The Relationship Between Direct and Data-Mediated Knowledge of the World

Kim Kastens, 12 January 2014

The Oceans of Data Institute has developed and is testing a hypothesized learning progression toward "data scientist," which involves passing through four domains: (A) children observe the real world with their human senses in an unstructured way, (B) students work with small datasets that they collected themselves, (C) students analyze and interpret large, professionally collected datasets in the context of well-structured problems, and (D) students or professionals analyze and interpret large, professionally collected datasets in the context of ill-structured problems (Figure 1).

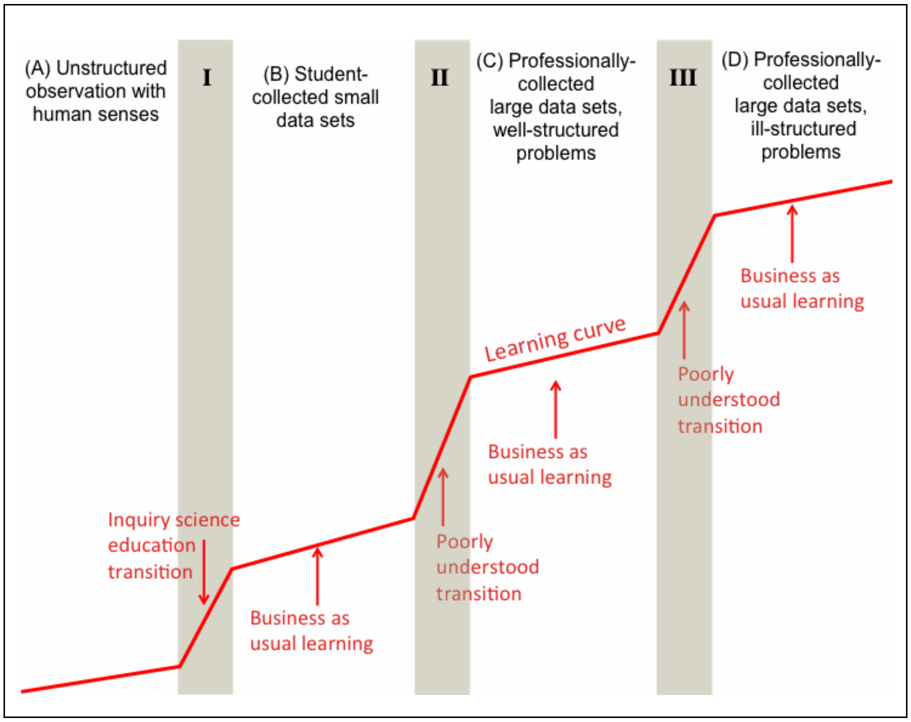

Figure 1 sketches the broad outline of a sequence of stages that could culminate in an individual who can use large, professionally collected datasets to solve the sort of ill-structured, complex problems that abound in adult life.

In domains (A) and (B), people are directly engaged with materials and phenomena of the real world, whereas in domains (C) and (D) they are engaged with representations.

While working in domains (A) and (B), learners are directly engaged with materials and phenomena of the real world, whereas in domains (C) and (D) they are engaged with representations. Our framework for thinking about this difference comes from developmental psychologist Lynn Liben's work on children's understanding of maps, a form of representation that is used for both scientific and practical applications. Liben (1996, 2006) presents an "embedded model" (Figure 2) of the relationships among a child, an external (i.e. not mental) representation, and those aspects of the real world represented by the represenation (the "referent").

In Liben's work, the external representation is a spatial representation, aka a map. In this white paper, I extend Liben's framework to any kind of visible representation, and in particular to data representations, aka data visualizations. Liben's work refers to the viewer/observer as "the child"; here I extend the model to "learners" of any age, be they in school or informal or professional settings.

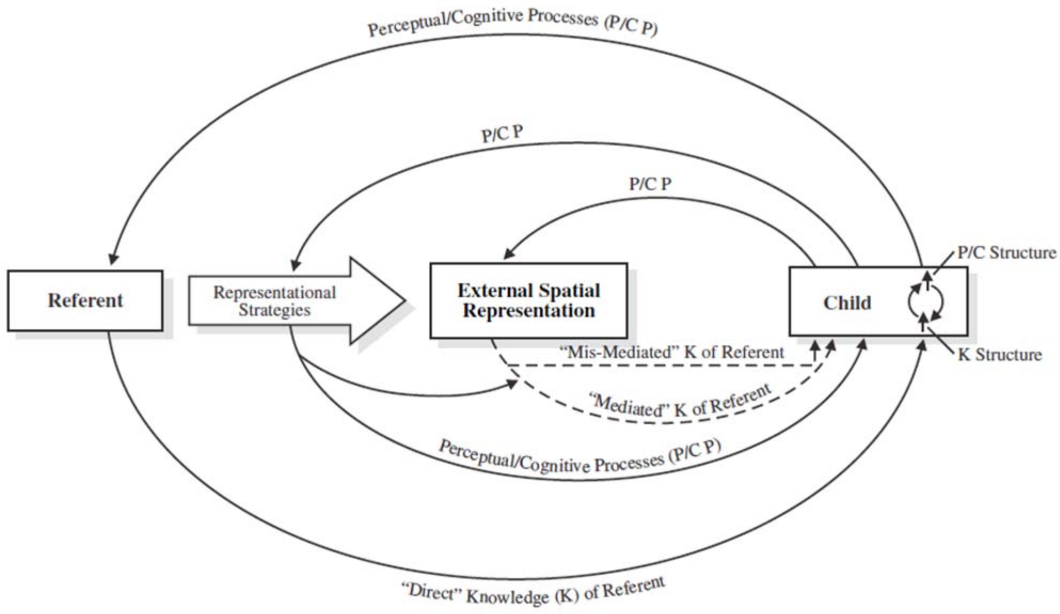

Figure 2: The work of Liben (1999, 2006), summarized in this figure, provides a framework for thinking about the relationships among data ("External representation"), that portion of the real world of which the data is a representation ("Referent"), and the data viewer/interpreter ("Child").

In Liben's model, the learner reaches out to interact with the real world through his or her perceptual and cognitive processes. The products of these interactions lead to "direct" knowledge of the structures, phenomena and processes of the real world (the outer yellow loop of Figure 3). Such direct interactions with the real world characterize domains (A) and (B) of the data-mastery learning progression (Figure 1). However, in domains (C) and (D) of the data learning progression, the learner is reaching out with cognitive and perceptual processes not to the real world referent, but rather to a human-made representation (inner orange loop of Figure 3). Obtaining direct knowledge of the world via the outer yellow loop of Figure 3 is an evolved ability that humans share with other animals; not so for obtaining knowledge of the world via external representations made by others.

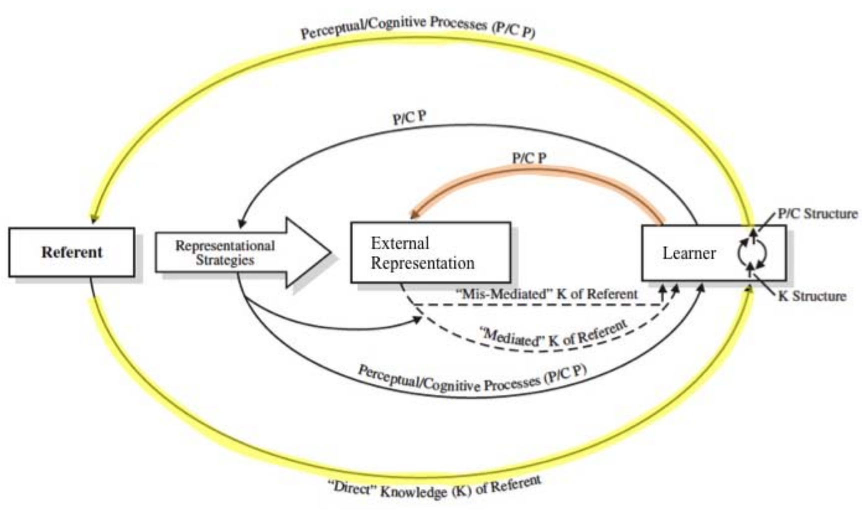

Figure 3: When learners acquire knowledge by direct observation of the world around them, they are using the outer, yellow-tinted pathway in Liben's model. But when they work with visualizations of data that they did not themselves collect, they are using the inner, orange-tinted pathway.

The external representation is connected to its referent by means of "representational strategies" (large green arrow of Figure 4). Liben's key insight was to point out that the representational strategies themselves must be explored through perceptual-cognitive processes; this process results in "direct knowledge of representational strategies" (the middle loop, green in Figure 4). If the representational strategies employed in generating a specific representation are well understand and properly applied, then the learner can gain veridical "mediated" knowledge of the referent by looking only at the representation, without looking at the referent. But if the representational strategies are not well understood or are wrongly applied, the learner gains "mis-mediated," and typically erroneous, knowledge of the referent.

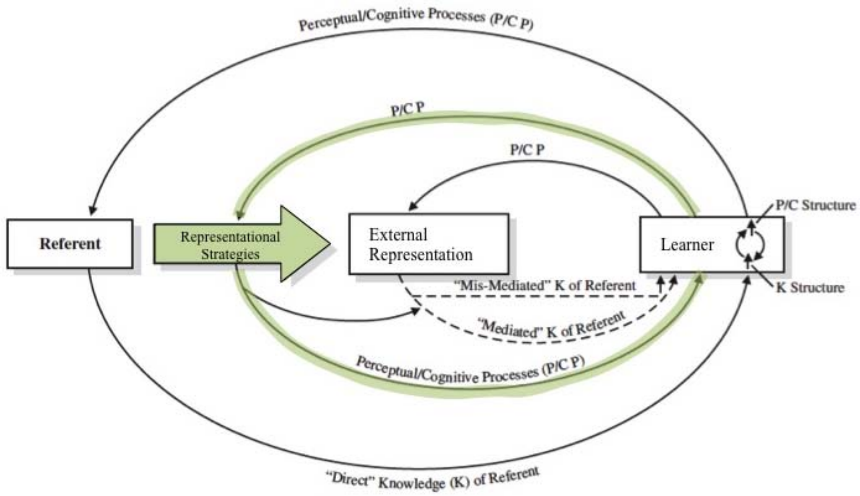

Figure 4: The middle loop of Liben's model, highlighted in green, shows that representational strategies must also be explored through perceptual and cognitive processes.

Liben's examples of representational strategies were drawn from cartography. For example, a child with a weak grasp of cartographic representational strategies might expect that a road represented by a red line on a map would correspond to a road that is actually paved with a reddish material in the real world. I would like to extend Liben's concept of representational strategies to encompass the myriad ways that scientists have devised to depict facets of enormously complex universe in forms that can be comprehended by the human mind. Representational strategies include using the two dimensions of a piece of paper to represent the two spatial dimensions of the surface of a planet (i.e. in a map), but they also include using one dimension of the paper to represent time, or using the rows and columns of a spread sheet to represent individuals and their attributes, or any of the other ways that researchers have devised to visualize data.

Moreover, as I see it, "representational strategies" also encompass the techniques and instrumentation that humans use in extracting or capturing the attributes of the real world in the first place. Latour (1986) and Goodwin (1994) refer to making "first inscriptions" (Latour, 1986; Goodwin, 1994), as the process of turning the raw, ephemeral material of nature into enduring, transportable human artifacts, the first step in the "cascade of inscriptions" that are the work of science. So my expansion of Liben's concept of "representational strategies" includes the instrumentation used to make first inscriptions (for example, a thermometer, or a theodolite or a sonar system) and the techniques for using them, plus the techniques used for turning the first inscriptions (aka raw data) into the visualization eventually seen by the learner.

Liben's concept of mediated and mis-mediated knowledge of the referent fits this expanded view of "representational strategies." For example, if a student thinks that a data visualization is a photograph, but in fact it is a representation of sonar data, then there is a risk that the knowledge she constructs from the representation will be incorrect, will be "mis-mediated knowledge."

The final element of Liben's model occurs within the learner (Figure 5, pink shaded arrows). Knowledge structures in the learner's mind interact with perceptual/cognitive structures, also in the mind, in two ways (at least). First, knowledge from previous learning episodes steers how the learner goes about directing his/her perceptual attention and cognitive effort in the current learning episode. This is why you tend to see what you expect to see, because you steer your attention and thinking guided by what you already know and don't know.

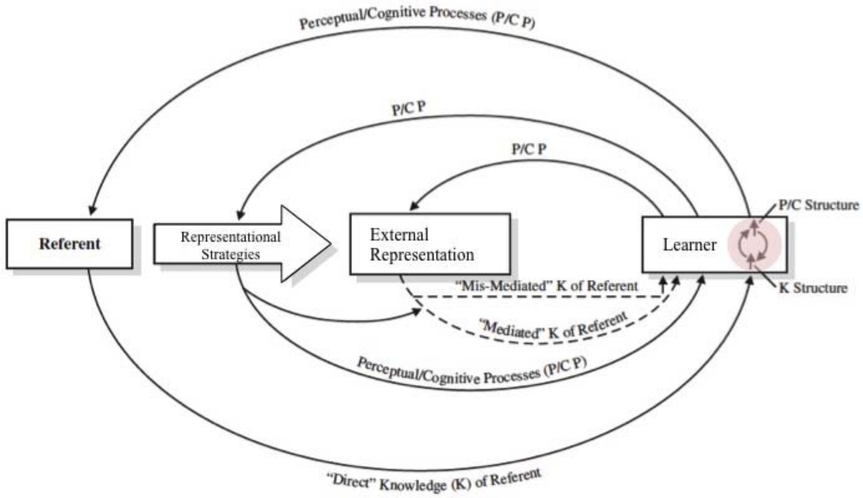

Figure 5: In the mind of the learner, knowledge structures interact with perceptual/cognitive structures (pink-tinted arrows). Prior knowledge about the referent system combines with visually available information from the representation, plus knowledge of representational strategies to enable the construction of new knowledge about the referent.

Secondly, inside the mind, previously acquired knowledge about the referent interacts with currently available percepts. Learning from data involves a chicken-and-egg problem. Without knowing something about the processes or structures of the referent (or about systems that are similar or analogous to the referent system) it is extremely difficult to make any kind of meaning from a data representation. To make any kind of sense about data from a river, for example, one needs to know some minimum basic attributes of rivers, for example, that they are long, skinny depressions or troughs in the earth's surface filled with water, and that the water is moving, that in general the water moves from higher elevation to lower elevation, etc. These river attributes constitute a kind of zeroth order mental model of riverness, without which little or no learning from river data can take place. Koslowski (1996) characterizes scientific reasoning as requiring "bootstrapping," such that "considerations of theory or mechanism constrain data, and data in turn constrain, refine, and elaborate theory" (p.86). You need to know something about the referent to understand the data--and yet data representations are the vehicle with which scientists communicate their understanding of the referent system (Kastens & Manduca, 2012).

To summarize, in domain (A) of our hypothesized learning progression (Figure 1), the learner is functioning entirely in the outer loop of Liben's model (yellow loop in Figure 3). In domain (B), the child is working to master one or more of the representational strategies that capture a complex real world referent into some kind of representation. However, any interpretation that he or she might be asked to do is strongly scaffolded by the presence of direct knowledge of the referent system acquired during the process of data acquisition. Upon entering domain (C) of the data learning progression, that scaffolding has been removed. In order to make inferences about the referent in domain (C), the learner must combine three different kinds of information:

1) information that is perceptually available in the data visualization,
2) previously learned knowledge of appropriate representational strategies, how they work, their affordances and limitations; and
3) previously learned knowledge of the structures and processes of the referent system.

Acknowledgement: many thanks to Lynn Liben for sharing her thoughts on how people understand and learn from external representations throughout many years of collaboration, and for reviewing this white paper.

References Cited

Goodwin, C. (1994). Professional vision. *American Anthropologist, 96*, 606-633.

Kastens, K. A., & Manduca, C. A. (2012). Fostering knowledge integration in geoscience education. In K. A. Kastens & C. Manduca (Eds.), *Earth & Mind II: Synthesis of Research on Thinking and Learning in the Geosciences, Geological Society of America Special Publication*. Boulder: Geological Society of America.

Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning.* Cambridge, MA: MIT Press.

Latour, B. (1986). Visualization and cognition: Thinking with eyes and hands. *Knowledge and Society, 6*, 1-40.

Liben, L. (1999). Developing an understanding of external spatial representations. In I. E. Sigel (Ed.), *Development of mental representation: Theories and applications* (pp. 297-321). Hillsdale, NJ: Lawrence Erlbaum Associates.

Liben, L. (2006). Education for spatial thinking. In K. A. Renninger & I. E. Sigel (Eds.), *Handbook of child psychology, sixth edition, volume four: Child psychology in practice* (pp. 197-247). Hoboken, NJ: Wiley.