hypothesis. The value change was assessed on a precision level, to be distinguished from the precision level of hypotheses. Three levels of precision were distinguished (in increasing order): specifying only that a variable will be changed, specifying only the direction of change, and finally, more precise ones, like multiplying a variable with a fixed factor. These levels are numbered 1 through 3 respectively.

In the context of the current study, subjects performed very good on selecting variables to do experiments with. In both assignments completed by the simulated learner subjects always picked correct instances of variables for performing experiments.

A relation was found between the assignment and the precision level of value changes. In Table 8 the experiment choice moments for which a subject actually performed a search operation in experiment space by specifying a value change are classified according to their precision level for each assignment separately. For the second assignment, that included general variables, subjects designed their experiments only at a low level of precision ($\chi^2{}_{(}2, N = 115) = 40.97$, $p < 0.001$). This might indicate that designing precise experiments is more difficult design when the hypothesis involved is at a high level of generality than for lower levels of generality. This supports our ideas on difficulty as they were formulated at the start of the current section.

We observed that, in some cases, subjects were led by the formulation of the last-stated hypothesis. They tried to copy the premise of the hypothesis as precisely as possible, which may not be needed to test the hypothesis. For example for the hypothesis: “if A increases, B also increases” it is not necessary to increase A to test the hypothesis, a decrement of A also suffices, but 23 (17%) times subjects stated explicitly that they wanted to follow the exact formulation of the hypothesis. Sometimes the simulated learner

*Table 8*. Frequency of usage of precision levels for experiments, as a function of the assignment.

| Assignment | Precision level 1 | 2 | 3 | Total |
|---|---|---|---|---|
| Instance | 28 | 15 | 18 | 61 |
| General | 54 | 0 | 0 | 54 |
| Total | 82 | 15 | 18 | 115 |

chose experiments which may test the hypothesis but did not follow its exact formulation. Mostly, subjects recognized that such experiments are just as good as the one they had chosen themselves, but 5 subjects stuck to their own experiment, and did not see the equivalence of the two experiments. This is illustrated by the following citations from a subject's protocol. The last-stated hypothesis is "if the amount of primary standard doubles, the error in the concentration of the titration solution divides by two":

> *He was, at his former experiment, at an amount of primary standard of 100 mg. And his hypothesis is that if you double the primary standard, thus 200 mg, then the sigma must become twice as small, so I would take the amount of primary standard 200 (s17, 6b).*

and after he had seen the experiment designed by the simulated learner (amount of primary standard becomes 50):

> *Er . . . So now I don't understand why he takes the mass (of the primary standard) twice as small, because the hypothesis stated that if you take the mass twice as large, then the sigma becomes twice as small. I wouldn't have done it this way. I would have taken 200 mg. (s17, 6a).*

## Conclusion

We started this article with presenting our model for describing discovery learning in complex domains. This model, following Klahr & Dunbar's (1988) SDDS theory, distinguishes a hypothesis and an experiment space but adds a structure to hypothesis space in terms of relation and variable spaces. On the basis of this structure we defined (classes of) search operations that can be performed in hypothesis space. Also a representation of the learner's domain knowledge in terms of a configuration of subspaces of hypothesis space was proposed. In this conclusion we will interpret the results of the empirical study from the viewpoint of our theoretical model. Then we will discuss some limitations and potential extensions to the current model. We will look at some constraints on the search in hypothesis space and explore our data in the context of these constraints. Finally, we will present some recommendations for the design of simulation learning environments.

*Subjects' discovery behavior in terms of the model*

In our study we specifically examined the search processes 'specification', 'restriction', and 'generalization', and we wanted to examine the background for applying or not applying these search operations. For the hypothesis space search operation of *specification*, we found that, especially for search operations starting from relations with a relatively high level of precision, subjects rarely spontaneously state they want to specify a hypothesis, but, once they have stated the intention to specify, they relatively often show that they know how to do this by formulating a new (more precise) hypothesis. For *restriction* the situation is different. In a number of cases subjects indicate that they see a need for restriction but they do not know how to do this, i.e., they are not able to find a conditional relation suitable to express their ideas. Especially, before subjects have seen a restriction in the recorded protocol they hardly give any indication that they know how to restrict a hypothesis. However, subjects relatively easily learned how to restrict hypotheses, that is, they extend their learner hypothesis space when confronted with an example of a restriction operation. This is illustrated by the fact that subjects recognized restrictions almost just as often as specifications and that a number of the subjects were able to apply restrictions once they had seen the simulated learner perform one.

In a study performed by van Joolingen and de Jong (1993) both precise and restricted hypotheses were rarely chosen. The current study shows that for the search operations leading to these hypotheses, different explanations are needed for their non-occurrence: specifications occur rarely because subjects see no reason to carry them out, restrictions occur rarely because subjects do not know how to perform them, since they seem to be unfamiliar with conditional hypotheses.

Subjects rarely stated spontaneously that they would perform a generalization. However, it was observed that in a number of cases subjects indicated that they wanted to extend the validity of a hypothesis to more instances of the same general variable. Apparently, they were not able to find the formulation of either the idea of generalization or the resulting general hypothesis or both. However, when confronted with a general hypothesis they showed little trouble in handling them: the general hypotheses were often recognized and the quality of conclusions drawn from experiments was not dependent upon the generality of the hypothesis that preceded it.

Our model predicts that experiments are harder to formulate when the corresponding hypothesis is formulated at a high level of generality, which also implies that these hypotheses are more difficult to test. In our data we found that subjects were capable of selecting the correct variables in experiment space, independent of the generality level of the variables of the accompanying hypothesis. So, in our particular case, these data do not support our idea of difficulty. Interestingly enough, however, we found that subjects in case of a general hypotheses specified experiments with a low level of precision as compared to a hypothesis with variables at a low level of generality.

*Limitations of and possible extensions to the model*

The model introduced in the first part of this article extends the applicability of the SDDS model by Klahr & Dunbar (1988). In the current section we discuss the generality of our model by putting forward some limitations of our current model.

One of the extensions we made to the SDDS model was the introduction of a variable hierarchy. This variable hierarchy allows for a description of the process of creating new variables that are generalizations of variables found directly in the domain. However, the model assumes that as a starting point, at least a number of variables has been created, of which it is possible to make observations in the domain. This first stage of discovery is important but seems quite resistant to being modeled in a satisfactory way. Most approaches in modeling scientific discovery take the same starting point as our model, e.g. Langley et al. (1983). In the case of discovery learning, problems with the creation of an initial variable space and to a lesser extent with the transformation of variable representations are not very important, since in general learners will encounter an artificial situation dedicated to furnish a successful discovery process, in which an initial variable space will already be present.

In the model we introduced the idea of a hypothesis set. The main aspect of a set of hypotheses that has been taken into account is its predictive power, i.e. the number and quality of the predictions that can be generated using the hypothesis set. This is, of course, the primary function of a set of hypotheses. A point that we did not address in the current paper is the internal structure of the hypothesis set. A main issue is the *parsimony* of the hypothesis set: the number of relations that is used to be able to generate predictions. Also, the *organization* of hypotheses is important: a set of hypotheses should not only contain relational information on variables but also reflect structural information of the domain. In van Joolingen (1995) a domain representation is described consisting of models which are small sets of relations. These models can be placed in two types of hierarchy, an is-a hierarchy and a part-of hierarchy. Using these models and the model hierarchies, we can extend our description of hypothesis space to account for an organization of sets of hypotheses which reflect structural information of the system that is being discovered.

The model as it is now is a descriptive framework of the discovery process in discovery learning environments. It allows for a precise description of search processes in hypothesis space. An important extension to the model will be a more detailed account of the conditions people use for selecting search operations. In the following sections we will identify a number of constraints on search in dual space search originating from several sources, and summarize some evidence found in our empirical study for these constraints. We will discuss domain specific prior knowledge, generic prior knowledge, internal and external goals, and personality factors (learner attributes).

*Domain specific prior knowledge*

In every learning setting, hence in discovery learning environments as well, a major factor influencing the learning processes is what the learner already knows (e.g., Ausubel et al., 1978). This accounts for both domain specific prior knowledge, as discussed here, and for generic knowledge, as will be discussed later. Alexander & Judy (1988), in a review, concluded that there is a strong relation between domain knowledge and problem solving approach. Learners with high prior domain knowledge in a certain domain also show a better problem solving approach. In similar vein, Klahr & Dunbar (1988) found a relation between the performance of search operations in dual search space and prior domain knowledge: learners who were able to state the correct hypothesis about the research question, generally applied a 'theorist' (model-driven) search strategy, whereas other students applied an 'experimenter' (data-driven) strategy.

In our model, the learner's prior knowledge is represented by the initial configuration of the learner hypothesis space and the effective learner search space and an initial marking ("true", "false" or "unknown") of the hypotheses within the latter space. Klahr & Dunbar's (1988) result can be restated by asserting that for hypotheses within the effective learner search space a theorist strategy will be applied, meaning that for these hypotheses experiment space will be searched for confirmation or disconfirmation. If the effective learner search space does not contain a suitable hypothesis (in the case that all hypotheses are rejected or that a more general or specific hypothesis is needed), the learner will need to extend the effective learner search space. In such a case an experimenter strategy will be applied to look for new hypotheses. Hence, we conjecture that search operations leading to a hypothesis outside the effective learner search space will not to occur before the learner has searched this space completely.

Support for this conjecture can be found in the Klahr & Dunbar study (1988) in which it was found that learners had difficulty in jumping from a hypothesis in the "counter frame" to one in the "selector frame". Both "frames" were subsets of hypothesis space. Within each set, the role of the parameter of the function investigated was similar, whereas it radically changed between the frames. Apparently, subjects initially restricted their search space to one "frame" and therefore could not easily perform a search operation which led them outside this frame.

This analysis stresses that the difficulty of performing search operations in hypothesis space is at least partly determined by the relative position of the new hypothesis to the effective learner search space and the learner hypothesis space. This does not mean that this completely determines the difficulty of search operations, it is possible that some search operations are intrinsically more difficult than others, for example, restriction requires the construction of a condition to a hypothesis which can be difficult for subjects unfamiliar to conditional reasoning.

*Generic prior knowledge*

Generic prior knowledge is (implicit and explicit) knowledge about the discovery process itself. It includes knowledge about the search operations themselves and how to perform them and also knowledge about search heuristics. Langley et al. (1987) describe a number of programs capable of discovering simple laws from data. Each of these programs includes one or more search heuristics for deciding which new hypothesis to put to the test. In our terms, these search heuristics are mostly concerned with hypothesis space search. Klahr et al. (1993) found a number of constraints in searching experiment space, including conservative search strategies, i.e. focusing on one dimension of change in designing experiments (see also Lavoie & Good, 1988; Tschirgi, 1980) and using the plausibility of the hypothesis to choose an experimental strategy. Heuristics can concentrate on the data available from the simulation, as do the heuristics included in the BACON family of programs, but also they may put emphasis on the structure of the hypothesis set, like heuristics aiming at a parsimonious description of a model.

Generic prior knowledge will influence the search through both search spaces, since this knowledge concerns the search process itself. So, unknown search operations will not be performed and search heuristics will drive the selection of specific search operations. In our empirical study, the influence of generic prior knowledge was shown by the difference between the performance of the two search operations 'specification' and 'restriction'. Subjects showed that they knew how to perform a specialization, and generally did so in the cases they saw need for it. In the cases they saw a need for restriction they couldn't perform it. Apparently, in the case of restriction they lacked the knowledge needed to perform the search operation. In experiment space search some subjects applied a heuristic of "hypothesis driven experiment design" by designing experiments which followed closely the premise of the hypothesis.

*Internal or external goals*

The discovery process can also be influenced by the goals the learners have. These goals may have an internal origin, for example the learner's ambition or curiosity, or may be externally imposed or generated, for example by assignments given by a tutor. Dunbar (1993) describes the dependence of the learner's search strategy on the learner's goals, for subjects working on a discovery environment for genetic mechanisms, with a main topic the control of some genes over others. He finds a crucial dependence between the two. A difference can be made between so-called find-hypothesis and find-evidence goals on the collection and interpretation of data. Furthermore, goals, once set, can block the discovery process. For instance, when learners had set a goal to find a specific kind of cause for a certain phenomenon, they were often unable to find causes of a different nature. In a study with secondary school children, Schauble, Klopfer & Raghavan (1991), found a difference between subjects applying an "engineering approach", aiming at an optimal behavior of the system investigated, and a "scientific approach" aiming at optimal understanding of the system. These approaches had crucial influence on the design of experiments and were susceptible to the setting of goals by a tutor.

In general, the results of Dunbar (1993) and Schauble, Klopfer & Raghavan (1991) would mean that search operations will be chosen by the learner in such a way that they lead to satisfaction of a goal, but do not go beyond that goal. Effectively, this means that the effective learner search space will not include the hypotheses that go beyond the learner's goals and experiment space will also be constrained as a function of the learner's goal.

In our study, the analysis of subject's reasons given for hypothesis space moves (or the non-performance of those) also showed the influence of the goals the learners set for themselves. At certain moments subjects were satisfied with the hypothesis that was stated and did not want to go to a higher level of precision, despite the fact that this was possible. The fact that subjects saw no need for specification clearly demonstrates the blocking effect that goals may have, as mentioned by Dunbar (1993). When learners had set a goal of finding a relation of certain precision, they would not search further, thereby effectively reducing the learner hypothesis space.

*Personality factors*

Finally, discovery learning may be influenced by a number of personality factors. In Goodyear, Njoo, Hijne & van Berkum (1991) a number of personality factors (learner attributes) is listed which may possibly influence discovery learning with simulations. Goodyear et al. distinguish intellectual attributes, academic motivation, cognitive style, and learning approaches as main categories of learner attributes. In the literature some phenomena have been observed which may be dependent of these factors. For example, in research with discovery environments, often the phenomenon of a confirmation bias has been observed (Wason, 1960; Gorman & Gorman, 1984; Gorman, Stafford & Gorman, 1987; Mynatt, Doherty & Tweney, 1977, 1978). This confirmation bias results in experiment designs aiming at confirmation of hypotheses, rather than disconfirmation with as a consequence that essential parts of experiment space remain unexplored. Van Joolingen & de Jong (1993) describe an analogous phenomenon for hypothesis space, a fear of rejection bias, resulting in a fear of stating hypotheses for which there is a relatively high chance of rejection. This results in hypotheses that are often so global that it is almost impossible to disconfirm them. For these hypotheses often no other than positive instances of the hypothesis can be found. These phenomena, confirmation bias and fear of rejection, may well be related to the personality factors indicated above.

In our study evidence was found for the occurrence of fear of rejection. In some cases learners disapproved of the hypothesis stated in the recorded session expecting that it would be false. In a substantial part of these cases this disapproval was based on (incorrect or partially incorrect) theoretical ideas, not on previous observed experimental results. These events indicate a fear-of-rejection based reasoning. However, the opposite of fear-of-rejection based reasoning, temporarily stating a hypothesis, motivated by the testability of the hypothesis was observed as well.

*Representing constraints in the model*

The constraints we identified in the preceding paragraphs can be represented in our extended dual search space model. We already mentioned the representation of prior domain knowledge in terms of an initial configuration of the subspaces of hypothesis space. The other constraints (generic prior knowledge, goals, and personality factors) can be represented in terms of conditional rules attached to search operations. These rules will determine if and when a certain search operation will be performed. In this sense, they act as constraints on the discovery process. Constraints can be subdivided in product constraints and generative constraints. A product constraint acts, as its name indicates, on the product of a search operation: a new hypothesis or hypothesis set. Generative constraints prevent search operations from being activated.

We can illustrate this by giving some examples:

- a search operation will only be performed if the learner knows how to perform it (generic prior knowledge, generative constraint);
- a search operation will only be performed if it does not lead to a hypothesis of a precision level beyond the learner's current goals (learning goals; product constraint);
- a search operation will only be performed by a learner with a fear of rejection bias, if sufficient evidence is available for the hypothesis before it is stated (personality factor; product constraint)

In order to be useful, constraints like these should of course be more specific and draw upon a representation of the learner (a learner model) containing information about the various aspects of the learner which are used in the rules. It is a challenging subject of research in this area to identify the rules determining the selection of search operations in hypothesis space.

The constraints representing the influence of prior knowledge, goals, and personality factors form the entry point for representing all kinds of social and psychological factors on the discovery process. For instance, goals can be influenced by motivational factors; fear of rejection may be related to anxiety (Leutner, 1993). It is beyond the scope of this article to enumerate all of these factors and their relation with discovery behavior, or even to make a proper start with this. We think, however, that our theory provides anchoring points for a description of the relation of these factors with discovery behavior. Further research should be directed at finding a limited set of constraints explaining much of observed learner behavior.

### *Recommendations for the design of discovery-based learning environments*

We conclude with some recommendations for the design of discovery-based learning environments. These learning environments embed a domain to be discovered in instructional support. Important examples of these types of environments are simulation-based learning environments, like the ones described in Shute & Glaser (1990), Reimann (1989), Schauble, Glaser, Raghavan & Reiner (1991), and de Jong et al. (1994). The extended dual search space model presented here provides means for predicting and measuring the effect of specific instructional support measures. Instructional measures can be modeled as affecting the attributes of the learner or other factors identified above. For example offering prior instruction and/or extra information during discovery may influence the learners prior knowledge, a tutor can set goals for the learner, motivating measures can reduce the effects of fear of rejection and specific tools offered to the learner, like a hypothesis scratchpad, can substitute the lack of knowledge of search operations. This means that instructional measures can explicitly be used in influencing search processes in hypothesis or experiment space. Also, detailed knowledge about the nature of the search processes and factors influencing them can lead to a more sophisticated design of instruction.

In the design of instructional measures for discovery learning environments, we should appreciate differences between learners, both at an individual level as well as on a group level. For example, in the study presented in this article, a difference was observed between the reasons for not choosing specifications and restrictions. The implication of this observed difference is that, for the group of learners observed, both search operation types require different types of support within a discovery environment. On the one hand, support directed at stimulating a learner to specify a hypothesis should concentrate on showing the learner the needs for specification, for example by asking to state a prediction needing a more precise hypothesis than the learner's current one. On the other hand, support for restriction should be more directed at showing learners how to restrict, e.g., by providing an example of a conditional relation.

At an individual level the model provides a facility to describe the knowledge of the learner. The representation of the learner's knowledge in terms of a configuration of search spaces may be made explicit in an internal representation of the learning environment. Using an instrument like the hypothesis scratchpad or, possibly, other means of retrieving information from the learner can provide the information necessary to build such a representation. Then this knowledge can be compared to a representation of the target model. In van Joolingen (1995), the system QMaPS is described, capable of performing such a comparison, yielding the difference between the learner's model and target model in terms of precision, scope and range. This information can be used to provide the learner with additional support to reduce this difference. An example of such support is to constrain the freedom of the learner in stating hypotheses and doing experiments, in order to draw attention to problematic areas of hypothesis and/or experiment space.

## Notes

1. Qin & Simon (1990) argue that the problem of finding variables to state hypotheses about, is not a real problem. They illustrate this point by the discovery of Kepler's third law, concerning the relation between the distance of planets from the sun and their period of revolution around the sun, where it seems that the identification of variables goes back to Aristotle, whereas the discovery of the law, i.e., the relation between the variables, was first made by Kepler. The problem with this viewpoint is that it assumes that the original concepts the discoverer used are identical to concepts that are still used in our days. However, in the days of the original discovery of a law, the meaning of concepts

used may well be different from their meaning in our days. In the case of Kepler, the concepts of space and time were not as developed as they are now, and, moreover, people in Kepler's age were less familiar than we to thinking in terms of a heliocentric view of the solar system.

2. To avoid confusion, in the current section, the word *learner* is reserved for the simulated learner performing the actions in the recorded session, whereas *subject* refers to the (live) subjects taking part in the experimental study presented here.
3. As indicated above, the simulated learner does not necessarily choose the best, if any, search operation available.
4. Citations from transcribed protocols were translated from Dutch. They are numbered according to subject (sN) choice moment number and statement before (b) or after (a) confrontation with the performance of the simulated learner. E.g., (s8, 5a) means: subject 8, choice moment 5, statement after the performance of the simulated learner had been shown to the subject.
5. "He" refers to the simulated learner. In the instruction given to the subjects, no reference was given to the sex of this non-existing person.
6. By "sigma" subjects mean error, or, more precisely, standard deviation. The term refers to the symbol used to represent this quantity.

## References


Alessi S.M. & Trollip, S.R. (1985). *Computer Based Instruction, Methods and Development*. Englewood Cliffs, NY: Prentice-Hall.

Alexander, P.A. & Judy, J.E. (1988). The interaction of domain-specific and strategic knowledge in academic performance. *Review of Educational Research* 58: 375–405.

Ausubel, D.P., Novak, J.D. & Hanesian, H. (1978). *Educational Psychology: A Cognitive View*. New York: Holt, Rinehart & Winston.

Borbrow, D.G. (1984). Qualitative reasoning about physical systems: An introduction. *Artificial Intelligence* 24: 1–5.

Böcker, H.D., Herczeg, J. & Herczeg, M. (1989). ELAB – An electronics laboratory, in D. Bierman, J. Breuker & J. Sandberg, eds., *Proceedings of the 4th International Conference on AI & Education* (pp. 15–25). Amsterdam: IOS.

Bruner, J.S. (1961). The act of discovery. *Harvard Educational Review* 31: 21–32.

Brown, J.S., Burton, R.R. & de Kleer, J. (1982). Pedagogical, natural language, and knowledge engineering techniques in SOPHIE I, II, and III, in D. Sleeman & J.S. Brown, eds., *Intelligent Tutoring Systems*. London: Academic Press.

Brown, J.S., Collins, A. & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher* 18: 32–42.

Collins, A. (1988). *Cognitive Apprenticeship and Instructional Technology* (Report No. 6899). Cambridge, MA: BBN Laboratories.

Collins, A., Brown, J.S. & Newman, S.E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing and mathematics, in L.B. Resnick, ed., *Knowing, Learning and Instruction: Essays in the Honor of Robert Glaser* (pp. 449–453). Hillsdale, NJ: Lawrence Erlbaum.

Collins, A. & Michalski R. (1989). The logic of plausible reasoning: A core theory. *Cognitive Science* 13: 1–49.

Duffy, T.M. & Jonassen, D.H. (1991, May). Constructivism: New implications for instructional technology? *Educational Technology*: 7–12.

Dunbar, K. (1993). Concept discovery in a scientific domain. *Cognitive Science* 17: 397–434.

Dunbar, K. & Klahr, D. (1989). Developmental differences in scientific discovery processes, in D. Klahr & K. Kotovsky, eds., *Complex Information Processing: The Impact of Herbert Simon* (pp. 109–145). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Friedler, Y., Nachmias, R. & Linn, M.C. (1990). Learning scientific reasoning skills in microcomputer-based laboratories. *Journal of Research in Science Teaching* 27: 173–191.

Glaser, R., Schauble, L., Raghavan, K. & Zeitz, C. (1992). Scientific reasoning across different domains, in E. de Corte, M. Linn, H. Mandl & L. Verschaffel, eds., *Computer-Based Learning Environments and Problem Solving* (NATO ASI series) (pp. 345–373). Berlin: Springer Verlag.

Goodyear, P., Njoo, M.K.H., Hijne, H. & Berkum, J.J.A. van (1991). Learning processes, learner attributes and simulations. *Education & Computing* 6: 263–304.

Gorman, M.E. & Gorman, M.E. (1984). A comparison of disconfirmatory, confirmatory and control strategies on Wason's 2-4-6 task. *Quarterly Journal of Experimental Psychology* 36A: 629–648.

Gorman, M.E., Stafford, A. & Gorman, M.E. (1987). Disconfirmation and dual hypotheses on a more difficult version of Wason's 2-4-6 task. *Quarterly Journal of Experimental Psychology* 39A: 1–28.

Greeno, J.G. & Simon, H.A. (1984). Problem solving and reasoning, in R.C. Atkinson, R.J. Herrnstein, G. Lindzey & R.D. Luce, eds., *Stevens' Handbook of Experimental Psychology; Vol. 2: Learning and Cognition* (pp. 589–673). New York: Wiley.

Hollan, J.D., Hutchins, E.L. & Weitzman, L. (1984). STEAMER: An interactive inspectable simulation-based training system. *AI Magazine* 5: 15–27.

Jonassen, D.H. (1991). Objectivism versus constructivism: Do we need a new philosophical paradigm? *Educational Technology Research & Development* 39: 5–14.

Jong, T. de (1991). Learning and instruction with computer simulations. *Education & Computing* 6: 217–230.

Jong, T. de & Joolingen, W.R. van (in preparation). *Discovery Learning with Computer Simulations*.

Jong, T. de, Joolingen, W.R. van, Scott, D., Hoog, R. de, Lapied, L. & Valent, R. (1994). SMISLE: System for multimedia integrated simulation learning environments, in T. de Jong & L. Sarti, eds., *Design and Production of Multimedia and Simulation Based Training Material* (pp. 133–165). Dordrecht: Kluwer Academic Publishers.

Jong, T. de & Joolingen, W.R. van (1995). The SMISLE environment: Learning with and design of integrated simulation learning environments, in P. Held & W.F. Kugemann, eds., *Telematics for Education and Training* (pp. 173–186). Amsterdam: IOS.

Joolingen, W.R. van (1995). QMaPS: Qualitative reasoning for intelligent simulation learning environments. *Journal of Artificial Intelligence and Education* 6: 67–89.

Joolingen, W.R., van & Jong, T. de (1991). Supporting hypothesis generation by learners exploring an interactive computer simulation. *Instructional Science* 20: 389–404.

Joolingen, W.R. van & Jong, T. de (1993). Exploring a domain through a computer simulation: Traversing variable and relation space with the help of a hypothesis scratchpad, in D. Towne, T. de Jong & H. Spada, eds., *Simulation-Based Experiential Learning* (pp. 191–206). (NATO ASI series). Berlin: Springer.

Klahr, D. & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science* 12: 1–48.

Klahr, D., Fay, A.L. & Dunbar, K. (1993). Heuristics for scientific experimentation: A developmental study. *Cognitive Psychology* 25: 111–146.

Kulkarni, D. & Simon, H. (1988). The process of scientific discovery: the strategy of experimentation. *Cognitive Science* 12: 139–175.

Kurland, L.C. & Tenney, Y.J. (1988). Issues in developing an intelligent tutor for a real-world domain: Training in radar mechanics, in J. Psotka, L.D. Massey & S.A. Mutter, eds., *Intelligent Tutoring Systems: Lessons Learned* (pp. 119–181). Hillsdale, NJ: Erlbaum.

Langley, P., Simon, H.A., Bradshaw, G.L. & Zytkow, J.M. (1987). *Scientific Discovery, Computational Explorations of the Creative Process*. Cambridge: MIT Press.

Lavoie, D.R. & Good, R. (1988). The nature and use of predictions skills in a biological computer simulation. *Journal of Research in Science Teaching* 25: 335–360.

Leutner, D. (1993). Guided discovery learning with computer-based simulation games: Effects of adaptive and non-adaptive instructional support. *Learning and Instruction, 3*: 113–132.
Mynatt, C.R., Doherty, M.E. & Tweney R.D. (1977). Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Quarterly Journal of Experimental Psychology* 29: 85–95.
Mynatt, C.R., Doherty, M.E. & Tweney R.D. (1978). Consequences of confirmation and disconfirmation in a simulated research environment. *Quarterly Journal of Experimental Psychology* 30: 395–406.
Newell, A. & Simon, H.A. (1972). *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall.
Njoo, M. & Jong, T. de (1993). Exploratory learning with a computer simulation for control theory: Learning processes and instructional support. *Journal of Research in Science Teaching* 30: 821–844.
Plötzner, R. & Spada, H. (1992). Analysis-based learning on multiple levels of mental domain representation, in E. de Corte, M. Linn, H. Mandl & L. Verschaffel, eds., *Computer-Based Learning Environments and Problem Solving* (pp. 103–129). Berlin: Springer Verlag.
Plötzner, R., Spada, H., Stumpf, M. & Opwis, K. (1990). Learning qualitative reasoning in a microworld for elastic impacts. *European Journal of Psychology of Education* 4: 501–516.
Qin, Y. & Simon, H.A. (1990). Laboratory replication of scientific discovery processes. *Cognitive Science* 14: 281–312.
Reimann, P. (1989). Modeling scientific discovery learning processes with adaptive production systems, in D. Bierman, J. Breuker & J. Sandberg, eds., *Artificial Intelligence and Education; Synthesis and Reflection. Proceedings of the 4th International Conference on AI and Education* (pp. 218–227). Amsterdam: IOS.
Schauble, L., Klopfer, L.E. & Raghavan, K. (1991). Students' transition from an engineering model to a science model of experimentation. *Journal of Research in Science Teaching* 28: 859–882.
Schauble, L., Glaser, R., Raghavan, K. & Reiner, M. (1991). Causal models and experimentation strategies in scientific reasoning. *The Journal of the Learning Sciences* 1: 201–239.
Shute, V.J. & Glaser, R. (1990). A large-scale evaluation of an intelligent discovery world: Smithtown. *Interactive Learning Environments* 1: 51–77.
Shute, V. J., Glaser, R. & Raghahavan, K. (1989). Inference and discovery in an exploratory laboratory, in P.L. Ackermann, R.J. Sternberg & R. Glaser, eds., *Learning and Individual Differences*. New York: W.H. Freeman.
Sime, J-A. & Leitch, R. (1992). A learning environment based on multiple qualitative models, in C. Frasson, G. Gauthier & G.I. McCalla, eds., *Intelligent Tutoring Systems, Proceedings of the 2$^n d$ International Conference ITS '92. Montréal, Canada. June 10–12 1992*. Berlin: Springer Verlag.
Simmons, P.E. & Lunnetta, V.N. (1993). Problem solving behaviors during a genetics computer simulation: Beyond the expert/novice dichotonomy. *Journal of Research in Science Teaching* 30: 153–173.
Simon, H.A. & Lea, G. (1974). Problem solving and rule induction: A unified view, in L.W. Gregg, ed., *Knowledge and Cognition*. Hillsdale, New York: Lawrence Erlbaum.
Spada, H., Stumpf, M. & Opwis, K. (1989). The constructive process of knowledge acquisition: Student modelling, in H. Maurer, ed., *Computer Assisted Learning, Proceedings of the 2$^{nd}$ International Conference ICCAL '89 (Dallas)* (pp. 456–500). Berlin: Springer Verlag.
Tschirgi, J.E. (1980). Sensible reasoning: A hypothesis about hypotheses. *Child Development* 51: 1–10.
Wason, P.C. (1960) On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology* 12: 129–140.
White, B.Y. & Frederiksen, J.R. (1990). Causal model progressions as a foundation for intelligent learning environments. *Artificial Intelligence* 42: 99–157.