Taylor & Francis
Taylor & Francis Group

Complex problem-solving: a field in search of a definition?

J. QUESADA†*, W. KINTSCH† and E. GOMEZ‡

†Institute of Cognitive Science, University of Colorado, Boulder, Muenzinger Psychology Building, Campus Box 344, Boulder, CO 80309-0344, USA
‡Department of Experimental Psychology, University of Granada, Campus Cartuja, S/N, Granada, Spain

Complex problem-solving (CPS) is as an area of cognitive science that has received a good amount of attention, but theories in the field have not progressed accordingly. The reasons could be the lack of good definitions and classifications of the tasks (taxonomies). Although complexity is a term used pervasively in psychology and is operationalized in different ways, there are no psychological theories of complexity. The definition of problem-solving has been changed in the past to reflect the varied interests of the researchers and has lost its initial concreteness. These two facts together make it difficult to define CPS or make clear if CPS should reuse the theory and methods of classical problem-solving or on the contrary should build a theoretical structure starting from scratch. A taxonomy is offered of tasks using both formal features and psychological features that are theory-independent that could help compare the CPS tasks used in the literature. The adequateness is also reviewed of the most extended definitions of CPS and conclude that they are in serious need of review, since they cover tasks that are not considered problem-solving by their own authors or are not complex, but ignore others that should clearly be included.

Keywords: Problem solving; Dynamic decision making; Micro-worlds; Expertise

1. Introduction

This paper is concerned with the definition of complex problem-solving (CPS). We have been interested in the relationship between the definitions of complex problem-solving and classical problem-solving. Can one actually draw a line between them? Is it necessary to build theories of CPS from scratch or are CPS tasks just a particular branch of problem-solving that can inherit theoretical breakthroughs from the past? The following pages try to show that the definitions of problem-solving, complexity and complex problem-solving are not well specified. It is difficult to build good theories in a field where the basic definitions are blurry.

Two books with exactly the same title and one shared editor, 'Complex problem-solving' (Sternberg and Frensch 1991) and 'Complex problem-solving, the European perspective' (Frensch and Funke 1995a), address completely different sets of problems with an extremely small overlap: Funke is the only author who has a chapter in both books and talks about the same kind of problems.

*Corresponding author. Email: quesadaj@psych.colorado.edu

The first collection (Sternberg and Frensch 1991) covers topics as broad as reading, writing, arithmetic, managerial problem-solving, legal reasoning in lawyers, mechanics and electronics problem-solving and program debugging. One could ask what binds these areas together. A possible answer is that they share a common factor that is not theoretical but methodological: the use of expert-novice comparisons. This book offers no definition of complexity, problem-solving or complex problem-solving; indeed, it might be difficult to craft one that covers the variety of subjects and tasks that the book reviews. The authors' 'implicit agreement' is to call 'complex problem-solving' anything that is not 'simple' problem-solving, understanding by simple the kind of tasks that were used in the laboratories and offered limited insight in the kind of problem-solving that people exhibit in their real lives. The second collection (Frensch and Funke 1995a) focuses on a completely different set of tasks, using standard psychology students as participants controlling simulations that are dynamic, such as cities, forest fires and ecologies. This book dedicates a chapter to definitions of complex problem-solving.

The two books are complementary, since, although they share the title 'complex problem-solving', they could not agree in a definition for the term and use very different approaches. However, neither book defines the term 'complexity', nor does either supply any meaningful way of relating the different tasks or domains. For that matter, one cannot say whether one task is more complex than another except by using common sense and that makes the comparison of results and conclusions from one task to another almost impossible.

In the 1970s and 1980s, the definition of problem-solving was expanded to cover practically every cognitive activity other than perception and memory.† Essentially, every goal-oriented activity was considered problem-solving (Burns and Vollmeyer 2000). It is believed that this over-stretching impairs the definition's utility. At some point in the 1990s, both European and American researchers decided that a line should be drawn between simple and complex problem-solving. Both of them seemed to agree about what can be considered classical, 'simple' problem-solving: puzzles that can be solved by simple reasoning programmes using pure logic. The other side of the line is what remains to be clarified.

Since it is proposed that the definition of the field needs some revision, one should present at the outset what is considered to be currently the most widely accepted definition of complex problem-solving from the European perspective; without doubt, some researchers would not agree with it, although the majority would. To avoid confusion, a provisional version of this consensus definition is presented here, which basically packs the main features shared by the definitions in Frensch and Funke (1995b) into a more compact presentation format:

CPS tasks are situations that are: (1) *dynamic*, because early actions determine the environment in which subsequent decisions must be made and features of the task environment may change independently of the solver's actions; (2) *time-dependent*, because decisions must be made at the correct moment in relation to environmental demands; and (3) *complex*, in the sense that most variables are not related to each other in a one-to-one manner. In these situations, the problem requires not one decision, but a long series, in which early decisions condition later ones. For a

---

† Memory retrieval is considered by some as a search activity that can be related to problem-solving as well.

task that is changing continuously, the same action can be definitive at moment t_1 and useless at moment t_2. In contrast, traditional, experimental problem-solving research has focused largely on tasks such as anagrams, concept identification, puzzles, etc., that are not representative of the features described above.

CPS tasks that fit this 'European' definition are also known as micro-worlds. An example of a micro-world is Firechief (Omodei and Wearing 1995), which simulates a forest where a fire is spreading. The participant's task is to extinguish the fire as soon as possible. In order to do so, the participant can use helicopters and trucks (each one with particular characteristics) that can be controlled by mouse movements and key presses. There are three commands that are used to control the movement and functions of the appliances: (1) drop water on the current landscape segment; (2) start a control fire (trucks only) on the current landscape segment and (3) move an appliance to a specified landscape segment. Firechief generates complex log files containing every action a participant has done.

Another micro-world is Duress II (Dual Reservoir System Simulation), a thermal-hydraulic scenario that was designed to be representative of industrial processes. Duress II has two redundant feedwater streams that can be configured to supply water to either, both or neither of two reservoirs. Participants control eight valves and two heaters to satisfy the goals of the system: to keep each of the reservoirs at a prescribed temperature (e.g. $40°C$ and $20°C$, respectively) and to satisfy the current water output demand (e.g. $5\,1s^{-1}$ and $7\,1s^{-1}$, respectively). The system is determined by the laws of conservation of mass and energy, and its state can be described using 37 variables (energies, volumes, valve positions, etc.) that are saved into log files.

Both micro-worlds (Firechief and Duress II) are representative of the features described in the consensus definition of CPS. However, most problem-solving studies using complex, dynamic tasks like the two described above used a product measure only, because of the difficulty of measuring performance in these tasks and comparing measures in a quantitative way (Christoffersen *et al.* 1996, 1997, 1998, Cañas *et al.* 2003).

A strict methodologist could say that the study of problem-solving outside simple tasks—whether using expert–novice comparisons or micro-worlds—cannot be considered experimental psychology. On one hand, the expert–novice approach samples from different pre-existing populations of experts and novices, rather than assigning participants to groups randomly and then developing expertise; this makes it a quasi-experimental approach (e.g. Cambell and Stanley 1963). On the other hand, in the micro-world approach, participants are assigned randomly to groups. Yet, the stimulus events are no longer under complete control of the experimenter, but instead they are also influenced by the subject's own behaviour (e.g. Busemeyer 2002a). This defies the basic assumptions of experimental control: all participants in the same group should receive the same treatment. If the stimulus is no longer under the control of the experimenter, then it is harder to guarantee what is happening.

In this sense, most methodologists would agree that expert–novice comparisons are not any stronger methodologically than the micro-world approach: both are quasi-experimental approaches most of the time. Nevertheless, the two approaches have had very different impacts in the past on the mainstream cognitive science literature, which values methodological rigor. Expert–novice comparisons are represented by many articles and book chapters in virtually every area of research. By contrast, the micro-world approach is barely published in mainstream journals and

tends to be found mostly in German-speaking journals and chapters, which have necessarily a smaller audience. Of the talks selected for problem-solving sessions at the 2002 Cognitive Science Society meeting, only one out of 12 presentations used micro-worlds.

One explanation for this difference in the impact achieved could be the differing ability of the two paradigms to produce well-rounded, publisher-attractive results and conclusions. Micro-world researchers tend to have serious problems in formulating clear-cut conclusions. The impressive amount of data obtained in process measures makes them difficult to interpret. When researchers use individual differences to explain performance changes, the difficulty of measure personality, motivational and social variables adds to the trouble of analysing the micro-world itself. On the other hand, the expert–novice approach most of the time produces conclusions that are crystal-clear. It almost guarantees statistically significant results, because the groups compared (expert and novices) are very different and tend to perform very differently when confronted with similar experimental situations (Sternberg 1995).

To return to the (lack of) definition of complex problem-solving, a recent paper by Gray (2002) has discussed the difficulties in finding a suitable definition. The field seems to be consolidated, as is shown by the existence of an important corpus of literature, including books and review papers, and the self-identification of the authors involved in developing and testing theories of CPS. This is not surprising, as some of the proposals in the field are significant departures not only from the mainstream problem-solving approach, but also from the mainstream experimental psychology methods and assumptions. The facts that (1) their proposals include novel solutions to an old dilemma in general science (external validity vs experimental control) and (2) in current society, humans spend most of their time interacting with complex systems and more knowledge about this interaction is demanded, makes their arguments inherently interesting and relevant.

This paper reviews the work that tries to delimit what has been called complex problem-solving, control of dynamic systems, control of complex systems or system control which is thereafter called just Complex Problem Solving (CPS). One would like to show that this general tag is misleading, because not all tasks that are used in the literature contain all the features that are cited in the consensus definition. For example, most of the tasks used in implicit learning are not *dynamic* or *time-dependent*, features 1 and 2 of the definition, although they have been traditionally lumped together in the same category with CPS (e.g. Buchner 1995). Also, the next few pages try to show that feature number 3, complexity, is not well-defined itself and should not be used in the definition of the area. Worse yet, some tasks that fit the definitions of complex problem-solving used in classical works in the field (e.g. Frensch and Funke 1995b) would not be included in more general definitions of problem-solving and some tasks that would be called complex by any common-sense observer are not captured in the current definition of CPS. However, this is a point that one would like to revisit after presenting this review of traditions, definitions of complexity and possible task taxonomies, because the arguments will make more sense after this overview of the literature.

Note that this paper is not a review of the theoretical explanations that have been proposed for CPS effects. The authors want to take a step back and reflect about what can be seen as the essential first steps in the establishment of a scientific field: definition of the terms and classification of the tasks. Both of these processes can be

performed either in a theory-independent way or from the standpoint of a theoretical framework. If the theoretical work from inside the CPS community has been sparse, some general theories and architectures have been borrowed and extended to explain complex behaviour that initially was not under their scope. A good example is the approach of the unified theories of cognition, SOAR (e.g. Newell 1990) and ACT (e.g. Anderson and Lebiere 1998). However, this is not the main objective of the paper.

In what follows, first a survey of research traditions that have used complex, dynamic tasks is introduced: naturalistic decision-making, dynamic decision-making, system control in implicit learning and the European version of complex problem-solving.

The next section focuses on how—indeed, whether—task complexity can be measured and remark that one still does not have a worked-out task complexity measure.

The succeeding section presents the need for a task taxonomy and two possibilities towards their definition: using only formal characteristics or concentrating on psychological features of the tasks. Of all the features studied, none is necessary and sufficient to define a task as 'complex'.

The concluding remarks argue that calling the field 'complex problem-solving' is limiting the development of theory.

2. Overview of CPS traditions

The literature on complex problem-solving can be divided into four blocks or traditions:

2.1 *Naturalistic decision-making (NDM)*

This branch of decision-making, inaugurated by Zsambok and Klein (1997) and Salas and Klein (2001) has been centred in field research, for example interviewing firefighters after putting out a fire or a surgeon after she has decided in real time what to do with a patient. Klein's 'recognition-primed decision making' (Klein *et al.* 1993, Klein 1997), the most extended model, is based on the idea that, in real-life tasks, the current goal and the current state of the environment are used to retrieve from memory actions that have worked under similar circumstances in the past. This tradition does not use simulated environments or micro-worlds. Researchers in it have criticized mainstream theories of problem-solving and decision-making as being inapplicable to everyday settings. This is a very fragmented field, where each researcher works on a particular 'real-world' task and the generalizability of the results is questionable (no taxonomy of NDM tasks has ever been developed). Theoretical developments in NDM are sparse and computational modelling is almost non-existent. The main theoretical contributions have been achieved when some researchers successfully adapted to NDM existing concepts and models as research vehicles. For example, the classical model of memory and categorization MINERVA II (Hinzman 1988) has been used to implement Klein's recognition-primed decision-making paradigm (Warwick *et al.* 2001, 2002).

2.2 *Dynamic decision making (DDM)*

This tradition (Brehmer 1992, Sterman 1994, Busemeyer 2002a) centred in simulations with economical environments like stock purchases or inventory control (Sterman 1989) is fairy well integrated with standard decision theory and tends to display a high level of formalization. The multi-stage betting paradigm explained, for example, in Rapoport (1975) is an extension of the standard single-stage betting paradigm. The research methodology has inherited the normative models approach: the decision-maker's actions are compared to a mathematical model that is supposed to be optimal. Systematic discrepancies between the optimal model and the decision-maker's behaviour are included in the psychological model. Unfortunately, contrary to what is an extended practice in decision theory, formulating and computing an optimal strategy is problematic in complex tasks and is possible only in the simplest dynamic decision-making situations (Slovic *et al.* 1977, as cited in Kleinmuntz (1985)). Some authors like Toda (1968) have focused mainly on decision problems for which analytical solutions can be obtained. Most of the work in the area has been done in discrete dynamic decision tasks that change only when the participant introduces a new set of inputs. Variables like time pressure have been successfully integrated in models like Busemeyer and Townsend's (1993) decision field theory.

2.3 *Implicit learning in system control*

This tradition has used tasks like the sugar factory (Berry and Broadbent 1984) or the transportation task (Broadbent *et al.* 1986) that are governed by comparatively simple equations. The sugar factory (Berry and Broadbent 1984) uses a simple computer-simulated scenario in which the participants control the production of sugar (P) by manipulating the number of workers employed (W). The sugar factory operates according to the following equation:

$$P_t = 2W - P_{t-1} + e \tag{1}$$

where $1 \leq W \leq 12$ and $1 \leq P \leq 12$. The values of W are multiplied by 100 and the values of P are multiplied by 1000 to represent the number of workers. According to the equation, the output of the system at time t, P_t, is determined completely by the last sugar output P_{t-1} and the present input, W_t (workers employed by the participant). However, a random factor e ($+1$, 0, -1) is added to the final result of the equation to make the task more difficult and prevent the participants from learning the rule.

The amount of research on implicit learning is staggering; in the last few years the system control research has stagnated and two other competing implicit learning paradigms (with much less external validity, but maybe better suited to demonstrate learning without awareness) have prospered at its expense. These paradigms are artificial grammar learning and serial reaction time. However, the body of research on system control remains very interesting. Although most of the tasks lack definitive CPS features like dynamics and feedback delays it supposes the lowest complexity, simplest way of studying some CPS phenomena and it has been the preferred option for modellers, so it is of great importance in the discussion.

The most common effects to be explained are related to how people improve their control of the system, but cannot verbalize that knowledge. The theorization and computational modelling in this branch of CPS are extremely rich. Models are based

on exemplar learning, rule learning and both (e.g. Dienes and Fahey 1995, Gibson *et al.* 1997, Lebiere *et al.* 1998).

2.4 *European complex problem solving (CPS)*

Mainly practiced in Germany, this tradition uses by far the most complex tasks. Initiated by Dörner (Dörner and Scholkopf 1991, Dörner and Wearing 1995), the European CPS tradition is the most radical departure of the mainstream problem-solving tradition, with tasks of more than 2000 highly inter-connected variables (*Lohhausen*, Dörner 1975, Dörner *et al.* 1983). In the past years, theorizing in CPS has been mainly disconnected from the mainstream lines of theory in psychology. In *Lohhausen*, participants play the major in charge of increasing the welfare of a small city. In *Moro* (Dörner and Scholkopf 1991), they are in charge of the survival of a semi-nomadic African tribe that raises cattle and has a very primitive economy. The goals in these systems are deliberately vague: What does 'welfare' mean?

Other tasks that could be identified in this tradition are the ones based on fire extinguishing (e.g. Artman 1999, Granlund *et al.* 2001). Although objectives and performance are clearer in these tasks, they have been used to evaluate group decision-making and problem-solving, which implies that their task complexity rivals *Moro* and *Lohhausen*'s.

Tasks like *Lohhausen* and *Moro* have good reason to be isolated from the general development of psychological theory: they offer a completely different view that often falls out the safe, accepted experimental paradigm in cognitive psychology. However, a large number of tasks that have been considered complex problem-solving are nowadays affordable for theory development and computer modelling (e.g. Putz-Osterloh 1993, Vollmeyer *et al.* 1996, Burns and Vollmeyer 2002, Schoppek 2002).

The common idea of the approaches presented here is to transport real-life complexity to the lab in a way that can be partly controlled. These traditions share a common denominator: the stimulus events are no longer completely under experimenter's control, but instead are also influenced by the subject's own behaviour (Busemeyer 2002a).

In summary, various traditions within cognitive science use complex, dynamic situations as tasks. The authors have commented to what extend they have developed theories and what they have in common. However, to define an area, one not only needs a set of tasks that are more or less similar in their complexity; but also a way of defining complexity. The next section describes some proposed solutions for this issue.

3. Measuring complexity

There are many cognitive phenomena that need a complexity metric and problem-solving is one of them. 'CPS' takes its name from two features that are ill-defined: complexity and problem solving. (As Burns and Vollmeyer (2000) noted the definition of problem-solving has changed from very narrow 'essentially the study of solving the tower of Hanoi' to very broad 'every goal oriented activity'.)

This section will review different complexity measures and their importance and applicability to complex problem-solving definition.

   Complexity does not equal perceived difficulty. It is well known that changes in cover story using exactly the same problem structure lead to very different perceived difficulty (Kotovsky *et al.* 1985, Kotovsky and Simon 1990). Most of the tasks used in complex problem-solving are not particularly difficult; even more, participants tend to find them engaging (Gray 2002). What interests one in this section is the problem structure. Obviously, a common factor in CPS tasks is their complexity. However, what does one mean with that? What follows is a review of several ways to measure complexity.

3.1 *Computational complexity*

In complexity theory, the time complexity function for an algorithm expresses its time requirements by giving, for each possible input length, the largest amount of time needed by the algorithm to solve a problem instance of that type (Garey and Johnson 1979). Problems that have an exponential time complexity are called *intractable*, or *NP-complete*. The traveling salesman is a popular example of an intractable problem† where one must find the path linking a set of cities that minimizes the distance travelled. In this problem, all known solutions take an amount of time that is exponential to the number of cities. Some other planning problems are also intractable. Can computational complexity theory be a good measure of complexity for CPS tasks? There are two main reasons to think otherwise. First, many computationally 'complex' problems do not fit the consensus definition of CPS and in fact are not perceived by humans as 'complex'. It is well known that people heuristically solve NP-complete problems (e.g. Taatgen 1999: chapter 3, Dietrich and Markman 2000). Indeed, people can solve NP-complete problems without complaining about difficulty and get stuck with problems that would be easily solvable for a computer. For example, Taatgen (1999: chapter 3) presents a scheduling task like this: there are two workers, each having 6 h. Task A takes 1 h, task B takes 2 h, task C takes 3 h and task D takes 3 h. How do you assign the tasks to the workers so that the following constraints are met: A before B, A before D and C before E? Following Garey and Johnson (1979: 239), one can prove that this task is NP-complete, but Taagen's participants were perfectly able to solve it. Secondly, even though in many CPS tasks a 'best' solution cannot be analytically derived, they are not necessarily NP-complete or intractable in computational sense‡ or at least the authors have not found any published proof of NP-completeness for those tasks. Note that, although NP-complete, the scheduling task does not fit in the consensus definition of complex problem-solving. At the same time, logic tasks like the Wason's (1968) selection task are extremely easy for computer algorithms but cause trouble to human participants.

---

† Problem difficulty was initially thought to vary with the size of the problem space. However, some problems are hard even if the problem spaces are small. For example, the Chinese ring puzzle, which requires freeing a metal bar from a series of rings, has an unbranching problem space, so that the solution can be reached simply by never reversing the direction. Nevertheless, this problem, with seven rings, is seldom solved in less than 30 min.
‡ However, Klein *et al.* (1993) have said that most complex, dynamic decision-making tasks are analytically intractable.

3.2 *Relational complexity*

Relational complexity (RC, Halford *et al.* 1998) is defined as the number of unique entities that one must process in parallel to arrive at a solution. Halford *et al.* propose that human adults cannot process without segmentation a relation that has more than four arguments. Halford *et al.* call this the problem dimensionality, four relations being *a processing capacity limit* that could conceptually steal the spotlight of number of chunks or amount of activation as determinants of working memory size.

Their idea is that some problems are complex because numerous elements are related to each other and cannot be considered meaningfully in isolation. These relations have to be at least partially processed in order to segment the main problem into sub-problems that can be processed serially. Complexity is expressed as the number of arguments of a relation. For example, unary relations have a single argument like in class membership: human (Bill). Binary relations have two arguments as in chases (cat, mouse) or correlates (temperature, heater setting). An example of ternary relation is ratio (distance, time, velocity) (Birney and Halford 2002). A quaternary relation can be proportion $a/b = c/d$. More complex relations may of course exist, but are, in the authors' view, less interesting, as human beings tend to be very limited in solving them.

Owing to high connectivity between variables, CPS tasks may require processes with high relational complexity. This proposal is really interesting, since it measures not the complexity of a task in itself but the requirements that the task imposes on the cognitive system. For example, imagine a simulated ecological system where more than four predator–prey relations are in equilibrium so that any variation in the population of one of the species would affect the rest. A successful participant in this task should handle at least quaternary relations in his head. Situations like that are common in micro-worlds like *Moro* (Dörner and Scholkopf 1991) or *Lohhausen* (Dörner *et al.* 1983). In Firechief, the direction in which the fire will spread can be estimated, but participants need to take into account wind direction and strength, flammability of cells close to the fire, location of control fires, relative humidity and more. In Duress, the volume of water contained in Reservoir 1 depends on the settings of at least five valves (A, A1, B, B1 and O1, see e.g. Christoffersen *et al.* 1998). The system used by Schoppek (2002) consists of four input and seven output variables, highly inter-connected. Participants controlling the system would very probably have to process relational instances of four or more arguments.

In systems like the ones cited above, the reducibility hypothesis (Lee and Anderson 2001) would appear not to hold, since by definition there is no way of decomposing quaternary relations without losing information. However, this is just a conjecture, as the authors did not analyse the complexity of the tasks cited in Halford *et al.*'s (1998) terms.

Basically, Halford *et al.*'s (1998) proposal was directed to *measuring working memory* capacity when performing different tasks, with some emphasis on developmental effects. However, they also showed that the concept of relational complexity is relevant to areas like transitive inference, problem-solving and sentence comprehension. In problem-solving, for example, they showed that the Tower of Hanoi can be solved without processing steps more complex than a quaternary relation (Loveday 1995, as cited in Halford *et al.* (1998: 823)). Some of the criticisms they

received cast doubts on the usefulness of the working memory capacity measurement idea. For example, Sweller (1998) argued that it is difficult to measure working memory without contamination from knowledge held in long-term memory. He argued that what constitutes a dimension will depend on the expertise of the problem-solver. An expert chunks together, as a single recognizable configuration, what a novice sees as separate dimensions. Such effects might lessen the utility of the relational complexity measure in knowledge-rich tasks like *Lohhausen* or *Moro*. As well, there is a pragmatic issue in this approach to complexity: the task has to be decomposed and the number relations detected in a purely analytical way and that is subject to the same criticisms discussed above.

3.3 *Size of the problem space*

Size of the problem space is one of the most obvious candidate measures of complexity. However, the size of the problem space is not, directly, a good measure of problem complexity, for at least two reasons.

3.4 *Having more decision alternatives does not hinder performance*

Bigger problems are normally characterized by a richer fan of options at any point of the problem space. Research on decision-making paradigms has showed that performance is not adversely affected by an increase in the number of alternatives (also known as the branching factor in the problem-solving literature) and seems affected only to a limited extent by an increase in attribute-based complexity. For example, Hogarth (1975) described a model of how people make decisions when the number of alternatives is increased. In this case, complexity is defined purely as number of alternatives, not as number of dimensions to be compared or difficulty of discriminating between alternatives. He showed that there is a non-monotonic relationship between time needed and complexity of the task, increasing with the number of alternatives until a certain maximum, after which people actually take less time in making a decision for more complex decisions.

Bruce and Johnson (1996) extended these results to the complexity in the set of dimensions to be compared, finding that performance was not significantly affected by the increase in this form of complexity.

Although these results come from non-dynamic, one-step decision-making paradigms, there are reasons to believe that similar results could be expected in multi-step paradigms.

3.5 *The number of states visited is always smaller than the total number of states available in a system*

The number of states visited is always smaller than the total number of states available in a system, so the sheer size of the system is not necessarily a good predictor of the size of the sub-set of visited states. For example, Firechief has 2 560 000 possible 'move' actions, but the total number of actions used by more than 200 real participants at several hours each is only 57.565, a difference of an order of magnitude. Buchner *et al.* (1995) reported that, in the sugar factory task, good controllers observe fewer different state transitions, only around the areas of the target values that needed to be obtained. Vollmeyer *et al.* (1996), using a similar but more complex

system called Biology Lab† did not measure the size of the problem space or the proportion of the space visited, but they reported that participants who explored the system without goals performed better, perhaps because they sampled more of the entire space; participants with concrete goals did not learn as much. Note that these results are explainable only if the state space visited is always less than the total available.

Note that Vollmeyer *et al.* (1996) did not measure the size of the problem space or the proportion of the space visited; however, Schoppek (2002) did. He assumed that the problem space of such systems can be represented by finite directed graphs. He reasoned that any variable that the participant can change form an 'input' vector and any variables that are changed form an 'output' vector too. To calculate the number of links in a problem space, the number of possible output vectors must be multiplied by the number of possible input vectors (because each possible input vector can be applied to all possible output states). The number of different input vectors equals the number of variables times the number of possible values of each variable, since input variables are independent. The number of output vectors are not that easy to calculate. Although conscious of that there might be better alternatives, Schoppek proposes this measure of the problem space is as a proxy for complexity. In small systems like the sugar factory, instance-based learning would be enough to control the system, as has been repeatedly shown in the literature (e.g. Dienes and Fahey 1995, 1998, Lebiere *et al.* 1998). Schoppek assumes that the instance knowledge acquired is based on triplets Output–Input–Output (what would be called a relational instance in Halsford's terminology). The probability that a subject would retrieve an appropriate triplet Output–Input–Output (O-I-O) in a given situation would depend highly on the size of the system. In bigger systems this instance knowledge would be far less important and the real determinant of performance would be structural knowledge: participants would have to learn the rules that govern the system.

One has two criticisms to Schoppek's measure. First, while some systems from the CPS tradition like the ones he used can be represented by finite directed graphs, one doubts that this representation is helpful (or even possible) for some other CPS systems. The directed graph generated by a snapshot decision in Firechief would be branchy enough to be useless. On continuous systems like Duress II, the representation is not even possible. If the continuous values are discretized, the idea would be applicable, but the graph would be distressingly large. Secondly, he assumes that the probability of retrieval of the relational instance required to act correctly in the system at time *t* depends highly on the size of the system. Schoppek uses the size of the problem space as an estimation of the usefulness of O-I-O knowledge for acting on a specific system: in 'big' systems, O-I-O knowledge would be less useful than structural knowledge. This assumption does not necessarily hold, as the literature on expert performance shows. Chess players can recognize tens of thousands of configurations and moves associated with them, doing retrieval in extremely short times with great accuracy (e.g. Chase and Simon 1973).

---

† Biology Lab has four input variables (temperature, salt, oxygen and current) that are connected one-to-many to four output variables (prawns, sea bass, lobster and crabs). The objective of the task is to reach a particular level of the output variables.

In summary, computational, relational and size-based definitions of complexity do not exhaust the meaning of complexity: CPS is still an ill-defined term. The relational complexity measure could be the beginning of a psychological theory of complexity, but, although widely used in developmental psychology, it is still not much used in studies of problem-solving and thinking. At this moment, it has been applied only to the Tower of Hanoi (Loveday 1995, as cited in Halford *et al.* 1998) and to the reasoning task of knight-knave (Birney and Halford 2002).

In some disciplines, the object of study cannot be defined clearly, but its manifestations can be classified in a scientifically useful way. It could be the case that CPS tasks are susceptible to meaningful classification. The next section explores this point.

4.   A taxonomy of CPS tasks

There is no single feature that is necessary and sufficient to identify a task as 'complex problem-solving'. This is unfortunate, because a theory that intends to explain and predict behaviour in CPS tasks should be targeting those features that make the task different from classic problem-solving. Perhaps what is needed is not a binary distinction but a whole taxonomy.

Why does one need taxonomy of CPS tasks? Is there any existing taxonomy of classic problem-solving tasks? Burns and Vollmeyer (2000) argued that there is a proliferation of experimental tasks which renders generalization difficult and that there is a set of phenomena, but not an underlying explanation of them. The underlying problem is generalization: Is what is learned in simple tasks generalizable and valid in more complex ones?

Up to this point, the Tower of Hanoi and similar puzzles that are the signature phenomena of Soar (Newell 1990) have been referred to as 'classic problem-solving tasks'. However, the authors do not want to draw a line between classical and complex problem-solving as if they were different species: there is no such a clear distinction. Some of the tasks that wear the label 'complex' are in fact formally simpler than some of the tasks that are considered 'classical'. For example, the sugar factory task (Berry and Broadbent 1984) is as simple as a linear equation relating three values (plus some random noise); yet it has been used as a prototypical example of a complex task. The classic Tower of Hanoi task may be formally at least as 'complex' without being so labelled.

The 'theory for classic problem-solving' is the one presented by Newell and Simon (1972) and expanded in successive versions of Soar, where the main idea is that problem-solving is search in a problem space. One has been interested in trying to explain when this theory is not successfully applicable to these tasks and why (Quesada *et al.* 2002). However, it has been argued that the classic problem-solving theory is not a complete theory at all but a language for thinking about problem-solving, since it makes no testable predictions (Burns and Vollmeyer 2000).

The need for a task taxonomy has been recognized in the literature for a long time now (e.g. Reitman 1964). One needs a taxonomy to assess the similarity between tasks and to generalize results from one type of task to another. A taxonomy of CPS tasks would show what behaviours are common over a large set of complex problem-solving tasks. Another advantage of a taxonomy is that it would help in decomposing complex systems into sub-modules. Although there is no guarantee of system

decomposability and most probably a feature of these really complex tasks is that they are not decomposable, it is a good idea to try to break down the system into independent components as a first step (Hunt 1991).

There are two basic ways in which one might classify complex problem-solving tasks: by their formal features and by their psychological features (that is, the processes and representations postulated in participants facing the task).

4.1 *Formal descriptions*

In applied settings like engineering problems, control of complex systems is a very popular topic and systems theory has developed vocabulary and concepts that can be useful to classify the tasks that psychologists have used. This section is based on some distinctions made in Vemuri (1978: chapter 2) and in Busemeyer (2002b). When dichotomic features are proposed in this list the more complex item is addressed first.

4.1.1 Time related

Time variant–time invariant (dynamic vs static systems): in static systems changes occur only whether participant intervenes. For example, in the Tower of Hanoi, disks do not move by themselves unless the participant acts on them. Other systems have at least 'an endogenous variable [that] at time t has an effect of its own state at time $t+1$ independent of exogenous influences that might add to the effect' (Funke 1993: 322). This feature has been called 'eigendynamik' or inner dynamic or simply dynamic. There are two kinds of dynamic systems: *event driven*, where the system changes in discrete steps, sometimes linked to simulated decision cycles (e.g. MORO, Dörner and Scholkopf 1991) and *clock driven*, where the events are triggered by an internal clock and some times are real-time simulations (e.g. high-fidelity flying simulators). Most event-driven systems are discrete-time systems and clock-driven systems tend to be continuous-time systems.

Continuous time–discrete time: since the dominant metaphor in cognitive modelling is the digital—discrete—computer and tools like differential equations are not commonly used in psychology, continuous-time systems tend to be discretized at some point for modelling. Tasks that are changing continuously are pretty common in real life, since the environment changes in real time. Driving a car is an example. Duress is a task that changes in real time. Some other tasks, like Biology Lab, change in discrete steps.

Degree of time pressure: Duress may vary in real time, but its participants experience no time pressure, as the system moves very slowly. In some other cases like the KA-ATC task or Firechief, the time pressure is intense and decisions have to be made quickly. Both types are representative of different real life tasks. Industrial processes, business actions and inter-personal relations may change very slowly, while other problem situations require rapid decisions.

4.1.2 Variable related

Number and type (discrete/continuous) of variables: some systems cannot be well described using number of variables. For example, Firechief uses a landscape of 400 cells, each one with different values of flammability, current fire size (if any),

humidity, presence or absence of control fire, etc. It would be misleading to assume that the humidity value of cell (1,1) in Firechief contributes to the overall complexity of the task as much as the 'amount of oxygen' does in Biology Lab, a system with only eight variables. In this case, the sheer number of variables does not seem to increase complexity linearly. Worse, two systems with the same number of variables do not need to be similar or even comparable. For example, a system with 16 dichotomic variables can be radically different from a system with 16 continuous-value variables. For a classification of German complex problem-solving tasks using number of variables, see Funke (1991).

Number and pattern of relationships between variables: again, a system with only eight variables highly inter-connected can be more complex than a system with double that number, but sparsely connected. Note that this refers only to the number and not to the kind of relationships, which is treated in the next point. The pattern of connectivity can be a determinant feature of the complexity of the system, as has been shown, for example, in Howie and Vicente (1998). Some systems are more complex because certain connectivity patterns are more difficult to detect and hide causal relationships. Kenny (1979, as cited in Judd and Kenny 1981) posits three conditions for asserting causality: (1) covariation, (2) time precedence of cause to effect and (3) lack of spuriousness, where spuriousness is defined as covariation that is due to shared common causes. To explain how the connectivity pattern can affect the detection of relationships between variables, we can use figure 1. Imagine that, at some particular time, the participant must figure out the relationship between variables A and B. In figure 1(a), A and B are related through a third variable C. If C is a hidden variable (not available in the system interface, represented as a shaded circle), then the participant could easily and wrongly infer that A and B are causally related if each time that she manipulates A, B is affected (figure 1(b)). In figure 1(c), the variable C affects both A and B but not simultaneously (first A, then B); if C is hidden, then the participant could infer that A causes B, again wrongly. Note that the intervening variables between A and B could form a long chain, making the discrimination difficult. Any of the variables C, D, E and F could be a hidden variable (figure 1(d)).
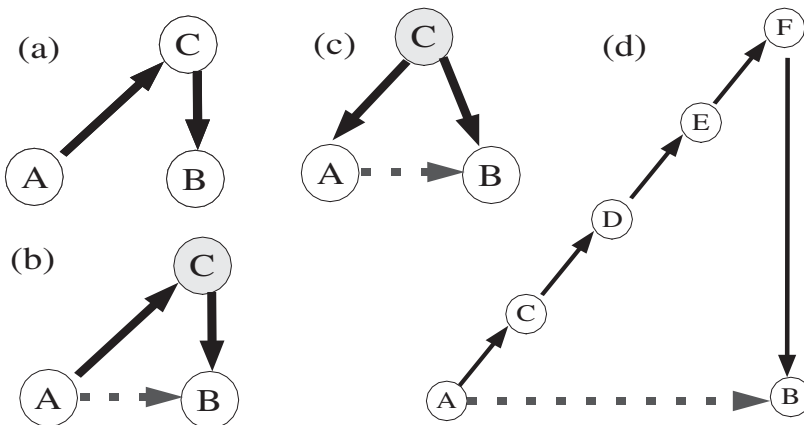

Figure 1.   Connectivity patterns that can make difficult the detection of causal relations in a system. Dotted lines represent indirect causal inferences that could be made observing the system.

Non-linear–linear: a system is said to be linear if the inputs and outputs are directly proportional over the entire range of measurement. Non-linearities are harder to model. Mathematically, one can estimate a solution for large linear equation systems, but it can be anywhere from very difficult to impossible to solve an equation system that contains non-linearities. Humans are known to be not very good at estimating non-linear relations. For example, Dörner (1989, Dörner and Scholkopf 1991) showed how both university students and experts tend to underestimate exponential developments. DeLosh *et al.* (1997) trained participants in a particular interval of linear, exponential and quadratic functions and asked them to extrapolate (predict) values to an interval where they had not been trained. Although participants captured the general shape of the assigned functions, notable individual differences were observed, as well as some systematic deviations particularly in the quadratic and exponential conditions.

This can be an interesting criterion for classifying CPS tasks; most of the control tasks in the implicit learning literature (e.g. Broadbent 1977, Broadbent *et al.* 1986), like some other control tasks based on equation systems (e.g. Funke 1992, Vollmeyer *et al.* 1996, Burns and Vollmeyer 2002, Schoppek 2002) are purely linear. However, other popular CPS tasks like Duress (Christoffersen *et al.* 1996), Firechief (Omodei and Wearing 1995), Lohhausen (Dörner *et al.* 1983) and Moro (Dörner and Wearing 1995) contain plenty of non-linearities.

4.1.3 System behaviour related

Opaque–transparent: an opaque system contains a layer of hidden variables that are never directly affected by the input variables, but have an effect on the output variables. Such a system never shows a complete portrait of its state. This feature seems to be very important in several definitions of complex problem-solving (Frensch and Funke 1995b: 14).

Stochastic–deterministic: in a stochastic system some events are triggered randomly, not related to any of the changes in the state of the system. When all the relations between variables are deterministic and all the events are triggered in a deterministic way, the system is deterministic. For example, one can imagine a trial of KA-ATC where the type and position of the planes is determined randomly. A system can also be stochastic if two or more variables are related using random noise. For example, the sugar factory task adds randomly 1, 0 or −1 in the function that relates workforce and sugar production. It seems that participant's effectiveness can be reduced by as much as half when a stochastic variable is introduced in an otherwise completely unmodified control task (Rapoport 1975), but these results are still unclear and under discussion (Kleiter 1975).

Most 'classic problem-solving tasks' are deterministic, while most complex problem-solving tasks are not deterministic. That is, in CPS tasks, taking the same action in the same state on two different occasions may result in different next states and/or different reinforcement values (Kaelbling *et al.* 1996). Nevertheless, if one declared non-determinism as 'conditio sine qua non' for CPS membership, then one would leave out tasks that have been traditionally called problem-solving, for example function learning, troubleshooting and problems in the spirit of the Tower of Hanoi and the rest of the signature phenomena of Soar. Some of the traditional implicit learning tasks in the system control literature (Broadbent *et al.* 1986, Lee

and Vakoch 1996) are deterministic and static independent decisions. A related taxonomy in dynamic decision-making, using criteria such as when previous decisions affect the environment, can be found in Edwards (1962).

Delayed feedback–immediate feedback: this is a very important point that has relevant consequences for modelling. Weak methods† depend on immediate feedback. One is referring to the explicit, formal feedback that can be obtained either directly (by reading the overall score at the end of the trial) or indirectly (by interpreting colour schemes or other indexes of performance that are absolutely evident). For example, if a truck is burnt out in Firechief, the participant can infer that he was not doing well in that particular action. The key point for this feature is whether the feedback can be traced back to the action that originated it or whether, on the contrary, there is no way of mapping each bit of feedback to particular actions. If each action can be mapped to a value of feedback, the system complexity is lower than if the actions and feedbacks are impossible to relate. Brehmer and Allard (1991) and Brehmer (1995) have shown that feedback delays seriously impair the performance of even informed participants.

As a summary of previous discussion, table 1 shows three possibilities for interactions among the criteria formally describing the problem. The table describes three possible complex problem-solving situations where the controller (human or model) acts in three possible idealized tasks, representative of some of the ones used in laboratories today. It is assumed that states and actions can be identified and, for clarity sake, they have been arbitrarily numbered. The interaction between the controller and the environment is depicted as a conversation.

The first column could be exemplified by simple tasks like the sugar factory. Following the formal taxonomy, the sugar factory would be a discrete time, time invariant (static) task. The number of variables is just two, the relationship is linear and they are both discrete with 12 values each. The system is transparent, although the dependency of the current state on previous amounts of sugar produced can be difficult to detect. It is stochastic, since there is a random error term. The feedback is immediate, since the error associated to each input value at any point can be calculated by subtracting the output value from the target value. Some implicit learning systems, like the one used in Lee and Vakoch (1996), are even simpler, because their equations do not include previous states (decisions are independent) and are deterministic (no random error term).

In a next step of complexity, not depicted in the table, one could find systems like Biology Lab and Linas (Putz-Osterloh 1993, Vollmeyer *et al.* 1996, Schoppek 2002) that contain more variables and equations and even inner dynamics, but do not change in real time. Feedback can be easily calculated and they tend to be transparent (no hidden layer of variables).

The middle column could be exemplified by multi-stage decision-making paradigms with time pressure, for example Pipes (see e.g. Gonzalez *et al.* 2004) in

---

† Weak methods (Newell 1980) are those used by problem-solvers when little is known about the task. As their name indicates, they are not very powerful, but often suffice to solve a problem if the space is not large or if extensive search is undertaken. An example of weak method is 'Generate and Test': Generate in any way possible (e.g. systematically or haphazardly) a sequence of candidate states, testing each for whether it is the desired state.

Table 1. Possible CPS situations (adapted and extended from Kaelbling *et al.* 1996).

	Static, transparent	Dynamic, transparent	Dynamic, opaque
Environment	You are in state 65. You have four possible actions	You are in state 65. You have four possible actions	You are in state 65. You have A_n possible actions
Agent	I'll take action 2	Uhm, I'm thinking. Can I wait?	Uhm, I'm thinking. Can I wait?
Environment	You received a reinforcement of 7 units. You are now in state 15. You have four possible actions	No! You received a reinforcement of 7 units. You are now in state 15. You have four possible actions	No! You received a reinforcement of R_{n+1} units. You are now in state 15. You have A_{n+1} possible actions
Agent	I'll take action 1	I'll take action 1	I'll take action 1
Environment	You received a reinforcement of -4 units. You are now in state 65. You have four possible actions	You received a reinforcement of -4 units. You are now in state 65. You have three possible actions	You received a reinforcement of R_{n+2} units. You are now in state 65. You have A_{n+2} possible actions
Agent	I'll take action 2	I'll take action 2	I'll take action 2
Environment	You received a reinforcement of 5 units. You are now in state 44. You have four possible actions	You received a reinforcement of 5 units. You are now in state 44. You have five possible actions	You received a reinforcement of R_{n+3} units. You are now in state 44. You have A_{n+3} possible actions

the condition of immediate feedback. Pipes simulates a water distribution system with deadlines for different tanks in the system. Decision-makers have to decide when to activate or deactivate pumps with restrictions on the number of pumps working at any time. There are 22 reservoirs with two pumps that can be connected or disconnected. That makes a total of $22 \times 2 \times 2 = 88$ possible actions at any time. The system to give immediate feedback in some experimental conditions, that is after each of the deadlines is passed, the system displays a number with the amount of water missed, which informs the participant about his performance.

The third column of the table shows dynamic, opaque systems where the number of possible choices is not known and so cannot be exhaustively searched. The amount of feedback that corresponds to each action is unknown as well. An example of these more complex systems is Firechief. Firechief is, according to the taxonomy, a time variant task that changes continuously in real time, with time pressure. It has a high number of variables, both continuous and discrete, as it develops spatially. Variables are related one-to-many and the basic equations that describe transformations in the cells according to the wind intensity, flammability, humidity, etc. are non-linear. Some information between the input and output variables is inaccessible to participants, so the system is opaque and it has some stochastic characteristics. Feedback is mainly delayed, as it is not easy to map the consequences of actions in a one-to-one way. Systems like Duress, Moro and Lohhausen would be described in this third column. Columns two and three are dynamic, fulfilling the criteria of the European consensus definition.

The distinctions made in table 1 are important, because the mechanisms needed to control these systems might be very different. Some models of cognition depend on immediate feedback or at least on feedback that is easily traceable back to actions. A system from column 1 can be controlled by very simple models. A system from column 3 would be very difficult to control even for a very advanced model. When such a model is developed, it tends to be extremely fine-tuned to the system, to the point that it can barely be considered a model of general cognition.

In summary, although there is a well-defined set of formal characteristics in which different CPS tasks vary, one would like to emphasize that none of the nine criteria is either necessary or sufficient to declare a task complex.

4.2 *Psychological descriptions*

Any psychological task description would depend on the theory of problem-solving. In most papers using CPS tasks, the cognitive processes to be used are not described, mainly because the theory has not reached that level of specification. Other fields of psychology, with perhaps better developed theories, do have meta-theoretical taxonomies based on psychological criteria (i.e. Humphreys *et al.* 1994 proposed a classification of memory tasks using data structures and access processes). However, the tasks used in the problem-solving literature are much more varied than the ones used in memory research. The experimental tasks used in Humphreys *et al.* (1994) were all based on word lists as inputs and the outputs could be a word or a decision. Problem-solving tasks can accept numerical values, mouse movements and clicks and virtually every activity that makes changes in the state of the tasks. Nor is the output set smaller or less varied. The division into computational primitives (data structures and processes) is also more difficult. To find common components between tasks like chess and the sugar factory or, for that matter, the real-life

problem of fixing a broken clothesline (Anderson 1990: 232–242) the analysis would need to descend to very molecular units of analysis or define a vocabulary almost infinite. A reduced goal would be to classify only the experimental tasks used in the literature of problem-solving, according to psychological criteria that tend to be valid independently of the theoretical framework proposed. The following is a list of those criteria.

4.2.1 Skill-based vs planning-based. This dimension could also be called reactive vs predictive. Planning-based tasks have the common feature that future states can be anticipated and this helps the participant designing a course of action. Other tasks do not have this feature and the agents have to react to the immediate circumstances to perform well. A similar description, widely used in the human factor literature, can be found in Rasmussen (1985). Rasmussen proposed that, when solving a task, an agent switches between three levels of control: (1) The skill-based level is automatic, unconscious and based on sensory-motor associations for well practiced tasks (skills); (2) for slightly more complex situations, a rule-based level involves a conscious application of a rule or procedure obtained by past experience or training; and (3) for unfamiliar or complex situations, the Knowledge-based level is used, involving planning and goal-driven deliberation of action at a conceptual level. In this view, task could be classified according to how demanding they are at each of these levels.

Examples of purely planning-based tasks are chess and the Tower of Hanoi (it is not an accident that they have been the basic research vehicles in the classic approach to problem-solving). Since in the tasks there are no time pressure changes in real time independent of the actions of the users and the time per decision is (virtually) unlimited, there is no advantage in performing actions very fast. In other words, the fact that a participant can perform actions very fast and/or allocate attention very quickly does not assure good performance.

Tasks where the speed-up of individual components (e.g. actions with keyboard and mouse) is enough to warrant an improvement in performance can be called reactive or skill-based.

Examples of a tasks that are mainly reactive are the AK-ATC task described in Lee and Anderson (1997, 2001) and the AMBR-ATC task described in Deutsch and Benyo (2001). This task is slightly more complex than AK-ATC, although still very simple compared to the real ATC task. It uses a graphic display where different planes enter or leave a sector of the ATC space where the participant has to assign them different tasks. It was designed to evaluate multi-tasking hypotheses. Both tasks require speed-up in perceptual, attentional and motor components to increase performance. The strategic, planning-based component is present, for example in assigning priorities for the different sub-tasks, but seems to be less relevant to performance than are the perceptual and motor skills.

Some tasks that fit the definitions of complex problem-solving used in classical works in the field (e.g. Frensch and Funke 1995b) would not be included in more general definitions of problem-solving. The Space Fortress (Mané and Donchin 1989) has an important skill-based component, but there are some parts that are strategic. Neither Firechief nor Duress is reactive. Non-reactive tasks tend to be more complex and perceived as more difficult in the sense that participants consider them an effortful cognitive activity.

4.2.2 Knowledge-lean vs knowledge-intensive.   Although this distinction appears to concern the tasks themselves, AI models have also been commonly divided using the same criterion into 'experts' and 'generalists' models. Expert models are applied to a very narrow area of study where they have lots of knowledge; they basically apply stored patterns with very little reasoning or combination of their representations. On the other hand, generalist systems have very little specific knowledge and base their power on the ability to represent and 'reason out' any domain using general rules (sometimes pure logic). The generalist systems have been applied basically to knowledge-lean tasks. In the late 1970s, the interest of the problem-solving community shifted from knowledge-lean to knowledge-intensive tasks, like for example algebra, physics, thermodynamics, chess, bridge, geometry, medical diagnosis, public-policy formation and computer programming (VanLehn 1989).

Newell (1990: 102), talks about two different types of processing: (1) deliberating to analyse the situation, possible responses, consequences, etc. and (2) retrieving knowledge that is already prepared. These two processes are depicting the same distinction that is discussed here. The Tower of Hanoi task requires not much knowledge other than the basic rules and operators described in the instructions to participants. This is a way to reduce the variability among participants: even an alien who shared no knowledge with a terrestrial human would be able to perform. However, the alien would have a hard time interpreting the function of a clothesline and, thus, with the problem of fixing it.

Note that, although they are predominant in the literature, the typical knowledge-intensive tasks described above are not included in the European definition of complex problem-solving. On the other hand, the tasks that do fit the definition (micro-worlds) can be performed by anybody and do not necessarily require previous knowledge of, for example, physics, medicine or algebra. A good number of them use abstract cover stories to avoid contamination from previous knowledge (e.g. Funke 1992, Schoppek 2002). Some others tasks like Firechief, Duress and Moro do use realistic settings, but it is assumed that participants will not have differential previous experience co-ordinating vehicles to extinguish fires, controlling thermodynamic systems or guiding humanitarian programmes to help third world countries. Also, the amount of information provided in the instructions tend to be sufficient to start working with the system. When the system is not completely defined in the instructions, the task of participants is to learn the missing bits of information that would enable them to perform well.

The knowledge-lean vs knowledge-intensive distinction can be interpreted in a different way, as the amount of knowledge that is required to operate the task and solve the problem and that was *not* there before starting the experiment. In this sense, most micro-worlds are knowledge-intensive tasks: people tend to learn a lot of new knowledge during long periods of controlling the system. On some tasks that are designed not as problem-solving test benches but as skill-acquisition environments, like the simplified ATC tasks, participants do not normally acquire new knowledge, but instead fine-tune the knowledge that was given in the instructions.

4.2.3 Learning vs no learning during problem-solving.   Newell and Simon's initial theory of problem-solving assumed that there is little or no learning during problem-solving (VanLehn 1989). Although this is an idealization, it enabled the creation of theories that ignored learning. On the other hand, if there is a pervasive effect in psychology, it is learning. Some experimental tasks can be conceived as learning-

independent: for example, 'one shot' decision problems that are performed in a single opportunity. Some decision-making tasks offer participants two or more options, but each episode is completely independent of the rest. This has been called the gambling paradigm. In those situations, practice effects are not in the experimenter's focus. However, many researchers would argue that problem-solving should be defined as a multi-step activity and that single-decision situations are not problem-solving.

Conversely, many multi-step tasks can be analysed under the idealization that learning is not taking place during problem-solving. For example, tasks like theorem proving, missionaries and cannibals and the Tower of Hanoi were explained by the initial theory of Newell and Simon under the no-learning assumption. The no-learning assumption could also be maintained and learning could be considered accessory when participants are working in an area where they are experts. Theories that seek to classify problem-solving situations should be able to make statements about this particular point.

4.2.4 Understanding-based vs search-based problems.

VanLehn (1989) argues that problem-solving behaviour has two components: understanding and search. Problems or tasks could be classified in function of the importance of these components. Reitman (1964) claimed that the difficult part of some problems is not search, but formulating the problem to begin with. He added that 'problems exist in the head (or more generally, memories) of particular problem-solving systems' (Reitman 1964: 289). In this view, representing a problem is a key part of solving it and the representation is the end result of comprehension processes. For example, in painting a picture or writing a fugue or a novel representation and comprehension are more important than search. A completely formal representation of these domains has been never achieved by any problem-solver. A well-structured problem is one that has been formalized by the problem-solver (Simon 1973). In that sense, the formalization process is a very important part of creating a problem representation and can be considered comprehension. Traditional tasks like the Tower of Hanoi are simple enough to be represented with negligible differences by different problem-solvers. While in these contexts the importance of search is evident and most of the theoretical weight has been carried on that assumption, the division between comprehension and search may be a key to classifying problems.

VanLehn (1989) argues that understanding does not run to completion before search begins, but the two processes alternate and even blend together. No studies were found oriented toward discerning what part of the processing time is dedicated to comprehension and what part is employed in search. However, that would be among the assumptions that particular theories of problem-solving would have to make. A common and strong meta-theoretical assumption is that the comprehension process yields a problem space. The mechanisms by which the problem space is generated and how it is represented are particular to each theory.

4.2.5 Decomposability under human constrains without information loss.

Newell and Simon maintained that most tasks are decomposable into smaller sub-problems, an idea that is at the heart of their problem-solving theory. They proposed Elementary Information Processes (EIPs), a concept since used by some other researchers. Payne *et al.* (1993) used Elementary Information Processes to calculate the mental effort of each decision strategy, such as reading an item of information, comparing two items

of information, multiplying or adding items of information, eliminating items of information, etc. If every single task could be divided using the same vocabulary of elementary information processes, the tasks could be classified in complexity and difficulty. However, it is not clear that there exists a set of Elementary Information Processes wide enough to describe exhaustively every activity involved in problem-solving. Different authors who have used the idea of EIPs have developed their own sets adapted to their problem-solving tasks (e.g. Huber 1980, Payne *et al.* 1993). Newell and Simon acknowledged that a common vocabulary of EIPs to describe any given task is hard, if not impossible, to achieve: 'What collection of symbols is primitive for a specific [Information Processing System] will vary with the particular application' (Newell and Simon 1972: 68).

For some researchers like Lee (personal communication), tasks that are not decomposable into components small enough to be easily handled in working memory are simply not learnable. This is a view that can be subsumed under the more general framework of Halford *et al.* (1998), where anything above quaternary relations are supposed to be out of the conventional limits of working memory. Again, problem-solving theories should specifically address this issue and should amplify the range of CPS tasks to be decomposed under their theoretical views.

4.2.6 Ill-defined vs well-defined.   The boundary between ill-defined and well-defined is vague. This feature is in the psychological descriptions and not in the formal descriptions section because it is clearly impossible to construct a formal definition of 'well-defined problem' (Simon 1973). The definition of ill-defined problems has been changing with different reformulations. For Newell (1969), ill-defined problems are those in which strong solutions do not exist and the solver has to use the so-called weak methods. For Simon (1973), the only way of defining an ill-defined problem is by contrast with a well-structured problem. A well-structured problem is one that has been formalized by the problem-solver. A pragmatic approach to this definition is that well-structured problems are those that can be handled by the general problem-solver (GPS). Although theorem proving and chess were always considered well-structured problems, Simon (1973) questions this assumption, to show how flexible the classification can be. Still, it seems safe to say that the sugar factory and Biology Lab are well-defined tasks, while Moro and Lohhausen are ill-defined tasks.

5.   Concluding remarks

Problem-solving has been traditionally a task-centred field. This is not surprising, since some authors (e.g. VanLehn 1989) think that 'task' and 'problem' are virtually synonymous. Still, there are no good task taxonomies and the definition of the area itself is blurry. These problems get accentuated when talking about complex problem-solving.

The European definition, proposed recently to restrict the scope of the field, left behind many problem-solving situations that could be considered complex (reading, algebra, etc.) because they did not fulfil the criteria of changing in real time, being opaque, etc. Curiously, many tasks that are not considered problem-solving by their own authors are included in this definition (e.g. KA-ATC, AMBR-ATC, Space Fortress).

The Kanfer and Ackerman ATC task fulfils all three requisites in the consensus definition (at least, one and two without any doubt, though the third is open to discussion), but does not fit some of the most popular definitions of problem-solving, for example Newell and Simon's (1972) that in its basic form states that a person is confronted with a problem when he wants something and does not know immediately which series of actions he can perform to get it. In the Kanfer and Ackerman ATC task, participants know perfectly well at any moment how to perform and what actions lead to a goal (e.g. land a 747 in a long runway); that is why Lee and Anderson do not mention problem-solving in their discussion, although they consider ATC as a complex skill. The ATC tasks are also characterized by very smooth power-law learning functions. In some other tasks and real life situations, knowledge reorganizations reflect progressive insights and learning curves do not follow well-defined power-laws.

Some problem-solving tasks that do not have any of the key features mentioned in the consensus definition are accepted in the CPS tradition without question, mainly for historical reasons (e.g. the sugar factory task), despite the concerns raised by some authors (e.g. Funke 1991). Thus, the new (1995) definition covers some activities that are not problem-solving (e.g. simplified ATC tasks) or complex (e.g. the sugar factory) and leaves out some activities that are (e.g. electronics troubleshooting, programming). Table 2 summarizes the problems with the current definition.

When can a complex system be decomposed into smaller sub-systems? That is exactly the question addressed by Halford *et al.* (1998). Their relational complexity measure could be the beginning of a psychological theory of complexity, which would be welcomed not only in problem-solving, but basically all areas of cognitive science. Yet, a problem arises: how does one know where the sub-systems are? Even the theory's own authors (Phillips 1997: 1) worry about this: 'The lack of formal analysis into relational information involved in cognitive tasks threatens to undermine [Relational complexity theory's] utility'. The predominant approach in psychology has been to do a logical analysis of the functions required to do a complex task and then assume that the cognitive system has specialized mechanisms to do them, as Hunt (1991) pointed out. This kind of analysis is *a priori* and presents several difficulties. It is, at the moment, out of the question to find a formal language

Table 2. Some example tasks and their classification according to the European definition of CPS and the classical definitions of problem-solving.

	According to the European definition of CPS, is this a complex task?	According to the general definitions of problem-solving, is this a problem-solving situation?
Sugar Factory (Broadbent *et al.* 1986)	No	Yes
ATC-simplified tasks (Deutsch and Benyo 2001, Lee and Anderson 2001)	Yes	No
Reading (e.g. Stanovich and Cunningham 1991)	No	Yes (according to Stanovich *et al.*)
Space Fortress (Mané and Donchin 1989)	Yes	No
Mechanics, electronics (e.g. Carlson *et al.* 1990)	No	Yes

to describe all problem-solving that is complete (that is that enables researchers to describe every possible problem-solving task). It is thought that *holistic* approaches to complexity have a better chance of success, in that they do not need to decompose tasks into sub-components and the authors have oriented their efforts in this direction (see Quesada *et al.* 2001, 2002).

Regarding the taxonomic efforts, the main point is that there is no simple way of classifying complex tasks and no single dimension can be considered necessary and sufficient to identify a task as complex. Formal descriptions give a set of features that can be used to describe CPS tasks from a strictly system-based point of view and this description is useful, although not complete for psychological theory development. On the other hand, the psychological descriptions proposed describe a framework that should be considered by any researcher who uses an old CPS task or designs new ones and should be used to develop hypotheses and experiments. Specific theories of CPS or general cognitive theories adapted to explain complex problems should be able to explain the position of each task in this taxonomy and changes in it with different experimental manipulations.

An alternative definition of complexity is not proposed here, because one does not think that the definition of problem-solving is completely worked out. All the phenomena studied as problem-solving fit just as well under other categories such as learning, reasoning and decision-making. However, one element that is not covered by those categories is planning. Perhaps one could restrict the definition of problem-solving to those tasks that involve planning (Burns, personal communication). The question of how to define complexity, problem-solving and their relations is still awaiting a good answer.

References

ANDERSON, J.R., 1990, *The Adaptive Character of Thought* (Hillsdale, NJ: Lawrence Erlbaum associates).

ANDERSON, J.R. and LEBIERE, C., 1998, *The Atomic Components of Thought* (Mahwah, NJ: Lawrence Erlbaum associates).

ARTMAN, H., 1999, Operation within and between hierarchical units in dynamic decision making. *Ergonomics*, **42**, 1404–1417.

BERRY, D.C. and BROADBENT, D.E., 1984, On the relationship between task performance and associated verbalizable knowledge. *Quarterly Journal of Experimental Psychology*, **36A**, 209–231.

BIRNEY, D.P. and HALFORD, G.S., 2002, Cognitive complexity of suppositional reasoning: an application of the relational complexity metric to the knight-knave task. *Thinking & Reasoning*, **8**, 109–134.

BREHMER, B., 1992, Dynamic decision-making—human control of complex-systems. *Acta Psychologica*, **81**, 211–241.

BREHMER, B., 1995, Feedback delays in complex dynamic decision tasks, In *Complex Problem Solving: The European Perspective*, J. Funke (Ed.), pp. 103–130 (Hillsdale, NJ: Lawrence Erlbaum).

BREHMER, B. and ALLARD, R., 1991, Dynamic decision making: the effects of task complexity and feedback delay, In *Distributed Decision Making: Cognitive Models for Cooperative Work*, J. Rasmussen and B. Brehmer (Eds.), pp. 319–334 (London: John Wiley & Sons).

BROADBENT, D.E., 1977, Levels, hierarchies and the locus of control. *Quarterly Journal of Experimental Psychology*, **32**, 109–118.

BROADBENT, D., FITZGERALD, P. and BROADBENT, M.H., 1986, Implicit and explicit knowledge in the control of complex systems. *British Journal of Psychology*, **77**, 33–50.

BRUCE, A.C. and JOHNSON, J.E.V., 1996, Decision-making under risk: effect of complexity on performance. *Psychological Reports*, **79**, 67–76.

BUCHNER, A., 1995, Basic topics and approaches to the study of complex problem solving, In *Complex Problem Solving: The European Perspective*, P.A. Frensch and J. Funke (Eds.), pp. 27–63 (Lawrence Erlbaum Associates: Hillsdale, NJ).

BUCHNER, A., FUNKE, J. and BERRY, D., 1995, Negative correlations between control performance and verbalizable knowledge: indicators for implicit learning in process control tasks? *Quarterly Journal of Experimental Psychology*, **48A**, 166–187.

BURNS, B.D. and VOLLMEYER, R., 2000, Problem solving: phenomena in search for a thesis. In *Proceedings of the Cognitive Science Society Meeting*, 13–15 August 2000, Pittsburgh, USA, pp. 627–632 (Lawrence Erlbaum Associates: NY).

BURNS, B.D. and VOLLMEYER, R., 2002, Goal specificity effects on hypothesis testing in problem solving. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, **55A**, 241–261.

BUSEMEYER, J.R., 2002a, Dynamic decision making. In *International Encyclopedia of the Social and Behavioral Sciences: Methodology, Mathematics and Computer Science*, N.J. Smelser and P.B. Bates (Eds), pp. 3903–3908 (Oxford: Elsevier).

BUSEMEYER, J.R., 2002b, Dynamic systems. In *Encyclopedia of Cognitive Science*, L. Nadel (Ed.), pp. 1039–1049 (London: Nature).

BUSEMEYER, J.R. and TOWNSEND, J.T., 1993, Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, **100**, pp. 432–459.

CAMBELL, D.T. and STANLEY, J.C., 1963, *Experimental and Quasi-Experimental Designs for Research* (Rand McNally & Company: Chicago).

CAÑAS, J.J., QUESADA, J.F., ANTOLÍ, A. and FAJARDO, I., 2004, Cognitive flexibility and adaptability to environmental changes in dynamic complex problem solving tasks. *Ergonomics*, **46**, 482–501.

CARLSON, R.A., KHOO, B.H., YAURE, R.G. and SCHNEIDER, W., 1990, Acquisition of a problem-solving skill: levels of organization and use of working memory. *Journal of Experimental Psychology: General*, **119**, 193–214.

CHASE, W.G. and SIMON, H.A., 1973, The mind's eye in chess. In *Visual Information Processing*, W.G. Chase (Ed.) (Academic Press: New York).

CHRISTOFFERSEN, K., HUNTER, C.N. and VICENTE, K.J., 1996, A longitudinal study of the effects of ecological interface design on skill acquisition. *Human Factors*, **38**, 523–541.

CHRISTOFFERSEN, K., HUNTER, C.N. and VICENTE, K.J., 1997, A longitudinal study of the effects of ecological interface Design on fault management performance. *International Journal of Cognitive Ergonomics*, **1**, 1–24.

CHRISTOFFERSEN, K., HUNTER, C.N. and VICENTE, K.J., 1998, A longitudinal study of the impact of ecological interface design on deep knowledge. *International Journal of Human-Computer Studies*, **48**, 523–541.

DELOSH, E.L., BUSEMEYER, J.R. and MCDANIEL, M.A., 1997, Extrapolation: the sine qua non for abstraction in function learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 968–986.

DEUTSH, S. and BENYO, B., 2001, The d-omar simulation environment for the ambr experiments. In *10th Conference on Computer Generated Forces and Behavioral Representation*, 15–17 May 2001, Norfolk Waterside Marriott, Norfolk, VA, SISO (Simulation Interoperability Standards Organization).

DIENES, Z. and FAHEY, R., 1995, Role of specific instances in controlling a dynamic system. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **21**, 848–862.

DIENES, Z. and FAHEY, R., 1998, The role of implicit memory in controlling a dynamic system. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, **51A**, 593–614.

DIETRICH, E. and MARKMAN, A.B., 2000, Cognitive dynamics: computation and representation regained. In *Cognitive Dynamics: Conceptual and Representational Change in Humans and Machines*, E. Dietrich and A. B. Markman (Eds.), pp. 5–29 (Mahwah: Lawrence Erlbaum).

DÖRNER, D., 1975, Wie menschen eine welt verbessern wollten und sie dabei zerstörten [How people wanted to improve the world]. *Bild der Wissenschaft*, Heft 2 (Populärwissenschaftlicher Aufsatz).

DÖRNER, D., 1989, *The Logic of Failure* (Henry Holt & Co.: New York).

DÖRNER, D. and SCHOLKOPF, J., 1991, Controlling complex systems; or expertise as 'grandmother's know-how', In *Toward a General Theory of Expertise*, edited by K.A. Ericsson and J. Smith, pp. 218–239 (Cambridge University Press: Cambridge).

DÖRNER, D. and WEARING, A.J., 1995, Complex problem solving: toward a (computersimulated) theory, In *Complex Problem Solving: The European Perspective*, edited by P.A. Frensch and J. Funke, pp. 65–99 (Lawrence Erlbaum: Hillsdale, NJ).

DÖRNER, D., KREUZIG, H.V., REITHER, F. and STAUDEL, T., (Eds) 1983, *Lohhausen: Vom Umgang Mit Komplexität* (Huber: Bern).

EDWARDS, W., 1962, Dynamic decision theory and probabilistic information processing. *Human Factors*, **4**, 59–73.

FRENSCH, P. and FUNKE, J., 1995a, *Complex Problem Solving: The European Perspective* (Lawrence Erlbaum: Hillsdale, NJ).

FRENSCH, P. and FUNKE, J., 1995b, Definitions, traditions and a general framework for understanding complex problem solving, *Complex Problem Solving: The European Perspective*, edited by P.A. Frensch and J. Funke, pp. 3–25 (Lawrence Erlbaum: Hillsdale, NJ).

FUNKE, J., 1991, Solving complex problems: exploration and control of complex systems, In *Complex Problem Solving: Principles and Mechanisms*, R.J. Sternberg and P.A. Frensch (Eds.), pp. 185–222 (Hillsdale: Lawrence Erlbaum).

FUNKE, J., 1992, Dealing with dynamic systems: research strategy, diagnostic approach and experimental results. *German Journal of Psychology*, **16**, 24–43.

FUNKE, J., 1993, Microworlds based on linear equation systems: a new approach to complex problem solving and experimental results, In *The Cognitive Psychology of Knowledge*, G. Strube and K.F. Wender (Eds.), pp. 313–330 (Amsterdam: Elsevier).

GAREY, M.R. and JOHNSON, D.S., 1979, *Computers and Intractability: A Guide to the Theory of Np-Completeness* (W.J. Freeman and Company: San Francisco).

GIBSON, F.P., FICHMAN, M. and PLAUT, D.C., 1997, Learning dynamic decision tasks: computational model and empirical evidence. *Organizational Behavior & Human Decision Processes*, **71**, 1–35.

GONZALEZ, C., LERCH, F.J. and LEBIERE, C., 2003, Instance-based learning in dynamic decision making. *Cognitive Science*, **27**, 591–635.

GRANLUND, R., JOHANSSON, B. and PERSSON, M., 2001, C3fire a micro-world for collaboration training in the rolf environment. In *Proceedings to (Sims 2001) the 42nd Conference on Simulation and Modelling, Simulation in Theory and Practice* (Scandinavian Simulation Society: Porsgrunn).

GRAY, W.D., 2002, Simulated task environments: the role of high fidelity simulators, scaled worlds, synthetic environments, and laboratory tasks in basic and applied cognitive research. *Cognitive Science Quarterly*, **2**, 205–227.

HALFORD, G.S., WILSON, W.H. and PHILLIPS, S., 1998, Processing capacity defined by relational complexity: implications for comparative, developmental, and cognitive psychology. *Behavioral & Brain Sciences*, **21**, 803–864.

HINZMAN, D.L., 1988, Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological review*, **85**, 445–463.

HOGARTH, R.M., 1975, Decision time as a function of task complexity, In *Utility, Probability, and Decision Making*, D. Wendt and C. Cvlek (Eds.), pp. 321–338 (Dordrecht, Holland: D. Reidel).

HOWIE, D.E. and VICENTE, K.J., 1998, Measures of operator performance in complex, dynamic microworlds: advancing the state of the art. *Ergonomics*, **41**, 85–150.

HUBER, O., 1980, The influence of some task variables on cognitive operations in an information-processing decision model. *Acta Psychologica*, **45**, 187–196.

HUMPHREYS, S.H., WILES, J. and DENNIS, S., 1994, Toward a theory of human memory: data structures and access processes. *Behavioral and Brain Sciences*, **17**, 655–692.

HUNT, E., 1991, Some comments on the study of complexity, In *Complex Problem Solving: Principles and Mechanisms*, R.J. Sternberg and P.A. Frensch (Eds.), pp. 383–395 (Hillsdale, NJ: Lawrence Erlbaum Associates, Inc).

JUDD, C.M. and KENNY, D.A., 1981, *Estimating the Effects of Social Interventions* (Cambridge University Press: Cambridge).

KAELBLING, L.P., LITTMAN, M.L. and MOORE, A.W., 1996, Reinforcement learning: a survey. *Journal of Artificial Intelligence Research*, **4**, 237–285.

KENNY, D.A., 1979, *Correlation and Causality* (Wiley-Interscience: NY).

KLEIN, G., 1997, The recognition-primed decision model: looking back, looking forward, In *Naturalistic Decision Making*, C. Zsambok and G. Klein (Eds.), pp. 285–292 (Lawrence Erlbaum Associates: Mahwah, NJ).

KLEIN, G.A., ORASANU, J., CALDERWOOD, R. and ZSAMBOK, C.E., (Eds) 1985, *Decision Making in Action: Models and Methods*, 1993 (Ablex Publishing Corporation: Norwood, NJ).

KLEINMUNTZ, D.N., 1985, Cognitive heuristics and feedback in a dynamic decision environment. *Management Science*, **31**, 680–702.

KLEITER, G.D., 1985, Dynamic decision behavior: comments on rapoport paper, *Utility, Probability, and Decision Making*. In edited by D. Wendt and C. Cvlek, pp. 371–375 (D. Reidel: Dordrecht, Holland).

KOTOVSKY, K. and SIMON, H.A., 1990, What makes some problems really hard: explorations in the problem space of difficulty. *Cognitive Psychology*, **22**, 143–183.

KOTOVSKY, K., HAYES, J.R. and SIMON, H.A., 1985, Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology*, **17**, 248–294.

LEBIERE, C., WALLACH, D. and TAATGEN, N., 1998, Implicit and explicit learning in ACT-R, In *Cognitive Modeling Ii*, edited by F.R.R. Young, pp. 183–193 (Nottingham University Press: Nottingham).

LEE, F.J. and ANDERSON, J.R., 1997, Learning to act: acquisition and optimization of procedural skill, In *Proceedings of the 3rd International Conference on Cognitive Modelling*, N. Taatgen and J. Aasman (Eds.) (Veenendaal, The Netherlands: Universal Press).

LEE, F.J. and ANDERSON, J.R., 2001, Does learning a complex task have to be complex? A study in learning decomposition. *Cognitive Psychology*, **42**, 267–316.

LEE, Y. and VAKOCH, D.A., 1996, Transfer and retention of implicit and implicit learning. *British Journal of Psychology*, **87**, 637–651.

LOVEDAY, W., 1995, *The Effect of Complexity on Planning in the Tower of Hanoi* (University of Queesland: Queensland).

MANÉ, A. and DONCHIN, E., 1989, The space fortress game. *Acta Psychologica*, **71**, 17–22.

NEWELL, A., 1969, Heuristic programming: ill-structured problems, In *Progress in Operations Research 3*, J. Aronofsky (Ed.), pp. 361–414 (Wiley: New York).

NEWELL, A., 1980, Reasoning, problem solving, and decision processes: the problem space as a fundamental category, In *Attention and Performance VII*, R. Nickerson (Ed.), pp. 693–718 (Cambridge, MA: Harvard).

NEWELL, A., 1990, *The Unified Theories of Cognition* (Cambridge, MA: Harvard University Press).

NEWELL, A. and SIMON, H.A., 1972, *Human Problem Solving* (Englewood Cliffs, NJ: Prentice-Hall, Inc.).

OMODEI, M.M. and WEARING, A.J., 1995, The fire chief microworld generating program: an illustration of computer-simulated microworlds as an experimental paradigm for studying complex decision-making behavior. *Behavior Research Methods, Instruments & Computers*, **27**, 303–316.

PAYNE, J.W., BETTMAN, J.R. and JOHNSON, E.J., 1993, *The Adaptive Decision Maker* (New York: Cambridge University Press).

PHILLIPS, S., 1997, Measuring relational complexity in oddity discrimination tasks. *Noetica*, **3**, 1–14.

PUTZ-OSTERLOH, W., 1993, Strategies for knowledge acquisition and transfer of knowledge in dynamic tasks, In *The Cognitive Psychology of Knowledge*, G. Strube and K.F. Wender (Eds.), pp. 331–350, Series Advances in psychology, Vol. 101.

QUESADA, J.F., KINTSCH, W. and GOMEZ, E., 2001, A computational theory of complex problem solving using the vector space model (part II): latent semantic analysis applied to

empirical results from adaptation experiments, In *Cognitive Research with Microworlds*, J.J. Cañas (Ed.), pp. 147–158 (London: Elsevier).

QUESADA, J.F., KINTSCH, W. and GOMEZ, E., 2002, A theory of complex problem solving using latent semantic analysis, In *24th Annual Conference of the Cognitive Science Society*, W.D. Gray and C.D. Schunn (Eds.), pp. 750–755 (Lawrence Erlbaum Associates: Fairfax, VA).

RAPOPORT, A., 1975, Research paradigms for studying dynamic decision making behavior, In *Utility, Probability, and Decision Making*, D. Wendt and C. Cvlek (Eds.), pp. 349–369 (Dordrecht, Holland: D. Reidel).

RASMUSSEN, J., 1985, The role of hierarchical knowledge representation in decision making and system management. *IEEE Transactions on Systems, Man, and Cybernetics*, **SMC-15**, 234–243.

REITMAN, W.R., 1964, Heuristic decision procedures, open constraints, and the structure of ill-defined problems, In *Human Judgments and Optimality*, M.W. Shelly and G.L. Bryan (Eds.), pp. 282–315 (New York: John Wiley and Sons).

SALAS, E.E. and KLEIN, G., (Eds.) 2001, *Linking Expertise and Naturalistic Decision Making* (Mahwah, NJ: Lawrence Erlbaum Associates).

SCHOPPEK, W., 2002, Examples, rules, and strategies in the control of dynamic systems. *Cognitive Science Quarterly*, **2**, 63–92.

SIMON, H., 1973, The structure of ill-structured problems. *Artificial Intelligence*, **4**, 181–201.

SLOVIC, P., FISCHHOFF, B. and LICHTENSTEIN, S., 1977, Behavioral decision theory. *Annual Review of Psychology*, **28**, 1–39.

STANOVICH, E. and CUNNINGHAM, A.E., 1991, Reading as constrained reasoning, In *Complex Problem Solving: Principles and Mechanisms*, J.R. Sternberg and P.A. Frensch (Eds.), pp. 3–61 (Hillsdale, NJ: Lawrence Erlbaum Associates).

STERMAN, J.D., 1989, Misperceptions of feedback in dynamic decision making. *Organizational Behavior & Human Decision Processes*, **43**, 301–335.

STERMAN, J.D., 1994, Learning in and about complex systems. System dynamics review *Systems Thinkers, Systems Thinking*, **10**, 291–330.

STERNBERG, J.R., 1995, Expertise in complex problem solving: a comparison of alternative conceptions, In *Complex Problem Solving: The European Perspective*, P.A. Frensch and J. Funke (Eds.), pp. 295–321 (Lawrence Erlbaum: Hillsdale, NJ).

STERNBERG, J.R. and FRENSCH, P., (Eds) 1991, *Complex Problem Solving* (Lawrence Erlbaum Associates: Hillsdale, NJ).

SWELLER, J., 1998, Can we measure working memory without contamination from knowledge held in long-term memory? *Behavioral & Brain Sciences*, **21**, 845–846.

TAATGEN, N., 1999, Learning without limits: from problem solving towards a unified theory of learning. PhD Thesis, University of Groningen.

TODA, M., 1968, Algebraic models in dynamic decision theory, In *Algebraic Models in Psychology*, C.A.J. Vlek (Ed.) (Leyden: Psychological Institute of the University of Leyden).

VANLEHN, K., 1989, Problem solving and cognitive skill, In *Foundations of Cognitive Science*, M.I. Posner (Ed.), pp. 527–579 (MIT Press: Cambridge, MA).

VEMURI, V., 1978, *Modelling of Complex Systems: An Introduction* (Academic Press: New York).

VOLLMEYER, R., BURNS, B.D. and HOLYOAK, K.J., 1996, The impact of goal specificity on strategy use and the acquisition of problem structure. *Cognitive Science*, **20**, 75–100.

WARWICK, W., MCILWAINE, S. and HUTTON, R., 2002, Developing computational models of recognition-primed decisions: progress and lessons learned. In *11th Conference on Computer Generated Forces and Behavioral Representation*, 7–9 May 2002, Norfolk, VA, SISO (Simulation Interoperability Standards Organization).

WARWICK, W., MCILWAINE, S., HUTTON, R. and MCDEMOTT, P., 2001, Developing computational models of recognition-primed decision making. In *10th Conference on Computer Generated Forces and Behavioral Representation*, 14–17 May 2001, Norfolk, VA, SISO (Simulation Interoperability Standards Organization).

WASON, P.C., 1968, Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, **20**, 273–281.

ZSAMBOK, C.E. and KLEIN, G., (Eds) 1997, *Naturalistic Decision Making* (Lawrence Erlbaum Associates: Hillsdale, NJ).

About the authors

Walter Kintsch is Professor of Psychology and Director of the Institute of Cognitive Science at the University of Colorado in Boulder. He came to the University of Colorado in 1968 after receiving a PhD in psychology from the University of Kansas and faculty appointments at the University of Missouri and the University of California at Riverside. His research focus has been on the study of how people understand language, using both experimental methods and computational modelling techniques. In co-operation with the Dutch linguist Teun van Dijk, he formulated the first psychological process theory of discourse comprehension in 1978. In 1988, this work was reformulated as a constraint- satisfaction process. His latest book 'Comprehension' appeared in 1998 and argues that many cognitive processes can be usefully conceptualized as comprehension processes. Kintsch is a member of the National Academy of Education and received the Distinguished Scientific Contribution Award of the American Psychological Association in 1992 and an honorary doctorate from the Humboldt University in Berlin in 2001.

Jose Quesada is a graduate student in the University of Granada and a research associate at the institute of Cognitive Science, University of Colorado, Boulder. His main interest is in computational theories of knowledge representation and problem-solving. His theoretical work is around Latent Problem Solving Analysis (LPSA), a theory that proposes that when facing complex environments, people represent the task using huge amounts of simultaneous small constraints, in a similar way as when people learn the meaning of words in a natural language. The theory has been tested in several simulated, laboratory environments such as firefighting and thermodynamic systems and also in real-life tasks such as piloting and landing a Boeing 747. In 2002, he was granted an European project to develop an automatic landing technique assessing system based on LPSA, using a high fidelity flying simulator. He administers the Latent Semantic analysis (LSA) web-site (lsa.colorado.edu).

Emilio Gomez is professor of Psychology at the Department of Experimental Psychology of Granada University. He teaches Psychology of consciousness and Motor Processes. His main interests are spatial attention and the shift of mental set. His research focus has been on attentional capture and on how to determine when mental effort is useful or disturbs a problem-solving situation. The main idea of his theoretical work is to study how attention works in complex environments. Latent Problem Semantic Analysis (LPSA) can help to build a measure of strategies switching, a kind of multi-dimensional cost shift of mental set. In 2002 he was granted, together with Jose Quesada, a European grant to develop an automatic landing technique assessing system based on LPSA. In 2003, he was granted a Spanish Government grant to Apply LPSA to the complex relationships between (vote) intention and action (real vote) in national elections, where cost of intention switching has been found.