

6 Solving Complex Problems: Exploration and Control of Complex Systems

Joachim Funke
University of Bonn/FRG

SOLVING COMPLEX PROBLEMS: EXPLORATION AND CONTROL OF COMPLEX SYSTEMS

Studying complex problem solving by means of computer-simulated scenarios has become one of the favorite themes of modern theorists in German-speaking countries who are concerned with the psychology of thinking. Following the pioneering work of Dietrich Dörner (University of Bamberg, FRG) in the mid-70s, many new scenarios have been developed and applied in correlational as well as in experimental studies (for a review see Funke, 1988). Instead of studying problem-solving behavior in restricted situations (like the "Tower of Hanoi" or "Cannibals and Missionaries"; cf. Greeno, 1974; Jeffries, Polson, & Razran, 1977), the new approach focuses on semantically rich domains that provide a touch of reality that was not inherent in the older research (see also Bhaskar & Simon, 1977). In the computer-administered scenario "LOHHAUSEN," for instance, subjects have to take over the regentship of a little town (Dörner, Kreuzig, Reither, & Stäudel, 1983). In other work, subjects take over the roles of a manager of a little shop (Putz-Osterloh, 1981), of an engineer in a developmental country (Reither, 1981), or of a pilot flying to the moon (Thalmaier, 1979). In general, the new approach deals with the *exploration and control of complex and dynamic systems by human individuals*.

This chapter is divided into four main parts. First, I give a working definition of what I mean by "complex problem solving" and suggest how complex tasks can be profitably analyzed and compared to each other across domains. Second, I summarize recent research on complex problem solving, analyze the main streams of current research, and discuss the underlying principles and mecha-

nisms uncovered so far. Also, I consider how people learn to solve complex problems and discuss expert-novice differences in complex problem solving. Third, I describe my own approach to studying complex problem solving in which it is conceptualized as a dynamic process of knowledge acquisition and of knowledge application. I briefly describe the so-called DYNAMIS project and the DYNAMIS shell for scenarios, and report the results of some studies within this framework. Finally, I give perspectives for future research.

DEFINITION OF COMPLEX PROBLEM SOLVING

I argue that complex problem solving can be understood by contrasting it with "simple," noncomplex problem solving in terms of the following, non-orthogonal criteria:

1. Availability of information about the problem, that is, transparency of the problem situation.
2. Precision of goal definition, that is, whether a goal is defined, and whether there are multiple goals, some of which may be contradictory.
3. "Complexity" of the problem as defined by the number of variables, the degree of connectivity among the variables, and the type of functional relationship (linear vs. nonlinear).
4. Stability properties of the problem, that is, time dependencies in the course of the problem-solving process ("Eigendynamik").
5. "Richness" of the problem's semantic embedding. Rich semantic embeddings often reduce the uncertainty to a large degree.

A complex problem-solving situation is one that can be characterized by the following features (with respect to the previously mentioned criteria):

1. "Intransparency": In complex problem-solving situations, only some variables lend themselves to direct observation. Often, only knowledge about "symptoms" is available, from which one has to infer the underlying state. This is a case of intransparency. Other cases of intransparency arise if variables can be assessed in principle, but their huge number requires selection of a few relevant ones.
2. "Polytely" (from the Greek words *poly telos* = many goals): Frequently, complex problem-solving situations are characterized by the presence of not one, but multiple goals. Problems can arise when some of the goals are contradictory (e.g., the manager who wants to make a lot of money, but has to pay high wages in order to find good workers), and a reasonable trade-off is required.
3. "Complexity of the situation": This feature concerns the number of identification and regulation processes involved. A complex problem-solving situation

is not only characterized by a large number of variables that have to be considered, but also by their complex connectivity pattern, by the possibilities to control the system, and by the dynamic aspects of the system. The growing complexity of situational demands may conflict with the limited capacity of the problem solver.

4. "Connectivity of variables": A high degree of connectivity describes a situation in which changes in one variable affect the status of many other, related variables. Complex problems often contain a high degree of connectivity, that is, it is very difficult to anticipate all possible consequences of a given situation.

5. "Dynamic developments": Complex problem-solving situations often change decrementally and worsen, forcing a problem solver to act immediately, under considerable time pressure. Also, spontaneous changes in the other direction are possible, causing less stress but making the situation less predictable.

6. "Time-delayed effects": Not every action shows immediate consequences. In complex problem-solving situations, effects often occur with time delay. This makes it necessary for the actor to wait patiently, in sharp contrast to the aforementioned situation, in which immediate action is required.

The features outlined differ not only from those traditionally emphasized in research on problem solving and thinking, but also from those employed in conventional intelligence tests. They do, however, allow for a more precise characterization of complex problem situations than do more traditional classifications, like the classification into well-defined and ill-defined situations. For example, Duncker's (1935) "radiation problem," although useful in studying analogical transfer (e.g., Gick & Holyoak, 1983), might not be classified as a complex problem according to the present classification scheme, because it lacks the feature of dynamic development as well as that of complexity.

Complex problem solving has also been a topic in recent man-machine research. With increasingly more automation and computerization, the operator of a complex technical system becomes a complex problem solver, rather than merely a controller (cf. Bainbridge, 1987). Process control tasks are used in the laboratory or observed in the field to address questions of systems design and of optimal training procedures of system-relevant knowledge. Because this research comes more from the applied, engineering point of view, however, it will not be reported in detail in this paper (see, e.g., Rasmussen, Duncan, & Leplat, 1987).

RECENT RESEARCH ON COMPLEX PROBLEM SOLVING

In the following, I consider some of the research on complex problem solving that has been conducted over the past 15 years. Following this review, I summarize: (a) the domain-specific and domain-general principles and mechanisms

underlying complex problem solving; (b) the acquisition of complex problem solving; and (c) expert-novice differences in complex problem solving.

Review of Studies on Complex Problem Solving

Because the research themes diverge and the domains that have been chosen are very heterogeneous, it is not easy to arrange the various studies in a systematic way. Even the simulation systems can only be compared superficially. For reasons of simplicity, the systems are, in the following, grouped according to their number of variables, a criterion which is sometimes seen as an essential indicator of complexity. Because no objective general measure of complexity exists, the number-of-variables criterion is just an expedient for orienting purposes. In this section, I give a short description of the major systems used in empirical research (for a more elaborated review see Funke, 1988).

Systems With up to 10 Variables. Systems with up to 10 variables are the most commonly used ones. Despite the fact that only a small number of variables is utilized, the complexity of these systems should not be underestimated. Table 6.1 gives an overview of the major systems in this category.

A major advantage of small systems is that all information relevant to the problem-solving situation can be displayed on a single computer screen, thus allowing the subject directly to interact with the system. For the small systems, the equations are given if known to the author. Systems are discussed in alphabetical order.

BLACK BOX. In Mackinnon and Wearing's (1985) BLACK BOX, subjects are asked to control an abstract, first-order feedback system for 75 trials. The behavior of the system can be described by a complicated formula (cf. Mackinnon & Wearing, 1985, p. 165). The subject's task is to maintain the goal value of a single system variable by controlling a single input variable. No information about system characteristics is given. BLACK BOX is a transparent system (no hidden variables) that has a single goal variable. The connectivity function is complex. There is no time pressure. The system develops dynamically. Effects of time-delayed feedback can be manipulated experimentally.

In an experimental study using BLACK BOX, Mackinnon and Wearing manipulated two factors: the value of the boundary function which amplified or attenuated the input value, and the intensity of feedback, operationalized via a short versus long "memory" of past inputs. For the data of 32 subjects, Mackinnon and Wearing found no significant effect of the amplification factor: Subjects were able to quickly adapt their inputs to different boundary parameters. In contrast, intensity of feedback did have a significant effect on subjects' problem-solving behavior. Results showed significantly better system control for the longer memory of past inputs. The authors concluded that a systems-analytical

TABLE 6.1
Overview of Simulation Systems: Systems With up to 10 Variables

Name	# of Variables	Reference
BLACK BOX	2	Mackinnon & Wearing (1985)
COLD-STORAGE DEPOT	6	Reichert & Dörner (1988)
ECONOMIC SYSTEM	4	Broadbent, FitzGerald, & Broadbent (1986)
ECOSYSTEM	6	Funke (1985)
GAS-ABSORBER	6	Hübner (1987)
HAMURABI	8	Gediga, Schöttke, & Tücke (1983)
INVENTORY PROBLEM	3	Kleiter (1970)
MINI-LAKE	6	Opwis & Spada (1985)
MOONLANDING	3	Thalmaier (1979)
PORAEU	8	Preussler (1985)
SIM002	10	Kluwe & Reimann (1983)
SINUS	6	Funke & Müller (1988)
SUGAR FACTORY	4	Berry & Broadbent (1987)
TRANSPORTATION	4	Broadbent (1977)
WORLD	4	Eyferth et al. (1982)

approach might be helpful in identifying the demands a problem-solving task makes upon the problem solver and might also make the comparison of different tasks and the ordered exploration of the range of possible tasks easier.

COLD-STORAGE DEPOT. In Reichert and Dörner's (1988) system, COLD-STORAGE DEPOT, subjects have to control a cold-storage depot by means of a steering wheel (u) with which the temperature of the depot (r) can be changed according to the following formula which is unknown to the subjects (s = outside temperature and v = delay factor; see Reichert, 1986):

$$r(t) = r(t-1) + (s(t) - r(t-1)) * 0.1 - q(t-1),$$

$$q(t) = (r(t-v) - u(t)) * 0.3.$$

This simulation system is transparent, has a single goal, dynamic development, time pressure, and, most importantly, time-delayed effects which require a careful control strategy.

In one of the studies using the COLD-STORAGE DEPOT, 54 student subjects had the opportunity to perform 100 interventions. Subjects were told that the automatic steering was defective and human control was necessary in order to prevent the food from being spoiled. The results of the study showed that only one-fifth of the subjects were able to run the depot successfully. The main difficulty for subjects was the time-delay of the nonlinear function relating subjects' interventions and the system's responses; some subjects recognized this delay and planned their actions adequately, that is, ahead of time, whereas other students changed their interventions immediately after receiving feedback. In-

terestingly, some of the "good" problem solvers were not able to verbalize the rules they were using so effectively. Reichert and Dörner developed what they called a "simulation of the simulation," that is, a psychological model that simulated the simulation game, which was able to produce a synthetic behavior almost indistinguishable from the behavior of real subjects.

ECONOMIC SYSTEM. Broadbent, FitzGerald, and Broadbent's (1986) ECONOMIC SYSTEM modeled an imaginary country, in which subjects can raise or lower the levels of taxation (R) and of government expenditure (G) in order to control the rates of unemployment (U) and of inflation (I) according to the following formulas (cf. Broadbent et al., 1986, p. 41):

$$\begin{aligned}U(t + 1) &= 12.8 - ((1 - R) * (G + 7650) / 730), \\I(t + 1) &= I(t) * (1.45 - 0.15 * U(t)).\end{aligned}$$

Broadbent et al. argued that their findings demonstrate a dissociation between verbal reports and actions. I return to this topic later when related work of the Broadbent group (SUGAR FACTORY, TRANSPORTATION) will be presented.

ECOSYSTEM. In ECOSYSTEM (Funke, 1985), subjects are asked to control the amounts of insects (Y1), leaves (Y2), and water pollution (Y3) in an ecosystem through the manipulation of poison (X1), vermin eaters (X2), and fertilizer (X3), according to the following system structure:

$$\begin{aligned}Y1(t + 1) &= 0.9 * Y1(t) + 1.0 * X2(t), \\Y2(t + 1) &= 1.0 * Y2(t) + 10.0 * X3(t), \\Y3(t + 1) &= 1.0 * Y1(t) - 0.1 * X1(t).\end{aligned}$$

ECOSYSTEM is a transparent, polytelic, complex, and dynamic system, in which time delay and connectivity can be manipulated as experimental variables.

The system simulates a total of five trials, each consisting of seven cycles. In the first four trials, subjects are encouraged to familiarize themselves with the system by actively exploring the system ("knowledge-acquisition phase"). In the last trial ("knowledge-application phase"), in contrast, subjects are asked to actively steer the system toward achieving a given goal state. Funke found that two critical system attributes, namely, the "connectivity of the variables" and the "degree of time delay," had a large effect on subjects' quality of the knowledge representation (a subject's diagnosed "mental model" of the system) as well as on the degree to which the goal was achieved, although the effects of time delay appeared to be weaker than the effects of connectivity. In a similar study, Fritz and Funke (1988) demonstrated differences between pupils with minimal cerebral dysfunction and matched controls with respect to discriminatory and integrational abilities in the process of hypothesis development and hypothesis testing.

GAS-ABSORBER. Hübner (1987) simulates a GAS-ABSORBER with one input variable (u) and three states (x):

$$x(t + 1) = \begin{bmatrix} 0.365 & 0.219 & 0.066 \\ 0.186 & 0.421 & 0.219 \\ 0.048 & 0.186 & 0.365 \end{bmatrix} x(t) + \begin{bmatrix} 0.019 \\ 0.100 \\ 0.389 \end{bmatrix} u(t)$$

The GAS-ABSORBER is a transparent, dynamic, complexly interconnected system with dynamic and time-delayed effects. The system's structure is, at a formal level, precisely defined and analyzed (see also Hübner, 1989).

In a study by Hübner (1987), two different learning conditions produced no differences with respect to the quality of control (measured as distance from a given objective point). However, distance to the goal at the beginning of the intervention phase proved to be very important: If the goal could be achieved in two steps, less input error was made than if the goal could be achieved in three steps. These results are consistent with those obtained in manual tracking studies (see Bösser, 1983).

HAMURABI. HAMURABI is the name of the absolute ruler of the agrarian state of "Summaria." In Gediga, Schöttke, and Tücke's (1983) system, subjects have the task of keeping alive as much of the population of Summaria as possible by using four manipulations: purchasing and selling arable acreage, deciding the area to be sown with corn, and determining the quantity of food required by each member of the community. Subjects run through two trials, each simulating a time period of 30 years. The system partly depends on randomly varying variables. Gediga et al. claim to have demonstrated that, on the one hand, problem situations with an exponential change over time were mastered by only a few subjects; on the other hand, the hypotheses of many subjects were in accordance with the complex problem situation and led to better performance.

HAMURABI is an intransparent, polytelic, complex system with dynamic components. Because of the random effects, it is not easy to determine the pure effects of subject interventions.

INVENTORY PROBLEM. Kleiter (1970) uses a situation in which a retailer stocks a product which spoils if it is not sold by the end of a certain period of time. For every unit sold, the amount won increases; the units not sold decrease the amount won. The formula combining the input supply (A) with the output of the demand (Z) and a random component (e) under an "optimism condition" is as follows:

$$Z(t + 1) = 0.25 * (A(t) - Z(t)) + Z(t) + e.$$

Under the "pessimism condition," a weight of -0.25 instead of 0.25 is used. The system is transparent, has a single goal, is low on complexity and connectivity, but shows dynamic developments. A random component makes subjects' performances more difficult to interpret.

For the "optimism condition," Kleiter found that the demand for the product increased when a higher amount than the last one was stocked; for the "pessimism

condition," it was the other way around. The results of 40 subjects working with this system for a maximum of 50 trials demonstrated that only 2 subjects in each condition were able to accumulate a win whereas 6 subjects in the optimism and 12 in the pessimism condition did not even reach the minimum win.

MINI LAKE. The ecosystem MINI LAKE (Opwis & Spada, 1985) is a biological population model (with isolated as well as integrated parts) that is transparent, complex, and interconnected, has multiple goals, dynamic development, and no time-delayed effects. Subjects managing the system are asked to infer the conditions of change that operate in the system to make predictions about future states. Subjects are given precisely designated objectives and are asked to take adequate action. They can change the amount of phosphate fertilizer (u_1) and fish biomass (u_2) to control for two kinds of phytoplankton (x_1, x_2) and zooplankton (x_3, x_4), according to the following matrices:

$$\dot{x}(t) = \begin{bmatrix} 0 & 0 & -0.244 & -0.122 \\ 0 & 0 & -0.110 & -0.220 \\ 0.036 & 0.018 & 0 & 0 \\ 0.016 & 0.032 & 0 & 0 \end{bmatrix} x(t) + \begin{bmatrix} 0.507 & 0 \\ 0.418 & 0 \\ 0 & -0.246 \\ 0 & -0.200 \end{bmatrix} u(t)$$

Opwis and Spada argue that the nature of reliable and valid problem-solving indicators is problematic in most systems: with unrestricted access to the system, unknown solubility of the task, and ignorance of subjects' internal mental representations of the system, the experimental examination of thought processes is virtually impossible. Opwis and Spada, therefore, use a research plan that allows control of these stated variables. A model based on subjects' individual knowledge data was able accurately to predict approximately 80% of subjects' answers to questions about the system.

MOONLANDING. Thalmaier (1979) uses the dynamic system MOONLANDING in which subjects have to control the landing maneuver of a space craft on the lunar surface. Thalmaier argues that the mathematical description of the problem type (e.g., the simulated system) and an understanding of the system's properties is a necessary prerequisite for understanding the behavior of experimental subjects. In his studies, Thalmaier found that 20 mathematics students who served as subjects were, indeed, able during a total of 20 practice landings, to recognize the dynamic aspects of the problem as well as its nonlinear development. Thalmaier concludes that subjects are not overtaxed from the beginning by nonlinear extrapolations. However, relative to an optimal steering strategy, subjects' difficulties in exploring and understanding the system were paramount. The successive recognition of the structure of the system through an input-output analysis had to occur first.

Empirical findings concerning this paradigm also come from a study by Funke

and Hussy (1984), who presented the MOONLANDING task in two different domains, in its original domain and as a similarly structured COOKING problem. Funke and Hussy predicted that experience with the two different reality domains would affect problem-solving performance. Twenty-four male and female subjects (assumed to be experts in the domains of MOONLANDING or COOKING because of sex-specific socialization) were used. The results, however, did not confirm the hypotheses. The main effects of the experimental conditions "domain" and "previous experience" on the dependent variable "quality of problem solving" were weak, and the expected interaction did not materialize. Statistical arguments did not allow an expanded interpretation of this finding.

In a similar study with the modified target-approach paradigm, Hussy and Granzow (1987; Hussy, 1989) showed that problem-solving quality (measured as distance to a target state) decreased as a function of the increasing number of variables as well as of nonlinear interweaving functions, and of lower problem transparency. Hussy and Granzow found a significant correlation between test intelligence and problem-solving quality—but only under transparent conditions with few variables, which seems to support earlier findings reported by Putz-Osterloh (1981).

MOONLANDING is low on complexity but contains dynamic components. Experimental manipulation of domain effects, of complexity, and of transparency effects demonstrate the usefulness of this scenario in analyzing different influences on problem-solving behavior.

PORAEU. PORAEU (Preussler, 1985) is a small nonlinear predator-prey model in which subjects have to anticipate the number of robbers and swags in a simulated ecosystem at each of the 35 discrete time points.

In a prediction experiment, Preussler crossed three semantic conditions (helpful: the growth of the robbers was bad for humans; hindering: growth of robbers was good for humans; or neutral: an abstract version of the system without semantic labels), two prognosis conditions (only robber values had to be predicted vs. the prediction of robber and swag values was requested), and two presentation forms (number of robbers and swags with or without graphical displays). Subjects did not receive any information about system variables and the connectivity structure. Because the author used more than 20 dependent variables in this experiment and tested more than 30 different hypotheses, it is difficult to summarize the results of this work in a few words. The main effects of the three factors on the predictive behavior were all nonsignificant; individual interactions, however, showed more distinct effects. Based upon an additional examination of response effects, the author concluded that individuals are not able to make predictions concerning exponential development trends. Interestingly, under conditions of graphical feedback, subjects approximated nonlinear developments by linear functions. Introducing graphical feedback by showing

the growth functions improved the quality of predictions, especially at later points in simulation time.

PORAEU is not a very complex system, but realizes a nonlinear dynamic development that is difficult for subjects to handle despite its transparency and the fact that only a single goal has to be controlled.

SIM002. Stimulated by a critical review of the studies on complex problem solving, Kluwe and Reimann (1983) derived an abstract system called SIM002. Kluwe and Reimann were less interested in pursuing the aim of simulating reality, rather, they wanted to develop systems that could be fit to many experimental inquiries. A more detailed description is given in the next section for the similar system, SIM00X.

SINUS. Funke and Müller (1988) were concerned with the effects of different demands of activity on the handling of an unknown dynamic system called SINUS. The system consists of living creatures from a distant planet called SINUS. The dependent variables are given the nonsense names, "Gaseln" (Y1), "Schmorken" (Y2), and "Sisen" (Y3); the independent variables are called "Olschen" (X1), "Mukern" (X2), and "Raskeln" (X3). The system has the following structure:

$$\begin{aligned} Y1(t + 1) &= 10.0 * X1(t) + 1.0 * Y1(t), \\ Y2(t + 1) &= 3.0 * X3(t) + 1.0 * Y2(t) + 0.2 * Y3(t), \\ Y3(t + 1) &= 2.0 * X2(t) + 0.5 * X3(t) + 0.9 * Y3(t). \end{aligned}$$

The task of the subjects is to explore the system and to control the dependent variables with respect to given goal states.

In their study, Funke and Müller manipulated (a) the possibility to actively or passively explore the system; and (b) whether the next system state had to be predicted or not. The amount of system knowledge subjects had acquired and the quality of problem solving served as dependent variables. Funke and Müller expected (a) the "interveners" to be superior to the pure "observers" with regard to amount of knowledge as well as to efficient operations; and (b) the "predictors" to accumulate more knowledge than the "nonpredictors." Subjects were 32 college students. Path-analytical evaluation of the data supported the expectations only partially: "interveners" were, indeed, better in dealing with the system, but seemed to know less than "observers" (cf. the similar dissociations reported by Broadbent et al., 1986; Putz-Osterloh, 1987). "Predictors" were more knowledgeable than "nonpredictors," but only in a special mode. Knowledge about the system was generally a good predictor of operating performance. Interestingly, there was a negative relation between the duration of the experiment and the quality of performance. Detailed analyses of so-called "experimental twins"—pairs of subjects who dealt with the same system situations—indi-

cated high interindividual variability, thus showing the relevance of person-specific ways of data-evaluation.

SINUS is a transparent, complex, interconnected, dynamic system with multiple goals and no time delays. It is an ideal instrument for experimental manipulation of system attributes.

SUGAR FACTORY. In Berry and Broadbent's SUGAR FACTORY (1984), subjects are asked to manage a small sugar-production factory in order to reach and maintain a given target production level. The size of the work force (W) can be varied in 12 discrete steps, yielding a level of production (P) according to the formula:

$$P(t + 1) = 2 * W(t) - P(t).$$

A second, mathematically equivalent task called PERSONAL INTERACTION used the same structure, but now the subject could choose between 12 styles of behavior (very rude, rude, very cool, cool, . . . , loving) in order to produce and maintain a target behavior in a fictitious person called Clegg. After two sets of 30 trials, results of a posttask questionnaire were correlated with control performance, yielding nonsignificant coefficients of about -0.50 (Exp. 1).

In a later study, the SUGAR FACTORY simulation was combined with the PERSONAL INTERACTION task (Berry & Broadbent, 1987; Marescaux, Luc, & Karnas, 1989) to make relationships more or less salient. The input variables were now the number of employed workers (W; 1-12) and the behavior toward the union chief (B; 1-10); the variables to be controlled were the level of sugar output (P; 1-21) and the behavior of the union chief (G; 1-10), according to the formulas:

$$\begin{aligned} P(t + 1) &= 1.8 * W(t) - 0.45 * B(t), \\ G(t + 1) &= 0.8 * B(t) + 0.45 * W(t). \end{aligned}$$

Results with this system illustrated the role of salience of relationships: No explicit knowledge about nonsalient relationships was acquired even when the system was handled very well.

TRANSPORTATION. In this system, first used by Broadbent (1977), subjects have to control the bus load (L) and vacant parking spaces (VS) in a fictitious city parking lot by manipulating the time intervals between bus arrivals (T) and the amount charged for use of the lot (F). The formulas are:

$$\begin{aligned} L(t + 1) &= 200 * T(t) + 80 * F(t), \\ VS(t + 1) &= 4.5 * F(t) - 2 * T(t). \end{aligned}$$

As in earlier mentioned studies, Broadbent (1977) reported a dissociation between the verbal statements of the subjects and their actual ability to control

the system. I discuss this phenomenon later in the context of development of knowledge.

SUGAR FACTORY, PERSONAL INTERACTION TASK, and TRANSPORTATION are all systems at the lower end of the complexity scale. They have no time delays, no intransparencies, no dynamic developments, and represent a situation with a single respectively a double goal. One might ask, therefore, if these systems would really represent complex problems.

WORLD. Eyferth, Hoffmann-Plato, Muchowski, Otremba, Rossbach, Spiess, & Widowski (1982) examine the coping possibilities—the “genesis of handling competence”—in a novel situation. **WORLD** exists as a series of pictures on a screen, upon which a few objects can carry out computer-controlled maneuvers and can move or interact with each other according to a set of fixed rules. The observer can use the keyboard to interrupt maneuvers and to become actively involved. The task is to understand the system rules and to manipulate the objects toward achieving a certain purpose. Four numbered squares move on the screen in various ways, changing after collisions. The observer can (a) vary the speed with which the squares move over the screen; (b) change the squares' directions of movement; and (c) stop the system. **WORLD** is a single-rule system, with dynamic development, no time delays, and a single goal.

The results of an exploratory study (Eyferth et al., 1982) indicated that subjects gradually construct a system representation and connect it to existing schemata.

Systems With up to 100 Variables. Table 6.2 shows the systems that belong to this category.

DAGU. DAGU (Reither, 1981) simulates the climatic, ecological, and ethnic situation of a fictional African developing area. Subjects' goals are to create better living conditions for the people of DAGU and to increase the population, but to prevent overpopulation. Seven operational areas (i.e., with the possibility for interventions) are used: food, animal fodder, birth control, medical supply, preventive actions against tse-tse flies, set-up of irrigation projects, and sale of produce. The results of Reither's research on DAGU are reported later in the section on expert-novice differences.

DAGU as well as the following DORI and EPIDEMIC systems can be classified as a highly interconnected, complex, dynamic, and intransparent system with multiple goals. The DORI and EPIDEMIC system are offsprings of the DAGU program and their features are very similar to DAGU's.

DORI. DORI simulates the living conditions of a nomad tribe in the Sahel region, whose livelihood depends on cattle rearing. Hesse (1982) compares a semantic version of DORI to a structurally identical, nonsemantic version in

TABLE 6.2
Overview of Simulation Systems: Systems With up to 100 Variables

Name	# of Variables	Reference
DAGU	12	Reither (1981)
DORI	12	Hesse (1982)
EPIDEMIC	13	Hesse, Spies, & Lüer (1983)
FACTORY	>20	Zimolong (1987)
FIRE FIGHTING	>10	Brehmer (1987)
MEDICAL DECISION	>10	Kleinmuntz & Kleinmuntz (1981)
MORO	49	Strohschneider (1986)
SIM00X	15	Kluwe, Misiak, Ringelband, & Haider (1986)
TAILORSHOP	24	Putz-Osterloh (1981)
TANALAND	54	Dörner & Reither (1978)
TANK SYSTEM	14	Moray, Lootsteen, & Pajak (1986)

which the variables were designated by Latin letters. In addition, Hesse crossed the semantic factor with a factor transparency, whose two levels were the presence or absence of a graphical display of the connections between the variables. Hesse found that notes were more heavily consulted in the abstract, nonsemantic conditions, but that subjects in the semantic group asked more pointed questions and organized their actions better. In general, the observed differences between good and poor problem solvers suggested a difference in strategy that was related to the content area. In the abstract condition, there was a positive relation between intelligence test scores and problem-solving quality. This result is in line with previously observed correlations between transparency and IQ in a study by Putz-Osterloh (1981).

EPIDEMIC. EPIDEMIC is a system that is very similar to DORI. EPIDEMIC, however, uses a different content area and also new individual connections. Subjects are asked to take charge of the health authority of a small town in the aftermath of an epidemic disaster (Hesse, Spies, & Lüer, 1983; the system variables and equations as well as the similarities to the DORI system are fully described in Spies & Hesse, 1987). Their decisions are aimed at reducing the number of illnesses. Subjects have a choice among seven possible interventions. EPIDEMIC's main concern is the effect of personal distress, which is realized by simulating two kinds of epidemics, each of which is presented to different subject groups. Whereas a reduced level of distress is supposedly induced by a simulated influenza epidemic, a higher level of distress is induced by a dangerous smallpox epidemic. In both cases, the same structural equations are used, only the semantic labels of the variables are changed. The findings of the experiment point out the effectiveness of the variable semantic content upon problem-solving quality:

the highly distressed students obtained higher quality values, worked harder, took more effective actions, and recognized effective measures more readily.

FACTORY. Zimolong's (1987) FACTORY is a real-time, interactive, computer-simulation program that simulates a manufacturing system containing up to seven machine stations. The spatial design of the machine places, the pathway of the material handling system through the production unit, and the launching point of the parts are arbitrarily adjustable. The characteristics of the individual machine stations can be changed in different ways. The screen image (showing the machines and their actual state) is updated every second; the subjects can check and maintain the state of the machinery in order to prevent breakdown of the factory. FACTORY is a highly dynamic and real-time environment, with many variables, time delay effects, and partial intransparency.

Empirical work by Zimolong (1987) showed that, after one hour of practice with the simulation system, risk-taking behavior (measured as time to expected tool wear failure) under conditions of complete human control was less developed than under conditions of limited control. Zimolong concluded from these results that the job design in a flexible manufacturing system should care for an active operator instead of automated conditions.

FIRE FIGHTING. Brehmer (1987) is interested in the mental models problem solvers develop on the basis of direct, interactive experience with a system. Brehmer describes a "dynamic decision problem" as one in which (a) a series of interdependent decisions is required to reach the goal; (b) the environment changes over time; and (c) the decisions change the state of the world, thus creating new decision problems. Based on a general computer program for simulating dynamic decision problems called DESSY (Dynamic Environmental Simulation System), the FIRE FIGHTING scenario simulates "the decision problems facing a fire chief who obtains information about forest fires from a reconnaissance plane" (p. 115). The information is displayed on a VDU, and the subject has command over eight fire-fighting units. The goal is to prevent the fire from reaching the base as well as minimizing the area that is burned down. FIRE FIGHTING is a complex, dynamic system with multiple goals and with time delayed effects.

Brehmer's studies (see also Brehmer & Allard, 1991) demonstrated that complexity (measured in number and efficiency of fire fighting units) had "little or no effect on performance, so long as the total efficiency of the units as a whole is kept constant" (p. 118). In contrast, delay of even minimal feedback had disastrous effects. Brehmer concluded that subjects do not manage to form any truly predictive model of the system, but, instead, base their reactions only on direct feedback.

MEDICAL DECISION TASK. Kleinmuntz and Kleinmuntz (1981) use a simulated medical decision task environment that is based on probabilistic relations between symptoms, diseases, and treatments. Within this scenario, a person (the doctor) is confronted with an ill patient complaining of three symptoms and suffering from one out of five possible diseases. The doctor can, at each point in time, request a test for any of 30 symptoms; in addition, she can choose among 12 different treatments. The task is dynamic insofar as the disease generally causes the patient to get progressively closer to death from trial to trial (linear trend), because each test for a symptom has a detrimental effect, and because the same treatments can have vastly different effectivenesses depending upon the disease of the patient. Comparing the strategies of (a) expected utility, (b) heuristic decision, and (c) generate-and-test, Kleinmuntz and Kleinmuntz found strategy (a) to be best, (b) slightly worse, and (c) less good. Data from human subjects were not reported.

MEDICAL DECISION TASK has conflicting goals and dynamic components, but is not very complex and interconnected.

MORO. Strohschneider (1986), Putz-Osterloh (1987), Putz-Osterloh & Lemme (1987), and Stäudel (1987) use the scenario MORO, which simulates the situation of a small nomad tribe in the southern Sahara. MORO is a polytelic, intransparent system with highly interconnected and dynamic variables, which partly show time-delayed effects.

In one of the studies using MORO, Strohschneider (1986) deals with the question of just how far this research instrument can be used to gather stable data and what evidence for the external validity of these data can be found. Concerning test-retest stability, Strohschneider concluded that behavioral indices (e.g., the number of questions posed) show a higher reliability than measures of the system's condition (e.g., the number of starving people). From an exhaustive debriefing of the subjects, Strohschneider concluded that subjects perceive the demands on their problem-solving ability as valid in the simulated scenario as in everyday complex problem solving.

SIM00X. SIM00X is a descendant of SIM002, which therefore will be described first. The system SIM002 consists of 10 system variables, whose relations are fixed in a first-order parameter matrix. The system states are displayed in the form of a histogram on the monitor of a personal computer, and the subjects can change as many of the variables as they want to at any given time. The goal of the problem solver is to reach a nominal value displayed on the screen; the difference between achieved state and goal determines the quality measure.

A more recent version of the system SIM002 is the system SIM00X, in which the number of variables have been increased to 15 and the system variables have

been arranged into groups. Unrestricted access to the system is followed by a step-by-step confining of the status display. At uneven intervals, subjects have to reproduce the previous system states or to anticipate the next ones.

The SIM00X systems are complex, interrelated, dynamic systems with partial intransparency and multiple goals. System characteristics can easily be changed for experimental purposes.

A central assumption of the work using SIM002 and SIM00X concerns the postulate of various states of construction of mental models, which are identified in individual studies of longer duration. The authors see the elapsing complex learning processes under the perspective of "chunk" construction. Because an ideal intervention into the system with respect to the stated aim can be designated at any time (because of the system construction), the process of learning can be described accurately. An increase in proficiency is coupled with a gain in time, which (as with chunk building) is open to large individual differences. At the end of a long steering period (200 simulation tacts per subject), the subjects have a verbalizable system knowledge with respect to the connections of the variables as well as to the specific qualities of individual variables.

TAILORSHOP. TAILORSHOP (Putz-Osterloh, 1981; see also Putz-Osterloh & Lüer, 1981; the systems equations are fully published in Funke, 1983) is a miniature system in which subjects take over the management of a tailor shop: By purchasing raw materials and modifying the production capacity in terms of workers and machines used, shirts are to be produced and to be sold at a profit. The goal is to describe and examine the sine qua nons of complex problem solving and of intelligence test tasks, and which problem-solving processes can be used to surmount these requirements. In addition, complex problems should be more strongly equated with everyday problem situations than intelligence tests presently are. In a complex problem, as opposed to an IQ-test item, the construction and derivation of problem-solving objectives requires a choice of actions leading to the achievement of the goals and the active search for information about relevant system variables. TAILORSHOP is an intransparent, complex, dynamic, interconnected system with imprecise goals and time-delayed effects.

A study by Putz-Osterloh and Lüer (1981) tested the hypothesis that test-intelligence and problem-solving performance are related through a comparison of a transparent with a nontransparent condition ($N = 70$ student subjects). The two experimental conditions were the presence or absence of an illustration, which presented the connections between the system variables. Only under the transparent condition did the authors find a statistically significant correlation between problem-solving performance and IQ. They interpreted this result as a criticism of common intelligence tests in which transparency is generally high. They argued that "real" problems are rather intransparent and highly complicated and therefore demand behavior that cannot be measured by intelligence tests.

TANALAND. The TANALAND system (Dörner & Reither, 1978) was one of the earliest simulation studies published. The ecosystem of an African landscape with various flora and fauna as well as human groups, the "Tupis" and "Moros," who live by cattle and sheep farming, is simulated. The 50 or more system variables are connected through a complicated process of "positive and negative feedback." Subjects are to assume the role of a technical agronomy advisor to improve the living conditions of the native population. The system is very difficult to handle. As Dörner and Reither have shown, almost no subject is able to succeed in this task. The observed failures mirror deficits of a more general nature. It appeared that the subjects did not possess enough cognitive ability to be able to cope with complex systems. The failure of "linear thinking" was proposed. In the realized systems, which were described in terms of such features as dynamics, complexity, connectivity, and opaqueness, thinking in the form of causal networks should be considered.

TANK SYSTEM. Interested in the acquisition of process control skills, Moray, Lootsteen, and Pajak (1986) use a tank system consisting of four subsystems. Each subsystem consists of one tank with input and output valves and a heater. Temperature and level of each tank are shown on a VDU in analog and digital form. The task is to control either one or all of the tanks with respect to given required set points for level, temperature, and flow rate; the required points should be reached as rapidly as possible. Each of the 12 trials is run until these goal points are reached. TANK SYSTEM is an interconnected, transparent, complex system with multiple goals and dynamic development.

Discussing the problems of data analysis, the authors concluded that it would make no sense to average individual data. Rather, the data should be analyzed separately for each operator. Looking at the graphs of the system variables as a function of time, they concluded further that operators develop good "mental models" of the system. "One aspect of the more complex skill is, therefore, the discovery of causal relations and their use to develop control tactics" (p. 498). Starting with closed-loop control, good operators later developed almost perfect open-loop control. Switching from the control of one tank to the control of four tanks simultaneously, learning slowed down and interference effects occurred. Despite the enormous variety in sequences used to achieve the required goal, strategies emerged that were related to the development of the mental model "which represents the dynamics and causality of the system and leads to more efficient control" (p. 504).

Systems with more than 100 variables. Only two systems (see Table 6.3) that have more than 100 variables and were used in scientific research are known to this author. One of them—LOHHAUSEN—is the most prominent example of the new way of studying problem-solving processes. (For a short description of LOHHAUSEN in English language see Dörner, 1987.)

TABLE 6.3
Overview of Simulation Systems: Systems With
More Than 100 Variables

Name	# of Variables	Reference
ENERGY SUPPLY	>2000	Vent (1985)
LOHHAUSEN	>2000	Dörner, Kreuzig, Reither, & Stäudel (1983)

ENERGY SUPPLY. The "Energieversorgung" (ENERGY SUPPLY) of private households in the Federal Republic of Germany was simulated in a large-scale system in which individual energy choice preferences were projected over time and space (Vent, 1985). The author was concerned with the effect of various presentation and feedback forms that stimulate certain ways of thinking (for instance, analytical or holistic thinking). In one condition, he presented the system's data numerically and in the other, graphically. The results supported the superiority of a visual-holistic way of thought over an analytic style of thought as measured by the quality of decisions. There have been no follow-up studies with this system.

LOHHAUSEN. If TANALAND was the first scenario to study complex problem solving, then LOHHAUSEN was its expansion. The simulated reality domain (a small city called LOHHAUSEN) contains more than 2000 variables. Dörner, Kreuzig, Reither, and Stäudel's (1983) comprehensive monograph introduced the five years of work on this unique study with the following sentences:

The following report states the results of a relatively long-term psychological experiment. We tried to find out something about the conditions and forms of actions in ambiguous and complex situations. For this we systematically observed 48 subjects over a relatively long period and processed the manifold results of these observations. (p. 13)

LOHHAUSEN, originally the name of the simulated town, has since become the name of a research program: a deepening of cognitive psychology through new paradigms of problem-solving research; paradigms that, in contrast to the traditional types of problems such as mind games or mental exercises, contain the characteristics of complexity and uncertainty. Subjects, who took over the role of mayor of LOHHAUSEN, were instructed to "take care of the future prosperity of the town over the short and long term," that is, over a simulated 10-year period. Testing was done in eight two-hour sittings. Approximately 100,000 data points per subject resulted, from which the authors hoped to successfully separate the important from the trivial, and accidental from meaningful information.

The analysis of the findings—with a few exceptions, such as case studies of

selected experimental subjects—was based upon aggregated data. The authors first agglomerated the objective and subjective measures of problem-solving quality to a single "General Quality Criterion," which made it possible to split the total sample into two extreme groups ($N = 12$) of good and poor problem solvers. Results comparing the two groups showed that different variables in the LOHHAUSEN system (such as earnings of the industry, funds of the town, stocks of the bank, production and trade data, number of inhabitants, and rate of employment) developed more detrimentally when worked on by the poor problem solvers than by the good problem solvers. Even the "good problem solvers" were not what their designation suggested, however: System experts (the experimenters) achieved even higher values on some of the variables.

The behavioral effects were less interesting than the connected thought, planning, and decision processes: Besides formal characteristics (e.g., the frequency and consistency of decisions) and content biases (e.g., "financial situation of the watch factory") of the experimenters' "gross protocol," there were interesting references in the "think-aloud protocols" that the subjects were encouraged to produce. Subjects' problem-solving quality and their test intelligence were found not to be correlated. Neither Raven's Advanced Progressive Matrices (APM) nor Cattell's Culture Fair Intelligence Test (CFT) correlated substantially with the solution quality. Rather, what correlated significantly with problem-solving performance was the experimenters' spontaneous judgment that a "subject makes an intelligent impression."

Although the authors were right in pointing out the shortcomings of classical IQ-tests (e.g., not taking information search into account), they themselves did not take a possible shortcoming of their own findings into account: the possible lack of reliability of their problem-solving measures. It is known that intelligence tests, when used repeatedly, produce homogeneous results. Also, the sample limitations (students with a restricted range of IQ scores) should not have been ignored when the results were interpreted.

Further findings of the LOHHAUSEN study were concerned with personality characteristics and their relation to solution quality. The construct of "self-confidence" has to be given a special mention in this context; it had a strong positive relation to complex problem solving and was introduced to set off "the total failure of intelligence tests." Also, prior knowledge was not a significant predictor of success.

The condensed theory of this comprehensive study contains a list of elementary information processing methods for dealing with complex problems such as, for example, component and dependence analysis as well as sub- and superordination processes. The construction and pursuit of partial objectives by a subject is subsumed under an intention management model. Based upon the emotional embedding of cognitive processes (Dörner, Reither, & Stäudel, 1983), the intellectual emergency reaction—a quick and general reaction of the cognitive system to unspecified danger situations—can be brought into connection with the

actual competence of the actor. Self-confidence can be used as an indicator for heuristic competence, which refers to the ability "to be able to create adequate ways of dealing even with unknown situations" (Dörner, Reither, & Stäudel, 1983, p. 436; cf. Stäudel, 1987). Central to the theory is the concept of control: Control competence guarantees action in uncertainty, and loss of control leads to the negative emotional consequences, which override problem-solving thought.

LOHHAUSEN not only stands for a new field of research in cognitive psychology; it is also an appeal against the prevalent "analytical procedure" in scientific endeavor. The examination of the highly complicated cognitive system of "mankind"—following Dörner—cannot be pursued using strictly experimental means because the isolation of a few chosen variables in a laboratory says little about the "normal" interplay of processes that are interactively embedded within other variables (see also Dörner, 1989). The demand for an intensified "collecting of beetles and butterflies," that is, the exact description of the observed phenomena, goes hand in hand with the search for an overlapping conceptual framework concerning the complete workings of the psychic system. (A first impression of this framework can be found in Dörner, Schaub, Stäudel, & Strohschneider, 1988).

At the end of this section, the question has to be raised as to what sense it makes to permanently create new systems. It is surprising, for instance, that no replication of the famous LOHHAUSEN study exists; in fact, many of the previously mentioned systems lack this basic scientific requirement. From this author's point of view, if new systems are, indeed, needed, existing systems should be modified, rather than new systems created, in order to fulfill certain experimental requirements. If different systems are used in different studies, results can neither be compared nor heterogeneous conclusions clearly attributed to certain system attributes. What is missing, then, is a descriptive schema of systems that allows us directly to compare different systems with respect to such attributes as complexity, connectivity, transparency, etc. The following section offers a taxonomy of influence factors that might help to organize the different studies.

Main Streams of Current Research

Main streams of current research as revealed in the material reviewed center around the following three topics, which might serve as a taxonomy of possible influential determinants: (a) personal factors (poor vs. good problem solvers), (b) situational determinants of complex problem solving, and (c) system characteristics.

The *role of personal factors* can be differentiated in three ways: (a) cognitive abilities, (b) emotional and motivational factors, (c) personality characteristics in a broader sense. Concerning the *cognitive abilities*, one would probably expect intelligence to play an important role in handling complex situations. "Whatever intelligence may be, *reasoning* and *problem solving* have traditionally been viewed as important subjects of it. Almost without regard to how *intelligence* has

been defined, *reasoning* and *problem solving* have been part of the definition" (Sternberg, 1982, p. 225). The empirical support for the effect of intelligence on the quality of complex problem solving is rather poor, however: most empirical studies report either low or even zero correlations. Correlations tend to increase, however, when the problem-solving situation is made more similar to the intelligence-test situation; that is, when the problem situation is made more transparent. Also, a more differentiated diagnosis of intelligence reveals higher correlations on subtest-level rather than on a global one (see Thomas, Hörmann, & Jäger, 1989; Hussy, 1989).

Concerning *emotional and motivational factors*, one has to acknowledge that in the course of action, many situations develop that might evoke emotional reactions; for example, critical events that demonstrate a person's inability to cope with the given situation. There are presumably many feedback loops between "pure" cognitive processes and these evaluation processes; in case of luck, or of good interventions, they could be stabilizing; otherwise, one might expect a lot of disturbances stemming from the noncognitive area. Dörner, Kreuzig, Reither, and Stäudel (1983) reported an "intellectual emergency reaction" for some of their subjects, a quick and general reaction of the cognitive system to unspecified danger situations. The effects of this reaction were (a) a general increase in activation, (b) an externalization of behavior control (reduction of situation analysis and growing use of dogmatic principles), and (c) the activation of unspecific, precautionary behavior.

Personality characteristics in a broader sense seem to have a great influence primarily in the beginning stages and especially under conditions of intransparency. In these cases, cognitive abilities and knowledge are less required than, for example, a stable personality that shows no overload due to the huge amount of uncertainty. One can imagine that people with high anxiety and/or low self-confidence will tend to retreat from these situations. It is, thus, evident that a lot of nonintellectual abilities are necessary to cope with uncertain situations. These abilities are *not* problem-solving qualities themselves, but, rather, reflect the importance of individual differences due to different state and trait personality characteristics.

The *role of situational determinants* is related to (a) the transparency of the situation, and (b) the concrete task demands with which a subject has to cope.

As previously mentioned, the *transparency of a situation* depends on the degree to which a subject has direct access to system information. This factor can be manipulated easily by the experimenter. Putz-Osterloh (1981), for example, used a diagram that displayed the relations among the system variables: under a transparent condition, subjects could see this diagram; under intransparency, they could not. Degree of (in)transparency of system connections is not the only way of manipulating this variable, however. Another frequently used method consists of varying how subjects get the information they want: under transparent conditions, subjects are shown all interesting variables on a VDU; the subject thus has direct access to the system. Under intransparency, the experimenter is the medi-

ator between system and subject; every time a subject wants some information (i.e., about the actual values of a certain variable), he or she has to ask the experimenter who will give an answer if possible.

Additional situational determinants are the *concrete task demands* subjects have to fulfill. Sometimes they have to control an (unknown) system right from the start, sometimes they are allowed to explore the system in a previous phase. Demands vary also with respect to goals: sometimes no goals are given at all (the finding of an adequate goal is part of the task), sometimes a few selected variables, and sometimes all variables, have to be controlled.

Brehmer (1989) conceptualizes the tasks of process control as "dynamic decision tasks"—in contrast to static or sequential decision tasks—with the following four characteristics: "(a) a series of decisions are required; (b) these decisions are interdependent; (c) the decision problem changes, both autonomously and as a consequence of the decision maker's action; and (d) the decisions are made in real time" (p. 144). Based on the assumption that the human decision-maker does not want to resolve discrete choice dilemmas, but, rather, attempts to achieve control, Brehmer (see also Brehmer & Allard, 1991) characterizes dynamic decision tasks more precisely, differentiating between (a) complexity (in relation to control); (b) feedback quality (the problem of indicators); (c) feedback delay (which implies feedforward instead of feedback control); (d) possibilities for decentralization (i.e., give control to local decision makers); (e) rate of change (controlling the economy of a country vs. flying a jet); and (f) the relation between the control process and the to-be-controlled process.

Understanding the *role of system characteristics* requires a differentiation between (a) formal aspects and (b) aspects with regard to the contents.

The *formal aspects* are related to the number of variables, their connectivity, the resulting stability of the system, the degree of time delays, and so on. Also, the distinction between time-continuous and time-discrete systems is useful as is that of linear versus nonlinear systems. The question of deterministic versus stochastic modeling of the domain has to be answered, too.

The *aspects with regard to content* are not so easy to specify. Primarily, one is concerned with the semantic embedding of the system in question, but also with the relation between the actually implemented structure and the structure which is assumed by the subject because of previous knowledge.

Principles and Mechanisms Underlying Complex Problem Solving

It is not easy to list the principles and mechanisms that govern complex problem-solving activity by a human operator. One may ask, in fact, if there are any special mechanisms and principles that are applied to complex problem-solving tasks, or if it is simply sufficient to look for the mechanisms applicable to the solving of simple problems. As Tversky and Kahneman (1974) have claimed:

"People rely on a limited number of heuristic principles which reduce complex tasks to simpler judgmental operations." (p. 1124) The reason to look for special principles comes from the new demands that the complex control tasks require. As mentioned in the introduction, complex problems have unique features that they do not share with simple problems. Therefore, one has good reason to assume that special mechanisms are needed to deal with these features.

On the other hand, there are also reasons to assume that a general model could be applied to this situation. Knaeuper and Rouse (1983), for example, suggest the application of the production-system formalism. They specify four tasks that an operator has to perform: (a) transition tasks (bringing the system into a certain state), (b) steady-state tuning, (c) detection and diagnosis of failures, and (d) compensation for failures. Therefore, they postulate, the operator needs knowledge about (a) how the system would evolve if left alone; (b) what effects the control actions would have; and (c) which of the four abovementioned tasks is the appropriate one.

For purposes of clarity, I first give a brief sketch of how—under ideal conditions—structural knowledge about an unknown system can be acquired and used; then the development and use of strategic knowledge is described. The separate presentations should not be taken as suggesting that the two aspects can be seen as unconnected parts, however.

The Development of Structural Knowledge. In the following, a normative stage model of structural knowledge acquisition is outlined for which empirical data have still to be delivered. Yet, despite the missing empirical base, the model might serve as a useful frame of reference for discussing the principles and mechanisms of solving complex problems in terms of knowledge acquisition and knowledge application processes.

Complex problem solving requires the development of structural knowledge, which describes the functional or causal relationships between variables. This explicit knowledge is the condensed result of a hypothesis-formation and hypothesis-evaluation process. It can be assumed that this knowledge starts from simple observation of contingencies between subsequent system states. At its first stage, such knowledge is restricted to the pure identification of a relation between at least two variables (relational knowledge). At the second stage, such observations lead to a more differentiated view that allows a statement about the sign of the relation (sign knowledge). At the third stage, finally, the precise influence factor can be specified (numerical knowledge).

The three kinds of knowledge are expected to be in a declarative format. The basic unit looks like the following quadruple:

$$H_t = \langle V_1, V_2, C, B \rangle,$$

that is, a hypothesis at time t (H_t) consists of four components: specifications of (a) a variable 1 (V_1), (b) a variable 2 (V_2), (c) a connectivity form (C), and (d) a degree of belief (B) in the hypothesis. For example, for a hypothesis on the

causal influence of "advertise" (V_1) on "demand" (V_2), the connectivity form (C) of a hypothesis could be "positive linear," the degree of belief (B) "high."

One could argue that this conceptualization of a hypothesis is similar to a schema with four slots. If one knows the semantics of V_1 and V_2 , for example, connectivity (C) and degree of belief (B) may be default values due to prior experience.

In addition to the explicit knowledge, implicit knowledge also emerges. In a series of studies, Berry and Broadbent (1984, 1987; cf. Broadbent & Aston, 1978; Broadbent et al., 1986) found evidence that, despite low explicit task knowledge (as measured by a questionnaire), subjects were able to control small systems with good performance. These results are not without problems: Haider (1989), on the basis of a simulation study, argued that, for the control of these systems, complete explicit knowledge is not necessary. However, one has to accept the possibility that subjects acquire a lot of information above the degree that is usually assessed by diagnostic procedures. The results of a recent series of studies by Sanderson (1989) point to the same conclusion. Using the TRANSPORTATION system originally introduced by Broadbent et al. (1986), Sanderson demonstrated association as well as dissociation effects between task performance and verbalizable knowledge depending on amount of practice, kind of display, and cover story. Interestingly, Sanderson's results contradict the common thesis that, with growing practice, verbal task knowledge decreases (e.g., Anderson, 1983).

The Development of Strategic Knowledge. It is by far more complicated to describe how strategic knowledge develops in the course of action. In most cases, this is essentially a problem of application, and not of developing the concepts. Like structural knowledge, the strategic procedures used by, and known to, a subject depend primarily on previous experience. Unlike structural knowledge, however, it is not as easy to construct experimental conditions under which the influence of previous strategic knowledge is minimized. Some kind of "naive experimentation," the use of concepts like "isolated (or systematic) variation of conditions," "Eigendynamik," "side effects," etc., are part of this strategic inventory which a subject possesses.

Learning to Solve Complex Problems

Learning to solve complex problems has two aspects. The first aspect concerns the improvement in handling a system over the course of repeated experiences (domain-specific learning). The second aspect concerns the potential transfer from one complex problem to another (domain-general learning). Generally, one is interested in improving complex problem-solving performance by some kind of training.

One example of this approach is a training study by Streufert, Nogami,

Swezey, Pogash, and Piasecki (1988). They used a design in which 56 subjects first had to work with one of two complex scenarios (either as coordinator of a disaster control center or as governor of a developing country) for six hours. Then, a first group (17 of the participants) received extensive training on domain-general rules (especially on structural management styles) and on their concrete operationalization. A second group (7 persons) received only the specific training, and a third, control group (31 persons) received no training at all. After that, all subjects had to work with the second system in order to assess training effects. As was expected, condition (a) yielded significant improvements in 8 out of 13 performance measures compared with only 4 improvements and 1 impairment under condition (b) and 5 impairments without any improvement under the no-training condition.

Many training methods for improving learning abilities exist (for a review see Derry & Murphy, 1986), including microcomponent training as well as meta-strategy approaches. But one has to consider that training an executive control mechanism that would automatically assess and combine learning skills whenever needed can only gradually be developed and requires time. Derry and Murphy, therefore, concluded that, "The choice of which taxonomy to use and which learning skills to train is a matter of selecting what is appropriate for the student population, the training time allowed, and the type of learning material involved" (p. 32). With respect to complex problem solving, one might expect that a lot of time is needed for improving behavior. Hays and Singer (1989) give a good overview of problems and possible solutions in designing and evaluating training systems.

Differences between Experts and Novices

Differences between novices and experts in "normal" problem solving were summarized by Mayer (1988), who stated "that experts and novices differ with respect to their tendency by using chunking in free recall, to use comprehensive plans in solving problems, and to classify problems based on their underlying solution requirements" (p. 572).

For complex problem solving, too, the observed differences between novices and experts are a relevant source for theorizing.

Reither (1981), using the DAGU simulation system, compared the results of 12 development-aid workers with 6 to 8 years of practical experience in third world countries with those of 12 subjects who were about to begin their first mission in developmental aid. Subjects working in groups of three had to create better living conditions for the people of Dagou and to increase their population smoothly. The results showed differences between novices and experts insofar as novices thought more in causal chains than in causal webs (i.e., thinking mainly in terms of "straight-on" main effects instead of taking possible side effects into their deliberations), showed more thematic jumps, and made more "metastate-

ments." The hallmark of expertise was experts' blind coping, that is, the fact that experts arrived at conclusions under all circumstances, thereby demonstrating a continuity of action under every condition. Interestingly, even the experts were not able to stabilize the critical variable "population size," however.

Putz-Osterloh (1987; see also Putz-Osterloh & Lemme, 1987) conducted a study in which she analyzed complex problem solving by experts (7 professors in economy, aged between 40 and 49) and novices (30 randomly selected students, aged between 19 and 27) who had to control the previously described business system TAILORSHOP for 15 months first and then the developmental scenario MORO for 20 months. Data analysis was concerned with the first six simulated months of each system, evaluating (a) the quality of interventions (analysis of subjects' behavior according to a complicated rating scale); (b) the frequency of use of domain-general strategies (according to a classification of verbal statements during thinking aloud); and (c) system knowledge (also derived from thinking-aloud protocols). Results indicated that professors of economy were better with respect to all three kinds of criteria in the TAILORSHOP situation. In the MORO situation, experts were better only with respect to strategies and knowledge, but not with respect to the quality of intervention. Putz-Osterloh concluded that experts differ from novices not because of different amounts of available data, but because of differences in processing these data: experts generate more correct hypotheses, more frequently and correctly analyze the relations between variables, and verbalize more often the expected effects of their planned interventions.

Critical remarks concern the operationalization of expertise, which is markedly confounded with age and therefore with life experience. Also, the dependent variables are primarily of a verbal nature, a kind of data that, in the opinion of this author, should not be the only source relied upon to characterize expert-novice differences.

SOLVING COMPLEX PROBLEMS AS KNOWLEDGE ACQUISITION AND KNOWLEDGE APPLICATION

In this section, I briefly describe my own work on complex problem solving. First, I describe the experimental setting, the scenario, that I used in various studies, and second, I summarize the theoretical framework underlying the studies.

Description of Research Within the DYNAMIS-Project

From the beginning, research about solving complex problems had to cope with a number of difficulties (see the critical aspects mentioned by Funke, 1984). One central difficulty was the measurement of problem-solving quality. Because there

was no "best" intervention (due to the partially nonlinear relationships between the variables for which no optimal solution could be found), one could never be quite sure if a subject's solution to a problem was really better or worse than any others. Therefore, problem-solving quality often was rated by "experts."

The line of research done in our laboratory adheres to the following principles (Funke, 1986):

1. It should always be possible to define the quality of a given problem solution by comparing it with an optimal solution strategy.
2. The problem-solving situations should take into account the aforementioned features of complex problem solving insofar as possible.
3. A differentiated diagnostic procedure should reveal the subject's development of hypotheses about the system. This implies repeated measurements and/or the use of thinking-aloud techniques under certain conditions.
4. There should be a clear distinction between a phase of knowledge acquisition (mainly realized by letting the subjects experiment with the system) and a phase of knowledge application, in which given states of the problem space should be reached by the subjects as quickly as possible.

The DYNAMIS Shell for Scenarios

Control of complex systems with dynamic behavior requires knowledge from the operator that has previously been acquired. To study the acquisition as well as the application of knowledge, we confront our subjects with computer-simulated scenarios. A universal tool for constructing these scenarios is the computer program DYNAMIS. It works like a shell, in that the user can implement in a simple way different types of simulation systems that follow one general frame of reference. This general frame of reference is a *linear equation system* (see, e.g., Steyer, 1984), which consists of an arbitrary number of exogenous (=x) and endogenous (=y) variables according to the following equation:

$$Y(t + 1) = A*Y(t) + B*X(t),$$

where $Y(t + 1)$ and $Y(t)$ are vectors representing the state of the y-variables at times $t + 1$ and t ,

$X(t)$: a vector representing the values chosen by the subject for the x-variables,

A, B: weight matrices.

The construction of the equation system follows theoretical considerations about the influence of certain system attributes on task complexity (e.g., the effect of self-dynamic, side effects or effects due to interdependencies). It is not intended to simulate a domain of reality adequately because this kind of simulation demands too many constraints on the attributes of the system. Consequently,

most of the simulation systems used in our research group are artificial ones. Regarding a distinction made by Hays and Singer (1989), one can say we do not want physical fidelity of our simulation systems but do want functional fidelity (for a taxonomy of simulation fidelity considerations see also Alessi, 1988).

Experimental Procedure

In our experiments, subjects pass through at least two phases. In the first phase, the *knowledge acquisition* phase, the subject is allowed to explore the system and its behavior (see also Moray et al., 1986). Subjects can take actions (i.e., make an intervention on one or more of the exogenous variables) and observe the resulting effects on the endogenous variables. From time to time, we measure the acquired knowledge by asking the subject for a graphical representation of his or her structural knowledge. In the second phase, called *knowledge application*, the subject has to reach a defined system state and try to keep the variables on the defined values. In this phase, we measure the quality of the operator's control by assessing the distance between the reached and the defined values for all endogenous variables. Some remarks on measuring structural knowledge and system performance seem necessary at this point because of their central role in each study. Empirical studies based on this procedure have been done by Fritz and Funke (1988, "ECOSYSTEM"), Funke (1985, "ECOSYSTEM"), and by Funke and Müller (1988, "SINUS").

Measuring Structural Knowledge and System Performance

Measuring knowledge and performance seems only on the first view to be an easy problem (for an overview, see Kluwe, 1988; Spada & Reimann, 1988). At a closer view, there are a lot of difficulties, some of which are mentioned and possible solutions presented. A good review of problems in diagnosing "mental models" is given by Rouse and Morris (1986).

Starting with *system performance quality*, the goal is to determine how well a given goal state is approximated by the operator's interventions. The classical approach requires the measurement of the deviation from the target system state by means of the root-mean-squares criterion (RMS). This indicator reflects the mean deviation, independent of sign, and weighs the individual deviations more heavily the further away they are from the target state. A good discussion of the frequently used RMS criterion can be found in Bösser (1983).

There is an aspect that reflects an ugly property of this kind of system performance evaluation, however. Assume that an operator has knowledge about the system. Reaching the goal is of little or no difficulty for him and the resulting RMS will be low and near zero. But what about the operator with little or missing knowledge? The resulting distance to the goal state as measured by RMS varies

with his (random) interventions. According to certain system characteristics, this would result in a large variety of measured distances. Therefore, different values of the RMS do not, in this case, reflect different degrees of quality of system performance. The argument here is one of different reliabilities of the RMS criterion for different states of an operator's knowledge, which is best in the case of correct knowledge (RMS indicating reliable values near zero) and worst in the case of pure random intervention (RMS indicating a huge range of values due to decreasing reliability).

One potential solution for this problem is a logarithmic transformation of the RMS. This transformation leads to an evaluation of distances which is more efficient: larger distances are now not weighed as more important but, rather, as less important. It does not matter if someone failed the goal by 10,000 or 100,000 points, as this is of the same importance as the difference between a deviation of 1 and 10. The transformation, thus, reduces the error variance that increases with the operator's distance to the goal state.

Measuring the *structural knowledge* an operator has acquired about a system requires also some kind of distance or similarity measurement, in this case between the assumed and the real existing structural relations. For this purpose, the operator marks on a sheet (or, in some versions, directly on the screen) the assumed causal relationships at certain points in time. The problems with this kind of measurement are:

1. Subjects differ with respect to their response tendency. Therefore, one has to count not only the hits (i.e., correspondence between assumed and existing relation), but also the false alarms.
2. Subjects differ with respect to the quality of knowledge they can talk about: Sometimes any relationship between two variables is assumed (relational knowledge), sometimes the sign of the relation is known (sign knowledge), sometimes even the numerical weights are known (numerical knowledge).
3. A false model can be useful for system control, at least within a restricted area of values. The functionality of a model is somewhat independent of its correctness.
4. Subjects take certain assumptions as givens but do not talk about them. For example, if a variable does not change over the time, the weight for that relation is assumed to be equal to one. The subject often does not find worth mentioning this kind of knowledge.
5. It is not clear if a subject follows only one single model or if there exist some models concurrently.

For problems (3) to (5), no solution can be given at present. Problems (1) and (2), however, can be solved by using a quantification of the following kind: For each specification of a subject, one first counts whether it belongs to one of the

three classes of knowledge (relational, sign, numerical) and whether it is right or wrong with respect to the implemented system. Then, for each level one can determine the "quality of system identification" (QI) as the difference between "hits" (HI) and "false alarms" (FA), weighted by some kind of "guessing" probability (P) according to the following scheme, which closely resembles the discrimination index P_r from the two-high threshold model for recognition memory (see Snodgrass & Corwin, 1988; the proposed "correction for guessing" goes back to Woodworth, 1938):

$$QI = (1 - p) * HI / \max(HI) - p * FA / \max(FA), \\ -P \leq QI \leq (1 - P).$$

The guessing probability for numerical parameters in a dynamic system could, for instance, be set equal to zero, so all hits count relative to the maximal number of hits. If one sets the guessing probability to 0.5 in the case of sign knowledge, then errors lead to a reduction in the QI index for that stage.

Note that it is not required that the operator have a complete and correct model of the system in question, because good control is possible with models that are partly incorrect (but see Conant & Ashby, 1970). The reason for this stems from the functionality a false structure can have: With respect to a restricted range of values, different models can be functionally equivalent (to the degree of control they allow). This might be one reason for the fact that in everyday life we often use wrong models that are nevertheless functional (see, for example, Kempton, 1986, who could show that up to 50% of Americans do not have a correct theory of home-heating control).

Relation to Research on Scientific Reasoning

Klahr and Dunbar (1988; see also Klahr, Dunbar, & Fay, in press) have recently developed an integrated model of scientific reasoning that seems to be applicable to our scenarios. At the core of their model is the Generalized Rule Inducer (GRI) (Simon & Lea, 1974), augmented by (a) a mechanism for the identification of relevant attributes (because in scientific situations these attributes are not given as in concept-learning tasks) and (b) a more complex "instance generator." Klahr and Dunbar argue "that scientific reasoning can be conceptualized as a search through two problem spaces: an hypothesis space and an experiment space" (1988, p. 7). These spaces result from the task a scientist has, may he be a skilled one or a naive one:

The successful scientist, like the successful explorer, must master two related skills: knowing where to look and understanding what is seen. The first skill—experimental design—involves the design of experimental and observational procedures. The second skill—hypothesis formation—involves the formation and evaluation of theory. (p. 2)

Working with the toy "BigTrak"—a computer-controlled robot tank that can be programmed (see also Shrager & Klahr, 1986)—Klahr and Dunbar asked their subjects to explore the functions of the device, especially the function of the RPT-key. A fine analysis of verbal protocols of their subjects (Experiment 1) showed that the hypothesis space was very small: Only eight different "common" hypotheses concerned with four attributes were found. A frame representation of the concept of the RPT key was important insofar as once a subject had constructed a particular frame, the task became one of filling in or verifying the contents of the slots in this frame.

With respect to the experiment space, Klahr and Dunbar showed that there are six different "regions" (combinations of values for two critical parameters) with different conclusions to be drawn. Incidentally, the experiment space is not the representation a subject has, but one that allows classifying subjects' experimental procedures.

Data from their Experiment 1 demonstrated that there were at least two groups of subjects: so-called "Experimenters" and "Theorists." The difference between the two groups stems from the strategy according to which they shifted their frames: If subjects switched their frame as a consequence of a certain experimental outcome, they were called "Experimenters"; if subjects searched in their hypothesis space and came to a shift, they were labeled as "Theorists."

This theoretical framework seems applicable to the research topic of this chapter insofar as the experimental situation Klahr and Dunbar were concerned with—exploration of a hitherto unknown object—is basically identical to the situation of handling an unknown system. Furthermore, Klahr and Dunbar conceptualize the process of knowledge acquisition in terms of hypotheses that are more or less deductively or inductively developed.

Aside from the similarities in the experimental procedures and in the theoretical frames of reference, however, there are some differences. The main difference can be seen in the way subjects' knowledge is measured. Klahr and Dunbar use primarily verbal data (see Bainbridge, 1979, for a comment on verbal data in this context), whereas in our procedure different approaches are taken to diagnose the structural knowledge a subject acquires. This difference is partly due to our "object" of exploration: subjects explicitly have to anticipate the next states of the system, they have to write down their hypotheses about structural relationships, and they have to control the system as well as possible.

PERSPECTIVES FOR FUTURE RESEARCH

Three tasks for future research are outlined briefly: (a) a differentiation between factors influencing complex problem solving resulting from the individual as well as from the situational and the system attributes; (b) reliability and validity research on complex problem-solving scenarios; and (c) adequate measurement

of the actual "mental model" and of the potential heuristics that complex problem solvers use, also over time.

Concerning the first task, separation of person, situation, and system influences on performance measures, the approach taken by Streufert et al. (1988) seems to point in the right direction: instead of using "free" simulations in which decisions can change a system's state quite drastically, they use a "quasi-experimental simulation technology," in which the system reacts in part independently from subjects' interventions such that each subject receives comparable informations and events. Despite this fact, subjects still believe that they have direct or delayed impact on the system. This technique should be extended further in order to standardize the conditions under which subjects' performance quality is measured independent from system attributes.

Concerning Task 2, reliability and validity aspects, there is a lot of work to do: up to now, mainly face validity exists. Jäger (1986) speaks of "uncovered checks" that have to be cashed in subsequent research (p. 274). It is simply not enough to show that there are no correlations to standard intelligence tests, because many reasons can account for that result. Rather, one has to show positive connections to other psychometric instruments as well as to external criteria. One possible line of research could be the use of the learning test concept (e.g., Guthke, 1982), according to which intelligence is not a static variable but, rather, is to be interpreted as "learning potential." It seems plausible that there exists a relation between learning potential and the ability to solve complex problems.

With respect to reliability, work also has to be done. One promising way of evaluating reliability of complex problem-solving indicators was recently proposed by Müller (1989). Following the concepts of consistency and stability developed by Steyer (1987), Müller applied this design to studies of complex problem solving. His procedure is as follows: at two times of measurement (TM_1 and TM_2), two parallel forms of a system (PF_1 and PF_2) are given to the same subjects, yielding the four observed variables y_{11} , y_{21} , y_{12} , and y_{22} (first index for the system parallel form, second index for time of measurement; see Fig. 6.1).

In this design (and when certain assumptions about uncorrelated residuals, etc., are met), it is possible, based on covariances, to determine the degree of measurement accuracy (=consistency) as well as the stability of the measured construct (the problem-solving "competence").

Concerning Task 3, the adequate measurement of the operator's mental model and his heuristics, one has to develop instruments that sensitively assess those relevant parts of human memory that are required for exploration and control. Whereas in the area of assessing structural (or declarative) knowledge some useful techniques exist, there are clear deficits in diagnosing the heuristic knowledge on which human problem solvers operate. Also, more attention should be given to developing measurement techniques that reveal the implicit knowledge of an operator.

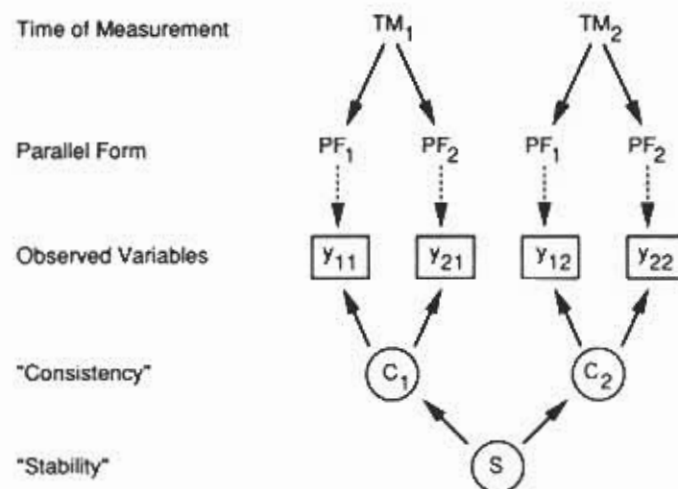


FIG. 6.1. Design of a study for assessing consistency and stability data (from Müller, 1989).

Concerning the general research strategy, it seems more useful to manipulate critical variables in systems that already exist than to create new systems. Only the strategy of analyzing the effects of small variations—the experimental method—can offer new insights into the principles and mechanisms that govern complex human problem solving.

ACKNOWLEDGMENTS

Preparation of this chapter was supported in part by a grant from the "Deutsche Forschungsgemeinschaft (DFG)" to the author (Az. Fu 173/1). Thanks to Edgar Erdfelder, Horst Müller, and Uwe Kleinemas for valuable comments on a draft version of this chapter. Special thanks to Peter Frensch who corrected my German English and provided many comments that made the chapter more precise and understandable.

REFERENCES

- Alessi, S. M. (1988). Fidelity in the design of instructional simulation. *Journal of Computer-Based Instruction*, 15, 40-47.
- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge: Harvard University Press.
- Bainbridge, L. (1979). Verbal reports as evidence of the process operator's knowledge. *International Journal of Man-Machine Studies*, 11, 411-436.
- Bainbridge, L. (1987). Ironies of automation. In J. Rasmussen, K. Duncan, & J. Leplat (Eds.), *New technology and human error* (pp. 271-283). Chichester, England: Wiley.

- Berry, D. C., & Broadbent, D. E. (1984). On the relationship between task performance and associated verbalizable knowledge. *Quarterly Journal of Experimental Psychology*, 36, 209-231.
- Berry, D. C., & Broadbent, D. E. (1987). The combination of explicit and implicit learning processes in task control. *Psychological Research*, 49, 7-15.
- Bhaskar, R., & Simon, H. A. (1977). Problem solving in semantically rich domains: an example from engineering thermodynamics. *Cognitive Science*, 1, 193-215.
- Bösner, T. (1983). Eine nichtlineare Regelstrategie bei der manuellen Regelung [A nonlinear control strategy at manual control]. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 30, 529-565.
- Brehmer, B. (1987). Development of mental models for decision in technological systems. In J. Rasmussen, K. Duncan & J. Leplat (Eds.), *New technology and human error* (pp. 111-120). Chichester, England: Wiley.
- Brehmer, B. (1989). Dynamic decision making. In A. P. Sage (Ed.), *Concise encyclopedia of information processing in systems and organizations* (pp. 144-149). New York: Pergamon Press.
- Brehmer, B., & Allard, R. (1991). Dynamic decision making: The effects of task complexity and feedback delay. In J. Rasmussen, B. Brehmer, & J. Leplat (Eds.), *Distributed decision making. Towards a basis for modeling* (pp. 319-334). Chichester, England: Wiley.
- Broadbent, D. E. (1977). Levels, hierarchies, and the locus of control. *Quarterly Journal of Experimental Psychology*, 29, 181-201.
- Broadbent, D. E., & Aston, B. (1978). Human control of a simulated economic system. *Ergonomics*, 21, 1035-1043.
- Broadbent, D. E., FitzGerald, P., & Broadbent, M. H. P. (1986). Implicit and explicit knowledge in the control of complex systems. *British Journal of Psychology*, 77, 33-50.
- Conant, R. C., & Ashby, W. R. (1970). Every good regulator of a system must be a model of that system. *International Journal of System Science*, 1, 89-97.
- Derry, S. J., & Murphy, D. A. (1986). Designing systems that train learning ability: From theory to practice. *Review of Educational Research*, 56, 1-39.
- Dörner, D. (1987). On the difficulties people have in dealing with complexity. In J. Rasmussen, K. Duncan & J. Leplat (Eds.), *New technology and human error* (pp. 97-109). Chichester, England: Wiley.
- Dörner, D. (1989). Die kleinen grünen Schildkröten und die Methoden der experimentellen Psychologie [The little green turtles and the methods of experimental psychology]. *Sprache & Kognition*, 8, 86-97.
- Dörner, D., Kreuzig, H. W., Reither, F., & Stäudel, T. (1983). *Lohhausen. Vom Umgang mit Unbestimmtheit und Komplexität* [Lohhausen. On dealing with uncertainty and complexity]. Bern: Huber.
- Dörner, D., & Reither, F. (1978). Über das Problemlösen in sehr komplexen Realitätsbereichen [On problem solving in very complex domains of reality]. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 25, 527-551.
- Dörner, D., Reither, F., & Stäudel, T. (1983). Emotion and problemlösendes Denken [Emotion and problem solving thinking]. In H. Mandl & G. L. Huber (Eds.), *Emotion und Kognition* (pp. 61-84). München: Urban & Schwarzenberg.
- Dörner, D., Schaub, H., Stäudel, T., & Strohschneider, S. (1988). Ein System zur Handlungsregulation oder—Die Interaktion von Emotion, Kognition und Motivation [A system for action regulation or—The interaction of emotion, cognition, and motivation]. *Sprache & Kognition*, 7, 217-232.
- Duncker, K. (1935). *Zur Psychologie des produktiven Denkens* [The psychology of productive thinking]. Berlin: Julius Springer.
- Eyferth, K., Hoffmann-Plato, I., Muchowski, L., Otremba, H., Rossbach, H., Spiess, M., & Widowski, D. (1982). *Studienprojekt Handlungsorganisation* [Project: Organization of action]. Berlin: Institut für Psychologie der TU Berlin (Forschungsbericht Nr. 82-4, korrigierter Nachdruck).
- Fritz, A., & Funke, J. (1988). Komplexes Problemlösen bei Jugendlichen mit Hirnfunktionsstörungen [Complex problem solving with pupils who have minimal cerebral dysfunctions]. *Zeitschrift für Psychologie*, 196, 171-187.
- Funke, J. (1983). Einige Bemerkungen zu Problemen der Problemlöseforschung oder: Ist Testintelligenz doch ein Prädiktor [Some comments to problems of problem solving research, or: An intelligence test is a predictor, isn't it?]. *Diagnostica*, 29, 283-302.
- Funke, J. (1984). Diagnose der westdeutschen Problemlöseforschung in Form einiger Thesen [Diagnosis of problem solving research in the FRG in form of some theses]. *Sprache & Kognition*, 3, 159-171.
- Funke, J. (1985). Steuerung dynamischer Systeme durch Aufbau und Anwendung subjektiver Kausalmodelle [Control of dynamic systems through development and application of subjective causal models]. *Zeitschrift für Psychologie*, 193, 443-465.
- Funke, J. (1986). *Komplexes Problemlösen. Bestandsaufnahme und Perspektiven* [Complex problem solving. State of the art and perspectives]. Heidelberg: Springer.
- Funke, J. (1988). Using simulation to study complex problem solving. A review of studies in the FRG. *Simulation & Games*, 19, 277-303.
- Funke, J., & Hussy, W. (1984). Komplexes Problemlösen: Beiträge zu seiner Erfassung sowie zur Frage der Bereichs- und Erfahrungsabhängigkeit [Complex problem solving: contributions to its measurement and to the question of the effects of context and prior experience]. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 31, 19-38.
- Funke, J., & Müller, H. (1988). Eingreifen und Prognostizieren als Determinanten von Systemidentifikation und Systemsteuerung [Intervention and prediction as determinants of systems identification and system control]. *Sprache & Kognition*, 7, 176-186.
- Gediga, G., Schöttke, H., & Tücke, M. (1983). Problemlösen in einer komplexen Situation [Problem solving in a complex situation]. *Archiv für Psychologie*, 135, 325-339.
- Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, 15, 1-38.
- Greeno, J. G. (1974). Hobbits and orcs: Acquisition of a sequential concept. *Cognitive Psychology*, 4, 270-292.
- Guthke, J. (1982). The learning test concept—An alternative to the traditional static intelligence test. *German Journal of Psychology*, 6, 306-324.
- Haider, H. (1989, March). *Die Bedeutung impliziten Lernens beim Problemlösen* [The importance of implicit learning in problem solving]. Paper presented at the "31. Tagung experimentell arbeitender Psychologen," Bamberg, FRG.
- Hays, R. T., & Singer, M. J. (1989). *Simulation fidelity in training system design. Bridging the gap between reality and training*. New York: Springer.
- Hesse, F. W. (1982). Effekte des semantischen Kontexts auf die Bearbeitung komplexer Probleme [Effects of semantic context on the processing of complex problems]. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 29, 62-91.
- Hesse, F. W., Spies, K., & Luer, G. (1983). Einfluß motivationaler Faktoren auf das Problemlöseverhalten im Umgang mit komplexen Problemen [On the effect of motivational variables at solving complex problems]. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 30, 400-424.
- Hübner, R. (1987). Eine naheliegende Fehleinschätzung des Zielabstandes bei der zeitoptimalen Regelung dynamischer Systeme [An obvious fallacy on the distance to the criterion in time optimal control of dynamic systems]. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 34, 38-53.

- Hübner, R. (1989). Methoden zur Analyse und Konstruktion von Aufgaben zur kognitiven Steuerung dynamischer Systeme [Methods for analysis and construction of tasks for cognitive control of dynamic systems]. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 36, 271-238.
- Hussy, W. (1989). Intelligenz und komplexes Problemlösen [Intelligence and complex problem solving]. *Diagnostica*, 35, 1-16.
- Hussy, W., & Granzow, S. (1987). Komplexes Problemlösen, Gedächtnis und Verarbeitungsstil [Complex problem solving, memory, and kind of processing]. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 34, 212-227.
- Jäger, A. O. (1986). Validität von Intelligenztests [Validity of intelligence tests]. *Diagnostica*, 32, 272-289.
- Jeffries, R., Polson, P. G., & Razran, L. (1977). A process model for missionaries-cannibals and other river crossing problems. *Cognitive Psychology*, 9, 412-440.
- Kempton, W. (1986). Two theories of home heat control. *Cognitive Science*, 11, 75-90.
- Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, 12, 1-48.
- Klahr, D., Dunbar, K., & Fay, A. L. (in press). Designing good experiments to test bad hypotheses. In J. Shrager & P. Langley (Eds.), *Computational models of discovery and theory formation*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Kleinmuntz, D. N., & Kleinmuntz, B. (1981). Decision strategies in simulated environments. *Behavioral Science*, 26, 294-305.
- Kleiter, G. D. (1970). Trend control in a dynamic decision making task. *Acta Psychologica*, 34, 387-397.
- Kluwe, R. H. (1988). Methoden der Psychologie zur Gewinnung von Daten über menschliches Wissen [Psychological methods for data acquisition about human knowledge]. In H. Mandl & H. Spada (Eds.), *Wissenspsychologie* (pp. 359-385). München: Psychologie Verlags Union.
- Kluwe, R. H., & Reimann, H. (1983). *Problemlösen bei vernetzten, komplexen Problemen: Effekte des Verbalisierens auf die Problemlöseleistung* [Problem solving with complex problems: Effects of verbalizing on problem solving quality]. Hamburg: Bericht aus dem Fachbereich Pädagogik der Hochschule der Bundeswehr.
- Kluwe, R. H., Misiak, C., Ringelband, O., & Haider, H. (1986). Lernen durch Tun: Eine Methode zur Konstruktion von simulierten Systemen mit spezifischen Eigenschaften und Ergebnisse einer Einzelfallstudie [Learning by doing: A method for construction of simulated systems with specific traits and results of a single case study]. In M. Amelang (Ed.), *Bericht über den 35. Kongress der DGfPs in Heidelberg 1986* (p. 208). Göttingen: Hogrefe.
- Knaeuper, A., & Rouse, W. B. (1983). A model of human problem solving in dynamic environments. In A. T. Pope & L. D. Haugh (Eds.), *Proceedings of the Human Factors Society* (Vol. 2, pp. 695-699). Santa Monica, CA: Human Factors Society.
- Mackinnon, A. J., & Wearing, A. J. (1985). Systems analysis and dynamic decision making. *Acta Psychologica*, 58, 159-172.
- Marescaux, P. J., Luc, F., & Karnas, G. (1989). Modes d'apprentissage sélectif et non-sélectif et connaissances acquises au contrôle d'un processus: évaluation d'un modèle simulé [Selective vs. unselective learning and knowledge acquisition about process control: Evaluation of a simulation model]. *Cahier de Psychologie Cognitive*, 9, 239-264.
- Mayer, R. E. (1988). From novice to expert. In M. Helander (Ed.), *Handbook of human-computer interaction* (pp. 569-580). Amsterdam: Elsevier Science Publishers.
- Moray, N., Looftstein, P., & Pajak, J. (1986). Acquisition of process control skills. *IEEE Transactions on Systems, Man, & Cybernetics*, 16, 497-504.
- Müller, H. (1989, March). *Zur Reliabilität verschiedener Indikatoren beim Bearbeiten komplexer dynamischer Systeme* [On the reliability of different indicators of dealing with complex dynamic systems]. Paper presented at the "31. Tagung experimentell arbeitender Psychologen," Bamberg, FRG.
- Opwis, K., & Spada, H. (1985). Erwerb und Anwendung von Wissen über ökologische Systeme [Acquisition and application of knowledge about ecological systems]. In D. Albert (Ed.), *Bericht über den 34. Kongress der DGfPs in Wien 1984* (pp. 258-260). Göttingen, FRG: Hogrefe.
- Preussler, W. (1985). *Über die Bedingungen der Prognose eines bivariaten ökologischen Systems* [On the determinants of forecasting a bivariate ecological system]. Bamberg: Memo No. 31 at the chair Psychology II at the University of Bamberg.
- Putz-Osterloh, W. (1981). Über die Beziehung zwischen Testintelligenz und Problemlöseerfolg [On the relationship between test intelligence and success in problem solving]. *Zeitschrift für Psychologie*, 189, 79-100.
- Putz-Osterloh, W. (1987). Gibt es Experten für komplexe Probleme [Are there experts for complex problems?] *Zeitschrift für Psychologie*, 195, 63-84.
- Putz-Osterloh, W., & Lemme, M. (1987). Knowledge and its intelligent application to problem solving. *German Journal of Psychology*, 11, 286-303.
- Putz-Osterloh, W., & Lüer, G. (1981). Über die Vorhersagbarkeit komplexer Problemlöseleistungen durch Ergebnisse in einem Intelligenztest [On the predictability of complex problem solving through intelligence test scores]. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 28, 309-334.
- Rasmussen, J., Duncan, K., & Leplat, J. (Eds.). (1987). *New technology and human error*. Chichester, England: Wiley.
- Reichert, U. (1986). *Die Steuerung eines nichtlinearen Regelkreises durch Versuchspersonen* [The control of a nonlinear system by subjects]. Bamberg: Memo No. 42 at the chair Psychology II at the University of Bamberg.
- Reichert, U., & Dörner, D. (1988). Heuristiken beim Umgang mit einem "einfachen" dynamischen System [Heuristics in the control of a "simple" dynamic system]. *Sprache & Kognition*, 7, 12-24.
- Reither, F. (1981). About thinking and acting of experts in complex situations. *Simulation & Games*, 12, 125-140.
- Rouse, W. B., & Morris, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin*, 100, 349-363.
- Sanderson, P. M. (1989). Verbalizable knowledge and skilled task performance: Association, dissociation and mental models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 729-747.
- Shrager, J., & Klahr, D. (1986). Instructionless learning about a complex device. *International Journal of Man-Machine Learning*, 25, 153-189.
- Simon, H. A., & Lea, G. (1974). Problem solving and rule induction: A unified view. In L. W. Gregg (Ed.), *Knowledge and cognition* (pp. 105-127). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, 117, 34-50.
- Spada, H., & Reimann, P. (1988). Wissensdiagnostik auf kognitionswissenschaftlicher Basis [Diagnosis of knowledge from a cognitive science point of view]. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 9, 183-192.
- Spies, K., & Hesse, F. W. (1987). Problemlösen [Problem solving]. In G. Lüer (Ed.), *Allgemeine Experimentelle Psychologie* (pp. 371-430). Stuttgart: Gustav Fischer.
- Stäudel, T. (1987). *Problemlösen, Emotionen und Kompetenz. Die Überprüfung eines integrativen Konstrukts* [Problem solving, emotions and competence. The evaluation of an integrative construct]. Regensburg, FRG: Roderer.
- Sternberg, R. J. (1982). Reasoning, problem solving, and intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 225-307). Cambridge: Cambridge University Press.
- Steyer, R. (1984). Causal linear stochastic dependencies: An introduction. In J. R. Nesselroade &

- A. von Eye (Eds.), *Individual development and social change: Explanatory analysis* (pp. 95–124). New York: Academic Press.
- Steyer, R. (1987). Konsistenz und Spezifität: Definition zweier zentraler Begriffe der Differentiellen Psychologie und ein einfaches Modell zu ihrer Identifikation [Consistency and specificity: Definition of two central concepts of differential psychology and a single model for their identification]. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 8, 245–258.
- Streufert, S., Nogami, G. Y., Swezey, R. W., Pogash, R. M., & Piasecki, M. T. (1988). Computer assisted training of complex managerial performance. *Computers in Human Behavior*, 4, 77–88.
- Strohschneider, S. (1986). Zur Stabilität und Validität von Handeln in komplexen Realitätsbereichen [On the stability and validity of action in complex domains of reality]. *Sprache & Kognition*, 5, 42–48.
- Thalmaier, A. (1979). Zur kognitiven Bewältigung der optimalen Steuerung eines dynamischen Systems [On the cognitive mastering of the optimal control of a dynamic system]. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 26, 388–421.
- Thomas, M., Hörmann, H. J., & Jäger, A. O. (1988). Systemwissen als Indikator der Bewältigung von Komplexität [System knowledge as indicator for coping with complexity]. In W. Schönplüg (Ed.), *Bericht über den 36. Kongress der Deutschen Gesellschaft für Psychologie in Berlin 1988* (pp. 301–302). Göttingen, FRG: Hogrefe.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124–1131.
- Vent, U. (1985). *Der Einfluß einer ganzheitlichen Denkstrategie auf die Lösung von komplexen Problemen* [The effect of a holistic strategy of thinking on the solution of complex problems]. Kiel: Institut for Pedagogic of Natural Science at the University of Kiel (= IPN working paper No. 61).
- Woodworth, R. S. (1938). *Experimental psychology*. New York: Holt.
- Zimolong, B. (1987). Decision aids and risk taking in flexible manufacturing systems. A simulation study. In G. Salvendy (Ed.), *Cognitive engineering in the design of human-computer interaction and expert systems* (pp. 265–272). Amsterdam: Elsevier Science Publishers.

7

Do Lawyers Reason Differently From Psychologists? A Comparative Design for Studying Expertise

Eric Amsel*
Rosanna Langer**
and Lynn Loutzenhiser***
University of Saskatchewan

Legal reasoning has a logic of its own.

—E. H. Levi (1949)

Lawyers are experts in solving complex problems in their domain of expertise. For example, how a case ought to be presented, a contract drawn up, and when and when not to go to court are types of problems best left to lawyers who are trained to deal with them. Generally speaking, people appreciate the value of legal expertise. Very few people hire a psychologist to solve their legal problems or a lawyer to solve their psychological problems, although, as any good episode of "L.A. Law" suggests, people sometimes confuse their psychological problems with their legal ones. Such confusions of problem-states notwithstanding, people see both lawyers and psychologists as equipped to solve different problems by virtue of their training and experience.

This basic insight is not challenged, or even addressed, in this chapter. We address a different question: Does the training and experience of lawyers equip them to solve the same problem differently than experts in other professions? This may not be the theoretical focus that the reader expected. However, we believe that this question better addresses the central issue of this book: How are complex problems solved, and what are the underlying principles and mechanisms of such problem-solving skills? By examining whether lawyers and other groups of novices and experts solve the same problems differently, we can begin

*Now at the Department of Psychology, Vassar College.

**Faculty of Law, University of Saskatchewan.

***Now at the Department of Psychology, University of Guelph.