Problem Spaces in Real-World Science: What are They and How Do Scientists Search Them?

Lisa M. Baker and **Kevin Dunbar**
McGill University Department of Psychology
1205 Dr. Penfield Avenue
Montreal, Quebec H3A 1B1 Canada
lbaker@ego.psych.mcgill.ca or dunbar@ego.psych.mcgill.ca

How do scientists think, reason, and represent their knowledge? We have been investigating these questions in a variety of simulated science and real-world science domains over the last decade (Baker & Dunbar, in preparation; Dunbar, 1993, 1995, 1996; Klahr & Dunbar, 1988). Having explored scientific thinking in a variety of domains, we are now able to address the issue of the types of problem spaces that contemporary experimental scientists use. In this presentation, we will focus on the problem spaces that scientists in immunology and molecular biology use when they reason about their research at laboratory meetings and at the bench. On the basis of our analyses, we propose that real-world scientists represent and conduct their science using three major spaces: a Theory Space, an Experiment Space, and a Data Space. Here we will specify the nature of the three spaces, the criteria for identifying these spaces, and discuss the nature of between- and within-space operations.

The Three Spaces

Theory Space. This space includes specific hypotheses about the scientist's current research, general theories relating to different scientific domains, and theoretical frameworks that guide research.

Experiment Space. This space consists of the knowledge needed to conduct Experiments. The Experiment Space contains a number of different types of knowledge such as experimental approaches, materials, procedures and controls. In a later section on the internal structure of spaces we will specify how these different aspects of experimental knowledge are related.

Data Space. This space contains scientists' representations of the output from experiments. The Data Space consists of the scientists' current representation of their data and other scientists' data. Our conception of the data space is very similar to that of Schunn and Klahr's (1995) "data representation space." Data is represented as having certain sets of features and this representation may change at a later point in time.

Criteria for Identifying Problem Spaces

We used two main criteria to identify the problem spaces that scientists work in. First, when scientists are searching within a space, their search can be represented as a choice among specific features using classic domain-general search heuristics as well as space-specific heuristics. Second, we identified a switch to another space when complex reasoning processes were required to translate from elements of one space into elements of another. When there are no complex translations between conceptual entities, we classify the entities as being in the same space. We will elaborate on these criteria in a later section on within- and between-space operations.

Internal Structure of Problem Spaces

Current-day scientists have complex theories, design and carry out elaborate experiments, and have multifaceted data to interpret. Our analyses indicate that each of the three spaces has a complex, often hierarchically organized internal structure of elements. These elements do not operate at the same "level" but may nevertheless be considered part of the same space. In order to clarify the nature of the internal structure of problem spaces, we will provide an example of search in the Experiment Space.

Our example is taken from immunology where a scientist might be exploring the way in which cells send signals to each other. The scientist may have decided to use a "blocking" approach to investigate the role of cytokines in cell production of growth factors. In this scientist's area of research, there is an accepted structure for blocking experiments. This structure can be said to provide a "frame" with certain "slots." For example, having chosen the blocking approach, the scientist knows that the experiment will involve blocking antibodies mixed with cells in culture. However, the scientist still has choices to make in order to fill in different slots in the design of this experiment. For example, she must specify procedures (how long to incubate the cells?), materials (which antibody will she use?), and control conditions (what antibody and cell controls should she use?).

Within the Experiment Space, experimental approaches or paradigms are situated higher than materials, procedures, and controls in a hierarchy. This is because the paradigm dictates the structure of the experiment. To the extent that experimental paradigms and features of experiments operate at different hierarchical levels, they can be referred to as hierarchically organized "subspaces" of the larger Experiment Space. The subspaces consist of features of the higher space that have been unpacked.

Within- and Between-Space Operations

We distinguish between three classes of operations: between-space, between-subspace, and within-space. Between-space operations involve relating items in different spaces; for example, relating data to theory or theory to experiments. These operations typically involve complex translations, as when a scientist translates a hypothesis about the role of cytokines into a "blocking" experiment. Thus, while there is a relationship between the current hypothesis and the current experiment, the current hypothesis does not fully determine what experiments will be performed (see Baker & Dunbar, 1996). Between-space operations make use of what might be called the scientist's "mental toolkit" of reasoning strategies, such as causal reasoning, induction, deduction, and analogy.

In contrast to between-space operations, between-subspace operations do not involve complex translation processes. Rather, a choice at one level of a within-space hierarchy directly determines features in the subordinate subspace. In our example, the decision to use a "blocking" experimental paradigm determined the structure of the experiment and what slots would need to be filled in.

Within-subspace operations are the types of operations that have been emphasized in models of heuristic search within a problem space. Within-space operations are aimed at searching the current space and making choices among elements in the space. For example, Baker and Dunbar (1996) specify the complex criteria used to search the Experiment Space.

Why N of Spaces is not Constant

Different researchers' multispace models have emphasized the importance of different spaces and have even argued for different numbers of spaces. It is not surprising that different researchers arrive at different values for the N in "N-space search," because researchers are studying different task environments. Newell (1989) has argued that humans construct problem spaces in order to solve particular problems. Further, Newell pointed out that in studying different tasks researchers would be able to identify different spaces. Following this analysis, we argue that at least some of the differences between the different numbers of problem spaces is due to the task demands of the different tasks that researchers have used. Thus, we see the existence of various spaces as arising from the interaction of the human cognitive architecture and the particular task environment under consideration.

In the case of real-world science, we would even go one step further and argue that the three spaces we have identified have been established and stabilized in part through the social interactions that go on in this task environment. Garrod & Doherty (1994) have shown how communication between subjects performing a task can result in task-specific terminology being generated and stabilized. Similarly, we argue that scientists stabilize the Theory Space, Experiment Space, and Data Space by referring among themselves to elements of these spaces as being of different kinds. There are many possible ways science could be and has been done; for instance, science can be observational rather than experimental. The three spaces that we have identified are a product of both the task environment and the representational practices of the general scientific community.

Given that we allow a role for both task demands and the social structure of science, we expect that both the number and types of problem spaces that scientists use can change depending on the task environments and social practices of science. While acknowledging the roles of these two constraints on the generation of problem spaces, we argue that the three-space model of scientific thinking best characterizes "real-world current-day experimental science."

Acknowledgments

This research was supported by a McGill Major Graduate Fellowship to the first author. The second author was funded by grant number OGP0037356 from NSERC and grants from FCAR and SSHRC.

References

Baker, L. M., & Dunbar, K. (in preparation). How scientists deal with unexpected findings: A discovery heuristic.

Baker, L. M., & Dunbar, K. (1996). Constraints on the experimental design process in real-world science. In *Proceedings of the 18th Annual Conference of the Cognitive Science Society*.

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

Dunbar, K. (1995). How scientists really reason: Scientific reasoning in real-world laboratories. In R.J. Sternberg & J. Davidson (Eds.), *Mechanisms of insight*. Cambridge MA: MIT Press. pp 365-395.

Dunbar, K. (1996). How scientists think: Online creativity and conceptual change in science. In T. B. Ward, S. M. Smith, & J. Vaid, *Conceptual structures and processes: Emergence, discovery, and change*. Washington, DC: American Psychological Association Press.

Garrod, S., & Doherty, G. (1994). Conversation, co-ordination and convention: an empirical investigation of how groups establish linguistic conventions. *Cognition, 53*, 181–215.

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

Newell, A. (1989). Putting it all together. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon*. Hillsdale, NJ: Lawrence Erlbaum Associates. pp 399-440.

Schunn, C. D., & Klahr, D. (1995). A 4-space model of scientific discovery. In *Proceedings of the 17th Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.