Quesada, J. (in press) Spaces for problem solving. In T. Landauer, D. McNamara, S. Dennis & W. Kintsch (Eds). Latent Semantic Analysis: A road to meaning. Erlbaum.

Spaces for problem solving

Abstract

Latent Problem Solving Analysis (LPSA) is a generalization of LSA where the basic units of analysis are problem states or actions, and the contexts are performance samples (normally whole trials). The basic idea is that the learning mechanisms that were proposed for representing semantics may well be valid to capture the representations that people generate when performing other complex activities. For example, operating thermodynamic systems such as power plants, making decisions in management or medical situations, flying planes, etc. may be described as moving around in a multidimensional problem space. This space is created after large amounts of experience. The consequences and theoretical value of these basic ideas have been tested in several experiments which are reviewed in this chapter.

1.1 Introduction: Latent Problem Solving Analysis (LPSA)

What if the mechanisms that have been proposed for learning the semantics of natural languages are more general, and some other aspects of high level cognition could be explained appealing to them? Landauer and Dumais (1997) anticipated that their approach is not necessarily reduced to words, and that other complex activities could be explained using the same representational approach taken in LSA. Quesada, Kintsch and Gomez (submitted) presented Latent Problem Solving Analysis (LPSA) as a theory of representation in experienced problem solving that takes up the idea that complex activities are represented in a similar way as we represent semantics. The basic idea is as follows. Consider that the events of interest are states of a system instead of words, and the contexts they appear in are trials instead of documents. Then, co-ocurrence information between events and trials can be used to infer the relationship between them, and represent the constraints of the problem. What LPSA proposes is that problem spaces are constructed from experience with the environment alone. The same inference mechanism could be used to learn the semantics of different problem solving domains. In problem solving domains, there are a high number of weak relations between tokens (tokens representing states of the system or actions to control it). Any event in a system (either an action or a state) is a "word" in traditional LSA parlance, and each context (e.g., a trial in an experiment, or a landing in a flying simulator) is the equivalent to a document. Then, one can construct a matrix of events by trials as a starting point for the inference algorithm (in this case, singular value decomposition, SVD). This matrix is very sparse, as one can infer from the long tails of the distributions presented in Figure 1: most events do not occur but in a few contexts.

We shall illustrate these ideas with some examples of complex problem solving tasks. Consider the complex, dynamic task Firechief (Omodei & Wearing, 1993, 1995). In this problem, participants have to extinguish a forest fire by allocating trucks and helicopters with mouse commands. Every time a participant performs an action, it is saved in a log file as a row containing action number, command (e.g., drop water or move) or event[1] (e.g., a wind change or a new fire), current performance score, appliance number, appliance type, position, and landscape type. Most of these variables are not continuous, but on a nominal scale, such as 'type of movement'. For more information on the structure of the log files, see Omodei and Wearing (1995).

LPSA can be trained on a corpus of thousands of trials collected from participants playing Firechief in different experimental conditions. Each trial is a log file of about 300 actions. Actions are created by concatenating the different numeric and nominal attributes that define them. In the Firechief case, the corpus was composed of 360,199 actions in 3441 trials. Among them, only 57,565 were different actions, which means that on average each action appears 6.25 times in the corpus. Note that LPSA represents *only* the information that actual people interacting with the system experienced not all possible actions in this problem. The coocurrence matrix is then submitted to singular value decomposition (SVD)

[1] Events are generated by the system, while actions are generated by the user. Events are also lines in the log file. Only 1-2% of the lines in a log file are events.

and recreated using only a small fraction (e.g., 100 – 300) of the dimensions. The resulting space is a problem space for Firechief.

As another example, consider the Duress task (e.g., Christoffersen, Hunter, & Vicente, 1997). Duress (DUal Reservoir System Simulation) is a thermal-hydraulic scenario that was designed to be representative of industrial processes. In this system, all variables are continuous. Duress has two redundant feedwater streams that can be configured to supply water to either, both, or neither of two reservoirs. Participants control 8 valves and 2 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 liters per second and 7 liters per second, respectively). The system is determined by the laws of conservation of mass and energy, and its state can be described using 34 variables. Running a SVD on the Duress state-context matrix using several years of logs produces a problem space. Duress spaces also contain a few thousand events, but it this case they are system states, a subset of the 34 rounded continuous variables. Rounding the value of continuous variables makes sure that the number of types is not excessive and the average token repetition is close to natural languages. Applying the same logic of an SVD decomposition of a trial by state matrix produces a metric space that can serve as a good approximation for a human problem space for Duress.

In these two examples, we can see that the events used to train the model can be either actions of states (they convey the same information). Also we can see that the variables that define the system can be continuous or discrete: for LPSA, the relevant information is the coocurrence relationships between the token these variable values form and the contexts they appear in.

Note that the knowledge extracted from experience is not explicitly represented as rules. For example, when working with Duress, LPSA does not infer the laws of conservation of mass and energy nor represent these laws directly as rules, but as fuzzy relations between past states of the system. In section 3 we shall show how these constraints as represented in LPSA are sufficient to mimic human similarity judgments and, predict future states with accuracy. Of course, people also extract rules (e.g., Anderson & Lebiere, 1998) and build models of the relations between variables (e.g., Glymour, 2001) explicitly when interacting with their environments. What we propose is that both the associative (LPSA-like) and the rule-based systems work in parallel helping each other. In fact, many theorists in cognitive science propose that our mind is separated into two such systems (e.g., Evans & Over, 1996; Kahneman, 2003; Sloman, 2002; Sloman, 1996; Stanovich, 2004).

Stanovich (1991) reviews a collection of theories that propose two different systems of reasoning. He grouped them as system1 and system2. System1 is characterized as automatic, unconscious, and relatively undemanding of computational capacity. System2 is controlled, reflected in IQ tests, rule-based, and involved in analytical cognition. The tasks that pertain to system1 are highly contextualized, whereas system2 tends to decontextualize (abstract) problems.

The two systems can produce contrary solutions to the same situation. Stanovich (1991) is concerned with the evolutionary interpretation of the two systems and their relevance to arguments about rational behavior.

Sloman (1996) tagged the two components "associative system" and "rule-based system." He used similarity-based thought and temporal similarity relations to draw inferences and make predictions that resemble those of a sophisticated statistician: "Rather than trying to reason on the basis of an underlying causal or mechanical structure, it constructs estimates based on underlying statistic structure. Lacking highly predictive causal models, this is the preferred mode for many forecasters, such as weather and economic ones" (Sloman, 1996, p. 4). In pilots, for example, the underlying causal or mechanical structure can be present, but, after extensive experience, it can be easier for them to operate using the statistical structure that they have extracted during practice. Thus, the organism faces important induction problems that can be solved in different ways. For some situations the "rule-based" system provides advantages, whereas in others it is the automatic, associative system the one that dominates induction.

The key step for induction in LPSA is the reduction of the dimensionality of the space. Imagine a hypothetical problem-solving task that, when performed from the beginning to the end in one of the N possible ways, traverses 300 states. To make it a really simple example, let us assume that every state is described using 6 dichotomous variables ($2^6 = 64$ possible states). Since we have 300 states in our sample of performance, there are $64^{300} = 7_E541$ possible paths in this task. Every sample would be represented as a matrix of $6 \times 300 = 1800$ values. After the dimension reduction used in LPSA, every sample would be represented as a vector of only 100 - 300 values.

The areas where LPSA has been applied differ from the traditional problem solving domains such as puzzles that have been used widely in psychology. Puzzles are aimed at system2, whereas system1 is best studied in the context of realistic, semantically rich tasks.

We propose that both language and complex solving use the same system of representation (multidimensional spaces) and inference mechanism (dimension reduction). Problem solving domains show similar distributional properties to each other, and to language. Figure 1 shows frequency distributions for three complex problem solving tasks, and one language corpus, TASA. When the log of the frequency of each type is plotted against the log of its rank, the result is close to a straight line. This implies a power law relation between rank of the item's frequency and actual frequency in the corpus (Zipf, 1949). Even though both the problem solving corpora and the TASA language corpus show departures from a perfect power law, the distributions show that there are very few high-frequency types, and very long tail really low-frequency types.

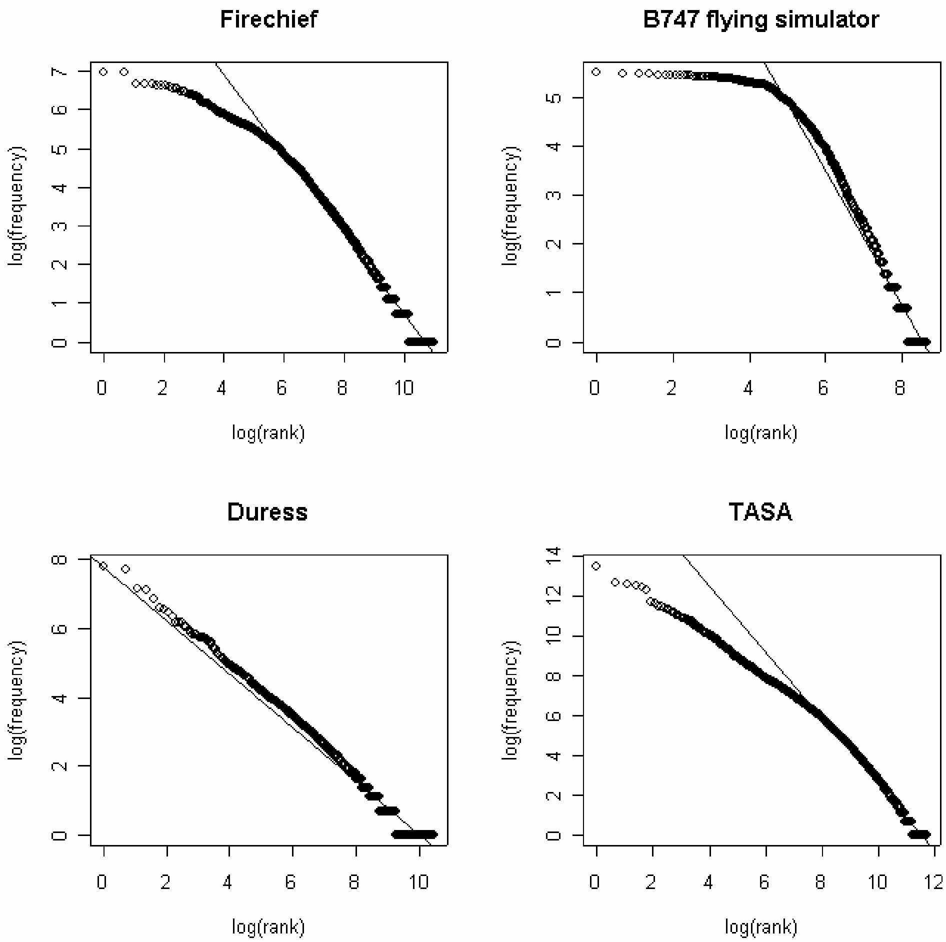

Figure 1: Different frequency-rank curves for three complex problem solving tasks, and one corpus of text (TASA). Even though the curves do not perfectly follow Zipf's law, they show that in the three problem solving domains, as in the language corpus, there are very few popular events, and a large majority of low frequency events.

Of course, there are many ways in which these frequency curves could exist as a consequence of different systems of representation. However, considering language and complex problem solving as two examples of a common representational system has advantages. First, it integrates language semantics with some types of problem solving in which one has to learn a large vocabulary of action-states. As we will see in this chapter, LPSA is an existence proof that a formalism that is useful describing a varied sample of complex tasks is possible. Second, it has an easy evolutionary explanation that our representational system could be adapted to this particular situation. If events in the environment (no matter what they are: words, faces, cities, etc) are distributed in a similar way, with many non-frequent events, and a selected few that are highly frequent, then it is rational to represent them using a common system.

Of course, we do not try to suggest that metric spaces learned from raw coocurrence data can explain representation in all types of problem solving and decision behavior. We view LPSA as an associative, intuitive system that works in collaboration with a conscious, rule-based system. People use both systems in different degrees depending on the problems they face. We focus here on the knowledge that can be obtained using coocurrence type-context only. However, there are other possibilities in a corpus-based approach to cognitive representation. Frequency (occurrence) and joint frequency (coocurrence) are two of the simplest statistics that can be derived from a corpus. Future proposals may use more fine-grained assumptions to infer structured representations from distributional information.

The other chapters in this book are examples of how a corpus-based approach to cognition is useful in the areas of semantics and language comprehension. In the next sections we argue that this corpus-based approach to cognitive representation holds promise for the areas of problem solving and decision making.

First, we review some issues in the literature, and how LPSA addresses them. Second, we review the experiments and simulations that present evidence for LPSA assumptions in the problem solving domain. Third, the discussion and conclusions are presented.

1.2 Some theoretical Issues in problem solving and LPSA solutions

We have presented the basics of the LPSA theory and its relations to LSA. This theory was developed as a response to four theoretical issues in the field of problem solving. In this section we present these issues, why we believe they are important and unaddressed by current theories, and how LPSA tries to address them.

(1) *The recursive definition of problem spaces*. From the very early attempts of the general problem solver (GPS, Newell & Simon, 1963) to the more contemporary approaches to expert and novice comparisons, the problem soling tasks selected have had a remarkable importance in shaping theory. However, as the field adopted new tasks, new theories needed to be proposed. Particularly distressing was the fact that each task generated an independent theory, often poorly specified by a flow diagram, and thus different results could barely be

integrated: "Theorists seem simply to write a different theory for each task" (Newell, 1980, p. 694). Newell proposed that the integrative metatheory could be built around the problem space hypothesis: "the fundamental organizational unit of all human goal-oriented symbolic activity is the *problem space*" (Newell, 1980, p.696).

The concept of problem space has been central to theory construction in the field of problem solving.[2] This concept has been used widely, and different researchers have interpreted it in different ways. It is natural that a useful concept experiences many incarnations in psychology. However, the definition and origin of problems spaces are often vague and mysterious. One contribution of the distributional approach to problem solving advocated here is that problem spaces are derived empirically from experience. In most theories of representation in problem solving, the problem spaces are hypothesized by experimenters, not derived automatically by an unsupervised procedure. Newell (1980) pointed out that one of the major items on the agenda of cognitive

---

[2] However, the problem space is a surprisingly ill-defined concept that has been changed and reworked in successive papers by their own proponents (Newell, 1980, 1990)(Newell, 1980, 1990), and by others (e.g., Laird, Rosenbloom, & Newell, 1986; see Quesada, unpublished, appendix J for a review of different attempts of definition)(e.g., Laird, Rosenbloom, & Newell, 1986; see Quesada, unpublished, appendix J for a review of different attempts of definition). Since most researchers in problem solving use it, it has been stretched and adapted in different ways to cover new situations, and some authors (e.g., Burns & Vollmeyer, 2000)(e.g., Burns & Vollmeyer, 2000) have issued warnings about this.

psychology is to banish the homunculus (little man residing "'elsewhere" that is responsible of all the marvelous things that have to be done to create the system's behavior). The homunculus is present in the classical problem-solving theory at least in one place: the intelligent creation of the problem space. This argument is circular: the generation of problem spaces is a symbolic cognitive task, and therefore it must be performed by means of a problem space. LPSA provides a solution to break this recursive explanation: the problem spaces are derived automatically from experience, there is no need for an intelligent agent in their creation.

(2) A related issue is *the increasing importance given to content (semantics) in problem solving*. In the 60s and 70s the problem solving studies were oriented to the detection and testing of general heuristics (e.g., mean-end analysis) by which a person could reduce search in a problem space (Newell & Simon, 1972). The possibility of an algorithm that could solve a very general class of problems was contemplated in programs such as the GPS (Newell & Simon, 1963). The idea that several problems could be abstracted away, removing any content to leave a common formal skeleton, was appealing. As research progressed, the ideal of a GPS drifted away. Two formally equivalent versions of the same problem with different cover stories could make the solution times vary enormously (Kotovsky & Fallside, 1989; Kotovsky, Hayes, & Simon, 1985; Kotovsky & Simon, 1990). If the representations and methods were the same, why would the two problems be so different in difficulty?

At about the same time, evidence on why content matters in problem solving came from the expertise literature. Experts could have extraordinary abilities that were restricted to a particular domain, being unable to extrapolate them to domains that were seemingly related (e.g., Adelson, 1984; Ericsson, Krampe, & Tesch-Römer, 1993; Ericsson & Staszewski, 1989). It seemed that what experts were capitalizing on was not better general-purpose search heuristics so much as by domain-specific knowledge (e.g., Chase & Simon, 1973). That is, the advantage of experts is tied to the semantics of the domain. LPSA offers a way of capturing the semantics of the domain in a formal way. It also explains how different domains can be represented using the same formalism. In LPSA, every event carries meaning since its representation is computed in relation to past experiences with other events.

(3) *Tasks are not representative of real-life situations*. Most experimenters during the 70's and 80's used simple tasks, such as the Tower of Hanoi, missionaries and cannibals, etc, to study problem solving in laboratory situations. However, many real life situations have certain characteristics that are not captured by these tasks. Many real tasks are (1) *dynamic*, because early actions determine the environment in which subsequent decision 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, and these decisions are, in return,

completely dependent on one another. For a task that is changing continuously, the same action can be definitive at moment *t1* and useless at moment *t2*. However, 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. Several researchers have started working on a set of computer-based, experimental tasks that are dynamic, time-dependent, and complex, called *microworlds,* and the area of thinking and reasoning that deals with them has been called Complex Problem Solving (CPS). We have discussed two examples of *microworlds*, Firechief and Duress. To test LPSA, we have chosen tasks that are representative of real world tasks, and have the characteristics mentioned above. Thus, in addition to the laboratory tasks we have used we have used a real-world task, a high fidelity landing simulator.

(4) *No common representational assumptions for different tasks: One task, one theory of representation*. This problem was also raised by Newell (1980): Certain modelers tend to analyze tasks in a way that is mostly dictated by their intuition of how people proceed when interacting with the environment. For example, Lee and Anderson (2001) used a simplified Air Traffic Controller task (ATC) in a controlled lab situation; they used three levels of description: (a) the unit-task level, (b) the functional level, and (c) the keystroke level. However, Lee and Anderson (2001) used no learning or representation theory to generate such decompositions: "[Our] task decomposition does not depend on any particular theory of skill acquisition. Rather, it solely relies on a pretheoretical task analysis" (p. 273). Although this decomposition can be well motivated and valid for the

task at hand, there is no attempt to abstract a task representation that can be used in more than that particular task. In psychology, the investigation of how people represent and interact with the problem solving tasks is called task analysis. Is task analysis the way to find common representational assumptions for different tasks? Schraagen, Chipman and Shute (2000) review 20 papers, and each of those presents on average several techniques for task analysis. These techniques include verbal protocol analysis, interviews, elaborated knowledge elicitation methods, similarity judgments, analytical descriptions of the task such as manuals, etc. It seems evident that there are no integrated theories about the task, nor are there consistent criteria for accepting or rejecting a particular task analysis. The distributional approach that LPSA exemplifies appears to propose a common measure for CPS tasks that can be applied to a wide variety of tasks, as we have already seen in the introduction.

1.3 Testing LPSA in problem solving

LSA was tested mainly by proving that the model could do some higher-cognition task (e.g., solving a synonym test Landauer & Dumais, 1997) as well as humans. In our testing of LPSA we have followed the same approach. However, we have also tested some basic assumptions of the metric spaces that LPSA proposes as problem spaces.

1.1.1.1 Human similarity judgments

People can evaluate the similarity between problem solving solutions. This is a skill that most people practice overtly in their daily life. Sometimes we have to compare other people's solutions to problems. For example, a teacher grading two student essays of the same test must compare them to each other, and maybe to an ideal essay that she would produce. Chess players study and compare the solutions (moves) of masters in difficult games as part of their training. Personnel selection departments compare the solutions of employees solving their daily projects for promotion.

If the model captures people's intuitions on similarity, then LPSA cosines for pairs of trials should correlate with human judgments of similarity. One of the simplest experiments that can be done to test LPSA is to select a few pairs of trials. Quesada, Kintsch and Gomez (submitted) did just that using Firechief. The videos were selected from a corpus of 1000 trials collected in past experiments with Firechief. LPSA was trained in the same corpus. All the trials had identical landscape, initial conditions, and events (fire appearances, wind changes). That is, the variance in every two videos was due only to participant interventions and their interaction with the system rules.

Quesada, Kintsch and Gomez (Quesada et al., submitted) presented experienced participants with 8 pairs of videos and were asked to assign a number from 0 to 100 for each pair reflecting the similarity of the two items in the pair. Participants could play the videos as many times as they wanted. Their similarity judgments correlated .94 with the LPSA cosines for these trials.

Testing for asymmetry

In LPSA, problem spaces are metric spaces. That is, similarity is represented as a metric distance, and it is symmetrical. However, Tversky (1977) demonstrated that in certain circumstances, human similarities are asymmetrical (e.g., the similarity of China to Korea is not the same as Korea to China). If problem solving spaces are metric spaces, then there must be no asymmetries in the similarity judgments for trials. Quesada, Gonzalez and Vakunov (in preparation) designed experiments where participants watched a screen with two replay videos played simultaneously, A and B. The A video was the reference video: participants had to focus on it, consider its characteristics, and made similarity judgments of the other video B with respect to the reference one. The instructions emphasized the importance of the reference video, asking them to use the A video as their comparison point, and make every judgment relative to it.

20 pairs sampled from the 380 possible permutations of 20 elements taken two by two, and other 20 reversing original pairs: that is the reference video was the other video. Participants were told that people vary in the way they solve the task and that we wanted them to say how similar the two videos are numerically (1-100). The same design was tested with two different complex problem solving tasks: Firechief and the 'water purification plant' (WPP) task (see e.g., Gonzalez, Lerch, & Lebiere, 2003).

Quesada et al. (in preparation) were not able to find any reliable asymmetries in any of these two tasks. At least at the trial level, similarity judgments of replay videos for this kind of problem solving situations are symmetrical. The failure to find the asymmetry could be due to the fact that Quesada et al. used very different materials from what Tversky (1977) used. However, Aguilar and Medin (1999) failed to replicate Tversky (1977) using the same design and type of materials in several experiments, which points in the direction that asymmetries in human similarity comparisons are small and may be hard to obtain. In any case, these results agree with the idea that people may represent problem solving situations within the metric constraints imposed by LPSA.

Context effects

Another assumption of metric models that Tversky (1977) challenged was that of context independence. Tversky found that adding additional elements to a comparison of two may change the perceived similarity between the two elements. Concretely, the extra elements would increase the salience of some features and tip the balance of importance that was obtained when the two items were alone, with no context. Tversky and Gati (1978, study 3) showed relevant context effects as well with countries as stimuli. Medin, Goldstone, and Gentner (1993, experiment 1) found that context can determine what features of an item are considered. In their experiment, people compared triplets of artificial, simple stimuli. The advantage of these stimuli is that they differ in only a few features. The open question is: can one find the same context effects using highly complex problem solving videos? Quesada et al. (in preparation) presented people with sets of four videos that could be played simultaneously. Participants had to rearrange four balls in a square; each ball represented a video. So the closer the balls were, the more similar they considered the problem solutions in the videos

(Goldstone, 1994). Quesada et al. (in preparation) used the average similarity judgments in the symmetry experiment (average of A->B and B->A) as distances to select the stimuli for the context experiment.

Using these distances, Quesada et al. (in preparation) calculated the closest neighbor of each of the 20 videos. Then, they created a list of random quintuplets [A B C P Q]. C was the target, A and B were two items with random similarities to C. P was the closest neighbor of A and Q was the closest neighbor of B. This design mimics the ones used before with countries (e.g., Tversky, 1977; Tversky & Gati, 1978, study 4) and schematic faces (Tversky, 1977). However, Quesada et al.'s (in preparation) study differs from the studies that used artificial stimuli with separable dimensions. In problem solving tasks such as Firechief and Duress it is very difficult for people to find natural separable dimensions Most changes are continuous and multifaceted in these tasks. Quesada et al. (in preparation) found no systematic context effects in judging similarities of problem solving episodes. Although this is only indirect evidence, it seems that the LPSA assumption of context independence hold at least while watching videos in the two complex tasks explored.

1.1.1.2 Expertise effects

LPSA can be used to model expertise effects. LPSA is a fully-specified computational model that combines the advantages of two major expertise theories: long-term working memory (LTWM, Ericsson & Kintsch, 1995) and the constraint-attunement hypothesis (CAH, Vicente & Wang, 1998). A description of these two theories and how they relate to LPSA follows.

Current theories disagree on what are the most relevant factors that contribute to the expertise. The LTWM theory claims that working memory has two different components: a short-term working memory (STWM), which is available under any condition, but of very limited capacity, and a long-term memory (LTWM), that is available only in the domain where one is an expert, but provides unlimited capacity. Thus, the LTWM theory breaks the dichotomy between STWM and long-term memory (LTM). STWM accounts for working memory in unfamiliar activities but does not appear to provide sufficient storage capacity for working memory in skilled complex activities. LTWM is acquired in particular domains to meet specific demands imposed by a given activity on storage and retrieval. LTWM is task specific. Intense practice in a domain creates retrieval structures: associations between the current context and some parts of LTM that can be retrieved almost immediately without effort. The contents of working memory act as the center of a focus that activates other contexts from LTM that are related to them thanks to the retrieval structures. A retrieval structure is defined as an abstract, hierarchical knowledge structure used to organize cues used in the encoding and retrieval of information. LTWM theory proposes that LTWM is generated dynamically by the cues that are present in short-term memory. During text comprehension, for example, where the average human adult is an expert, retrieval structures retrieve propositions from LTM and merge them with the ones derived from text.

The retrieval structures have to be proposed *ex professo* for each domain for which the theory is to be applied. For example, the retrieval structure that chess masters use to encode and retrieve chess information was proposed to be a chessboard (Ericsson & Kintsch, 1995, p. 237)[3]. On the other hand, the retrieval structure that a waiter uses for memorizing orders is assumed to be a spatial description of the table[4]. This situation is less than optimal, as when one define a retrieval structure for each task, one is falling victim of the problem "one task, one theory" that we described before (Newell, 1980). In LTWM (at least the first instantiation of 1995) the retrieval structures are specific for each task and different from each other; in some cases, the retrieval structures have a hierarchical component, whereas in some other cases, there is a strong spatial component; this lack of definition has been criticized as "vague" by Gobet (1998), among others.

---

[3] There are other process theories of expertise that also propose the concept of retrieval structure, such as the Elementary Perceiver and Memorizer model (EPAM, e.g., Richman, Staszewski, & Simon, 1995), but we have omitted them from the discussion for brevity. In EPAM the concept of retrieval structure has changed with the theory. In chess situations, it was also supposed to be a chess board, although reimplementations such as CHREST (Gobet, 1993) proposed that the retrieval structure had two components: what Gobet called "Hypothesis," the longest set of which is the pattern containing the largest quantity of information up to that point, and an "internal representation", which was a schematic representation of the chess board. For expert digit memorizing, the retrieval structure proposed was treelike, with chunks of size three, four, and five (running times) and an algorithm that assembled the structure recursively.

[4] The retrieval of the similar contexts in text comprehension was not formalized until Kintsch (1998) proposed how to use the combination of LSA and CI for that. Formalization of the retrieval structures in LTWM for other tasks have not been proposed by the authors (although see Gobet, 2000a).

LTWM, like most theories of expertise, is a process theory. That is, it specifies the psychological mechanisms that explain the problem solving: it is a theory of ”how.”) An alternative view is the constraint Attunement Hypothesis (CAH , Vicente & Wang, 1998) which is a product (i.e., input-output) theory of expertise, where the question to answer is ”what” conditions are needed to observe expertise effects, rather than “how”. The CAH theory proposes an important distinction between *intrinsic* and *contrived* tasks. *Intrinsic* tasks are definitive features of the domain of expertise, for example, blindfolded chess, memorizing dinner orders, and memorizing digits. A *contrived* task is one that is not part of the domain of expertise, but designed to fulfill some experimental purposes. For example, chess players just play chess, and remembering random chess configurations is not part of the task. Vicente and Wang consider that most of the tasks used in the memory expertise literature are *contrived*, not *intrinsic*, and hence are not explained by LTWM and other process theories. In CAH environmental constraints are represented by the Abstraction Hierarchy. This is a hierarchy of mean-ends relations that each experimenter must construct by hand for each task. The main proposal of the CAH is that the amount of structure in the environment determines how large the expert’s advantage can be, and that these constraints can be represented by the abstraction hierarchy.

Quesada (unpublished, chapter 6) used the data from a six-month long, six-participants experiment reported in Christoffersen, Hunter, and Vicente (1996;

1997; 1998) to generate an 'LPSA-simulated expert' with three years of experience with the DURESS problem.

To test the 'LPSA-simulated expert' created after averaging the experience of the six participants, the reference task selected was prediction. Prediction plays a very important role in humans' interaction with the environment. Some scientists argue that many features of the cognitive system (such as representation, memory, and categorization) can be conceptualized as tools that help to predict the next states of an organism's environment (e.g., Anderson, 1990).

The methodology that Quesada (unpublished, chapter 6) used was to test how well a prediction could be generated using the nearest neighbors of a target slice of behavior. For example, in a trial of Duress, how much of the end can be predicted using the information from the beginning of a control episode? Or, when we are reading a passage of text, how much of the information contained at the end can an experienced reader anticipate? To implement this idea, one possibility is to define a cutting point that divides the predicting and predicted parts of the passage, and manipulate their sizes while evaluating the quality of the prediction. For brevity, let us assume that the cutting point we define is the point that leaves ¾ of the trial behind.

The model could predict the last quarter of any trial using the first three quarters with an average accuracy of .8. When the system was given an experience of only six months the predictions fell down to less than .3. If the system is trained with three years of practice in an environment with no constraints (that is, not governed by rules of conservation of mass and energy), the predictions were (as expected) very bad, and comparable to the novice level. LPSA's explanation for these results unifies CAH and LTWM. LPSA can explain both amount of practice effects (main tenant of LTWM) and amount of environmental structure effects (main tenant of CAH). At the same time, LPSA can be considered a computational implementation of the retrieval structures that LTWM proposes.

What is a retrieval structure in our theory? In LPSA, the retrieval structure is implemented as the closest area in the problem space to the situation at hand. If the current context is represented as a point in this space, the retrieval structure is an area of close points that are retrieved from memory when the situation is analyzed. Note that these retrieval structures are created empirically from the statistics of the environment in LPSA, and then they are not a-priori. The procedure to define the retrieval structures in text comprehension situations is the same as the one for the very different situation of problem solving in the thermodynamic task, which is certainly an advantage of the theory since the retrieval structures do not change with the task to be modeled.

Another line of evidence for LPSA as a theory of expertise comes from *landing technique assessment*. Quesada, Kintsch and Gomez (submitted) used a high-fidelity flying simulator to collect 400 landings where the landing conditions were manipulated systematically, and created a vector space with them. Two instructors evaluated the landing, one of them sitting in the copilot seat, and the other one watching plots of relevant variables in real time (complete- and reduced-information raters respectively).

The reason for this different exposure to information was to use the knowledge of the raters to filter the variables of interest from the more than 10,000 available in a landing flying simulator! The reduced-information expert plotted only 6, and even with this incredibly small set he could rate landing technique just fine. Thus, the LPSA model was trained with the variables that the reduced-information rater used in his plots to evaluate the landing technique. Then, the nearest neighbors of any new landing were used to generate automatic ratings. Each new landing acted as a probe to retrieve past landings that were similar to it. The grade for a new landing was simply the weighted average of the grades retrieved for its neighbors.

Quesada et al. (submitted) created several corpora modifying the number of dimensions (100, 150, 200, 250, 300, 350, and maximum dimensionality, 400) and the number of nearest neighbors used to estimate the landing ratings (from 1 to 10). Another manipulation was the inclusion or exclusion of a time tag, and the type of weighting scheme used (log entropy vs. none). This way, the possible combinations of levels were (7 x 10 x 2 x 2) = 240. For each of these combinations of levels, Quesada et al. (submitted) used leave-one-out crossvalidation to calculate the ratings for the landing excluded. The estimated ratings for each of the 400 landings were then correlated with the real ratings. The combination of levels that best correlated with both humans was selected, and that was: Corpus with 200 dimensions, 5 Nearest Neighbors, no weighting, no time tagging).

The average agreement between human raters was not very high (correlation of only .48). However, this correlation is in line with others reported for e.g., Clinical Psychologists (.41), Stockbrokers (<.40), Grain Inspectors (.62) and Pathologists (.50) (Shanteau, 2001, p. 237, table 13.3).

The ratings that the model generated agreed with both humans .46 and .39 respectively, about as much as the two human graders agreed with each other (.48).

One of the LPSA assumptions is that experienced humans perform dimension reduction to represent their environments. The equivalent model (5 nearest neighbors, no weighting, no time stamping) without performing dimension reduction (that is, using 400 dimensions, which is the shortest dimension of the matrix) correlates with humans (on average for all criteria) only .26, which can be interpreted as evidence for dimension reduction in the representation.

Conclusions

Problem solving, mental models, and reasoning are explanatory concepts employed in cognitive science to account for performance in complex tasks. LPSA shows that simple ideas such as similarity-based processing and pattern matching could have a role even in cognitively complex tasks.

LPSA assumes that representation takes place in a space that has fewer dimensions than the external (distal stimuli) space represented. Thus, the representation is simpler than the represented world. Chater and collaborators (e.g., Chater, 1999; Chater & Vitanyi, 2003) present a view of a cognitive system that compresses data, simplifying the representations. This compression may be lossy, (i.e., the system partly throws away information in the process) but the resulting representation may be more adaptive (i.e., it may predict future states of nature better, or produce better inferences, because accidental, inessential information is not present).

LPSA also assumes that humans retrieve the most similar past experiences to the current one automatically. The response to the current situation (for example, in grading a landing) occurs partially because the ratings of past landing which are similar to this one "come to mind", and the response is a composite of those ratings.

This automatic retrieval of past situations takes place at the same time that people use more effortful processes to solve the task. We propose that both the associative (LPSA) and the rule-based systems work in parallel helping each other, and that people use one or the other more depending on the situation at hand. If complex, dynamic tasks like the ones described here, we propose that the associative system is used often.

LPSA provides the means for modeling the constraints in different domains in a comparable manner. What we propose is to use large, naturalistic corpora of problem solving activity to generate problem space representations. The particular representation LPSA generates is a metric space via dimension reduction. The great advantage of LPSA is that it can deal with truly complex problem solving tasks, and with large corpora that provide realistic estimates of human problem solving behavior.

To conclude this chapter, we would like to use Simon's (1981) parable of the ant and the beach. Simon's (1981) wondered why the path that an ant describes when walking on a beach is so complex, and how this complexity can be ascribed to the ant. However, he continued, it could be the case that the complexity is in the beach itself. In other words, the cognitive system's complexity can be a reflection of the structure of the environment (Anderson, 1990; Anderson, 1991). A formal description of the constraints of this environment can be very well the best approximation to the cognitive representation that the mind uses. We view LPSA as a computational description of the beach obtained by analyzing the paths of thousands of ants.

references

Adelson, B. (1984). When Novices Surpass Experts - the Difficulty of a Task May Increase with Expertise. *Journal of Experimental Psychology-Learning Memory and Cognition, 10*(3), 483-495.
Aguilar, C. M., & Medin, D. L. (1999). Asymmetries of comparison. *Psychonomic Bulletin & Review, 6*(2), 328-337.
Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
Anderson, J. R. (1991). Is human cognition adaptive? *Behavioral & Brain Sciences, 14*(3), 471-517.
Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, New Jersey: Lawrence Erlbaum associates.
Burns, B. D., & Vollmeyer, R. (2000). Problem Solving: Phenomena in search for a thesis. In L. R. Gleitman & A. K. Joshi (Eds.), *Proceedings of the cognitive science society meeting* (pp. 627-632). NY: Lawrence Erlbaum Associates.
Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing*. New York: Academic Press.
Chater, N. (1999). The search for simplicity: A fundamental cognitive principle? *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology, 52*(2), 273-302.
Chater, N., & Vitanyi, P. (2003). Simplicity: a unifying principle in cognitive science? *Trends in Cognitive Sciences, 7*(1), 19-22.
Christoffersen, K., Hunter, C. N., & 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., & 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., & Vicente, K. J. (1998). A longitudinal study of the impact of ecological interface design on deep knowledge. *International Journal of human-Computer Studies, 48*(6), 729-762.
Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*(2), 211-245.
Ericsson, K. A., Krampe, R. T., & Tesch-Römer. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*(3), 363-406.

Ericsson, K. A., & Staszewski, J. J. (1989). Skilled memory and expertise: Mechanisms of exceptional performance. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 235-267). Hillsdale, NJ: Lawrence Erlbaum.

Evans, J. S. B. T., & Over, D. E. (1996). *Rationality and reasoning*. Hove, England: Psychology Press.

Glymour, C. (2001). *The Mind's Arrows: Bayes Nets and Graphical Causal Models in Psychology*. Boston: MIT Press.

Gobet, F. (1998). Expert memory: a comparison of four theories. *Cognition, 66*(2), 115-152.

Goldstone, R. (1994). *An efficient method for obtaining similarity data*.

Gonzalez, C., Lerch, F. J., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science, 27*, 591-635.

Kahneman, D. (2003). A perspective on judgment and choice - Mapping bounded rationality. *American Psychologist, 58*(9), 697-720.

Kotovsky, K., & Fallside, D. (1989). Representation and transfer in problem solving. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon*. Hillsdale, NJ: Erlbaum.

Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology, 17*(2), 248-294.

Kotovsky, K., & Simon, H. A. (1990). What makes some problems really hard: Explorations in the problem space of difficulty. *Cognitive Psychology, 22*(2), 143-183.

Laird, J., Rosenbloom, P., & Newell, A. (1986). *Universal subgoaling and chunking*. Boston: Kluwer.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review, 104*, 211-240.

Lee, F. J., & Anderson, J. R. (2001). Does learning a complex task have to be complex? A study in learning decomposition. *Cognitive Psychology, 42*, 267-316.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for Similarity. *Psychological Review, 100*(2), 254-278.

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

Newell, A. (1990). *The unified theories of cognition*: Harvard University Press.
Newell, A., & Simon, H. A. (1963). GPS: A program that simulates human thought. In E. A. Feigenbaum & J. Feldman (Eds.), *computers and thought* (pp. 279-293). Oldernbourg KG.
Newell, A., & Simon, H. A. (1972). *Human Problem Solving*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc.
Omodei, M. M., & Wearing, A. J. (1993). Fire Chief (Version 2.3): University of Melbourne.
Omodei, M. M., & 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.
Quesada, J. F. (unpublished). *Latent Problem Solving Analysis (LPSA): A computational theory of representation in complex, dynamic problem solving tasks.* Unpublished Phd. dissertation, Granada (Spain).
Quesada, J. F., Gonzalez, C., & Vakunov, P. (in preparation). Testing symmetry and context effects in similarity judgments for complex problem solving solutions.
Quesada, J. F., Kintsch, W., & Gomez, E. (submitted). Latent Problem Solving Analysis (LPSA): A theory of representation in complex problem solving. *Cognitive Science*.
Richman, H., Staszewski, J., & Simon, H. A. (1995). Simulation of expert memory using EPAM IV. *Psychological Review, 102*(2), 305-330.
Schraagen, J. M., Chipman, S., & Shute, V. J. (2000). State-of-the-art review of cognitive task analysis techniques. In J. M. Schraagen & S. Chipman & V. L. Shalin (Eds.), *Cognitive task analysis* (pp. 467 - 487). Mahwah, New Jersey: Lawrence Erlbaum associates.
Shanteau, J. (2001). What does it mean when experts disagree? In E. Salas & G. Klein (Eds.), *Linking expertise and naturalistic decision making*. Mahwah, NJ: Lawrence Erlbaum Associates.
Simon, H. A. (1981). *The sciences of the artificial*. Cambridge, MA: MIT Press.
Sloman, S. (2002). Two systems of reasoning. In T. Gilovich & D. Griffin & D. Kahneman (Eds.), *Heuristics and Biases: The psychology of intuitive judgment* (pp. 379-396). cambridge: cambridge university press.
Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*(1), 3-22.

Stanovich, E., & Cunningham, A. E. (1991). Reading as constrained reasoning. In J. R. Sternberg & P. A. Frensch (Eds.), *Complex problem Solving: principles and mechanisms* (pp. 3-61). Hillsdale, NJ: Lawrence Erlbaum Associates.
Stanovich, K. E. (2004). *the robot's rebelion: finding meaning in the age of Darwin*. Chicago: University of Chicago Press.
Tversky, A. (1977). Features of similarity. *Psychological Review, 84*(4), 327-352.
Tversky, A., & Gati, I. (1978). Studies of similarity. In E. Rosh & B. B. Lloyd (Eds.), *cognition and categorization* (pp. 79-98). Hillsdale, NJ: Lawrence Erlbaum associates.
Vicente, K. J., & Wang, J. H. (1998). An ecological theory of expertise effects in memory recall. *Psychological Review, 105*, 33-57.
Zipf, G. K. (1949). *Human Behaviour and the Principle of Least-Effort*. Cambridge MA: Addison-Wesley.