Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Reinterpretation and the Perceptual Microstructure of Conceptual Knowledge

Permalink

https://escholarship.org/uc/item/4fk4t7m2

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 11(0)

Author

Shrager, Jeff

Publication Date

1989

Peer reviewed

Reinterpretation and the Perceptual Microstructure of Conceptual Knowledge

Cognition Considered as a Perceptual Skill

Jeff Shrager
Xerox PARC
Palo Alto, California

Abstract

In this paper I argue that conceptual knowledge has significant perceptual content, based upon evidence from studies of theory formation and from recent experimental work on learning in complex physical domains. I outline a theory of "perceptually grounded" conceptual knowledge, and briefly outline a computational model of learning about lasers, in which student's "qualitative" understanding of lasers rests primarily upon his perceptual experience in the domain.

Introduction

The goal of this research is a detailed computational theory of conceptual representation and use. Previous theories (e.g., Schank & Abelson, 1977) have been based primarily on symbolic representational substructure. Pavio (1986) and others have argued for a mixed cognitive representational framework, but these theories still rely upon an independently meaningful symbolic representation. In this paper I argue from our studies of "reinterpretation" during theory formation that our conceptual knowledge must be *perceptually grounded*, and propose a theory of conceptual representation based upon such grounding.

Studies of Theory Formation

Shrager & Klahr (1986) gave college students a programmable toy vehicle called the "BigTrak" and asked them to "figure it out" without instructions or advice. In the course of about one-half hour, subjects undertake numerous steps of theory refinement and reformulation. In some of these events subjects seem to reformulate their theory of the BigTrak in ways that introduce new terms and representational principles.

Consider the segment of protocol in Table 1 (studied in detail in Shrager & Klahr, 1986). By programming the BigTrak with "$CLR, \rightarrow, 1, \uparrow, 2, GO$" (FC122-127) FC caused the toy to move six degrees clockwise and then two feet forward. From previous behavior we believe that FC thought that this would make the BigTrak move one foot to the right and then one foot forward. Figure 1 shows (a) what we think FC expected, and (b) what the BigTrak really did.

A reformulation step takes place at this point of mismatch between FC's expectation and the behavior of the BigTrak. Around FC127 we hypothesize that he does the following:

1. *Recognizes* that the behavior of the BigTrak matches his expectations when mediated by the concept of vector-addition;
2. *Introduces the concept* of vector-addition into his theory of the BigTrak, including "resultant" and "component" terms, and the associated representational principles; and

```
115:        Does it...I don't know maybe it remembers things or
116:        something so that...it just did the same thing I told it
117:        to do last time even though I pushed different buttons.
118: CLR
119: CLR    Alright.  I guess you can like oh I see (?) program
120:        steps into it or
121:        something like that. So if you push ummm...
122: CLR
123: →      right one
124: 1      then forward
125: ↑      two
126: 2
127: GO     Went straight and right a little bit.
{Here the BigTrak turns right 6 degrees and moves forward 2 feet.}
128:        Oh. I see that's the resultant
129:        thing maybe.  I don't know.
```

Table 1: A part of FC's BigTrak Instructionless Learning Protocol.

3. *Reformulates his knowledge* of the BigTrak and its operation in accord with the new terms and representational principles introduced in the preceding step.

This reformulation results in FC having a theory of the BigTrak which is a combination of his previous knowledge and his general knowledge of vector-addition. FC's introduction of vector-addition seems to be *rapid*; to apply *as a unit, without intermediate problem-solving*, to his understanding of the device; and to *augment* and serve as a *reorganizing principle* for his understanding of the device and its behavior.

Problems with Concept Use in Theory Formation

We previously proposed that theory changes of this sort involve a cognitive mechanism that we called "View Application" (Shrager, 1987), whose role was to *reinterpret one's knowledge in terms of newly uncovered abstractions* (i.e., "views"). Implementing View Application in a symbolic representational framework leads to two particular problems:

The Paradox of Recognition: How can views containing novel terms and relations be recognized as applicable to the current domain if some of those terms and relations are not *already* available in the learner's current theory? We seem to depend upon *prearticulated sensory data* for view selection, but we must wait for the perceptual framework given by a view in order to obtain these articulations.

The Framework Alignment Problem: How can semantic contact be made between terms and relations in the learner's current theory and those in a novel view without common terms shared between theory and view, or rules of translation between terms in the theory and those in the view? In the worst case, terms introduced by the new view may simply be incommensurable with those in the theory to be reformulated in accord with the new view.

Since, for instance, the vector-addition view is the locus of the vector and resultant terms and representations, the subject must have noticed these terms and this representation in the activity of the BigTrak *before* choosing the vector-addition view. However, I previously claimed that the view application step introduces these terms and representations into the learner's theory. This is an example of the paradox of selection.

The combination step of View Application suffers from the framework alignment problem. When the view applier begins to reformulate the learner's current theory according to the new view, it must make "representational contact" between aspects of the view (say, the individual vectors) and the aspects of the learner's current theory (say, movement commands). That is, *differently represented terms which are about the same real-world thing(s)* must be located and their relationship made available to the view applier.

These difficulties seem to stem from our tendency to think of views and theories in terms of schematic internal knowledge in the form of models composed of categorical terms and relations. These categorical entities (which when under interpretation are generally referred to as "symbols") are captured in computational models in the form of scripts, frames, schemas, views, etc. The connection that must hold between the world and the symbolic structures in order that they are *operational* is usually ignored or relegated to the "peripheral" roles of perception or motor activity. This over-reliance on internal and ungrounded knowledge has led theories of mental model formation to be overly rigid, entirely missing the ability to reinterpret experience as *experience per se is nowhere to be found!*

Perceptually Grounded Conceptual Knowledge

The theory of grounded representation rests upon the the following fundamental claim:

Perception and perceptual experience form the basis of conceptual knowledge.

Specifically, we replace symbolic representation in frames, views, scripts, etc, with a set of "*synchronization routines*" that mediate between traces in one modality (say echoic, visual, or motor traces) and traces in another (or the same) modality.[1] Knowledge thus consists of *skills* of *identifying* (and often *naming*) relevant features and concepts, and more importantly, *skills* for *acting* (i.e., executing appropriate actions) with respect to these entities.[2]

The basic approach to the framework alignment problem and the paradox of selection, provided by the grounded representation framework, is that knowledge that is "carried" in different representational frameworks can be compared by understanding how they differentially interpret the experiences that compose their grounding. A central cognitive role is given to experiences themselves (or to quasi-perceptual traces of experiences themselves).[3]) Not only is a picture worth ten-thousand words, but it may be described in perhaps ten different ways, at say a thousand words per description. If each of these thousand-word descriptions is "grounded" on the picture, then we can compare these different descriptions to one-another by reference to the picture itself.

---

[1] *Modalities* are the substrate of *representations*, but representational structures operate under interpretation. Both algebra and linguistic inference rules, for instance, might be represented in a quasi-linguistic modality; and both static images and physical animations might be represented in an iconic modality.

[2] This approach is reminiscent of the dual-coding approach of Pavio (1986), but Pavio proposes only associations between codes, whereas the present theory makes the stronger proposal of inter/intra-modality synchronization *procedures*. Our theory is a cognitive analog to the theory of visual routines proposed by Ullman (1984).

[3] By the term "quasi-perceptual traces" I mean some poorly-understood combination of deep motor representation and animated-imagery. However, as I haven't any real idea what this deep quasi-perceptual modality might be like, my computational implementations use bitmap animations (ala Funt, 1980). There is a difficult issue here of how a *procedure* or its input, output, or parameters exist such that they can be "examined"

The approach to paradox of selection, suggested by the present theory of grounded representation, is similar: As all knowledge is grounded to experience (or to traces, as above), one can find the desired features for selection in the experiences themselves (or the traces) – among one of the those thousand-word descriptions of our picture. Thus, we do not have to rely upon finding these terms in the learner's present theory.

Let us return to the event where FC's recognizes the BigTrak's movements as a vector resultant. Note that the triangle made by the two component arms (1a and 2a in Figure 1) and the resultant (2b) form approximately the "image" that most of us who have taken formal trigonometry associate with vector-addition. The claim is, then, that what we know of vector-addition includes a procedural recognizer for this image and that it is through this path that we come to "recognize" the possible use of the process of vector-addition in the BigTrak's activity.

To see how perceptual representation helps with framework alignment consider the process of introducing the notion of "memory" into one's theory of the BigTrak. This commonsense concept may be suggested by observing that the BigTrak "did the same thing I told it to do last time even though I pushed different buttons" (Table 1: FC115-118). The application of this view involves reunderstanding the procedure of pressing keys on the BigTrak as storing things in the memory, and the internal activity of the BigTrak as reading out the contents of that memory, and executing it. The quasi-perceptual nature of knowledge gives a straightforward account for this reinterpretation: The representation of these procedures, and of our understanding of what goes on inside the BigTrak's memory is "active" in a kinesthetic or animated sense (or, more likely, both – but certainly far from actual visual perception or physical motion). The BigTrak's mechanism is thus thought of as actively placing (iconic representations of) the BigTrak's actions into (iconic representations of physical) memory slots that have (quasi-perceptual) spatial organization with respect to one-another. Similarly, introducing vector-addition into FC's knowledge of the BigTrak entails bringing in the *procedural skills* of locating and reasoning with aspects of vector addition. (See Shrager, in press, for further details of this sort of model.)

Studying the Perceptual Content of Conceptual Knowledge

We are presently developing paradigms which will both help to reveal the specific quasi-perceptual content of conceptual knowledge, in accord with the above theory, and to provide support the theory. Here I describe a study of learning about laser physics (quantum optics) which serves both goals to some extent.[4]

Overview: We wish to observe the development of students' interpretation of quantum optics via learning about lasers. We wish to specify the "conceptual" (actually, in the present theory, quasi-perceptual) chunks and skills that the students pick up and use in learning and explanation in this domain. We employ a number of methods, ranging from eye-tracking during study and explanation, to reconstructive tasks (as Chase & Simon, 1973), mostly augmented by verbal protocols. In the present paper I report preliminary results from a reconstruction task.

Method: Undergraduates with little physics background were given approximately three hours of instruction in laser physics over four sessions one week apart, including reading textual materials with figures, and answering summary questions. At the end of each session the subject was asked to copy twenty 8.5x11 figures from the page on which they were drawn, onto the next page of the test booklet. The subject could look back as many times as necessary to complete the drawing, but was required to turn the pages fully either to draw or to look back. Complete protocols were

---

[4] This domain is a good one in many ways. Although everyone knows approximately what a laser is, and is interested in them, almost no one knows how they work. Furthermore, quantum physics is a rich domain but is relatively separate from real world experience and so is easily manipulated.

only collected for the first subject ("J"); here I will focus on her behavior.

Figure 2 shows the stimulus design and predictions from the copying task. Four of the figures in the copying task came directly from the textual materials that the subject had studied (all textual labels in the figures were deleted). Six were *analogs* of the figures in the laser text (drawn from a different textbook). The rest came from a book on computer vision, and were unrelated to laser physics.

Results: J is able to haltingly but correctly answer the relatively difficult questions after each session, sometimes by reference to the text figures. One study question (appearing after 46 minutes of study) asks: "In your own words explain what role the mirrors play in making a laser work. Do the same for stimulated emission." J says:

> We have this uh this discharge tube that has gases inside it...a gas mixture inside it and there are two mirrors on either end. One is like 100 percent reflective and the other is like 95 percent reflective or whatever. [...] Okay, they are spontaneously emitting photons and after a while they keep on doing this and eventually one will hit the mirror and it will bounce back ... it'll hit the mirror head on and it'll bounce back and it will keep on doing this and after a while um it'll start collecting other photons and like stimulating emission, creating them to give off more photons and then soon [...]"

Figure 3 contains the originals (left) and J's first copies (right) for one image from each category in the copying task. *A* is a portion of a figure from the text (less labels) showing a laser cavity in operation. *B* is an analogous representation of a laser cavity in operation, but one which J did not see in her reading. *C* is a figure taken from computer vision. J required 157 seconds and 2 page-turns to copy *A*, 195 seconds and 4 turns for *B*, and 166 seconds and 5 turns for *C*. The summary statistics in Figure 2 (from our preliminary analyses for J) confirm the expected trends.[5]

During her copying of *A*, J says: "Here we have d and e ... the photons and the mirrors. [...] This is an example of that stimulated emission and these [referring to some of the dots] are just spontaneous emission[...]." When she draws the mirrors she says: "Here we have mirror one and here was have another mirror [...]." At one point, when she draws the second from the left photon vector on part "d" of the figure, she says: "Oh no'd better make it look in another way to show uh spontancity [...]." Most of her reference is in terms of laser dynamics: "photons" "stimulated emission", etc. She clearly recognizes the time sequence implicit in the ordering of the similar parts of the figure.

During her copying of *B*, J says: "[...] these look like springs or something. [...] There are two identical, I guess I identical [...] except for three lines here...wavy lines...c and b, I wonder if that has any significance whatsoever." At one point she says: "I guess I should draw some of the dots...I don't know why I just feel like without them this picture, I don't know if it would make much sense [....]" Most of her reference is in terms very close to the image structure: "wavy lines" "dots" "inside", etc.

During her copying of *C*, J says: " Oh, this looks like uh one of those games that you used to play where you have a bunch of marbles and you have to get em through [....] nebulous shape here [....] one circle going in here, one going off [....] going in through this hole [....]" Most of her reference is in terms very close to the image structure: "circles" "here" "holes", etc.

---

[5] There is much complexity in interpreting the quantitative results of this study. We must balance for figure complexity and repeated exposures. The analyses presented here are not so controlled and thus can only be considered preliminary.

Observations: From her explanation, J seems to have learned the "conceptual" material in the text. Also, although she recognized and "conceptually interpreted" figures that she had seen, but not a novel figure (unsurprisingly) or a close conceptual analog to one she had seen *even though she was trying to interpret these other figures in terms of lasers.* At several points during the copying tasks J says (in paraphrase): "I'm trying to figure out what this picture has to do with lasers."

From her verbal protocol it is clear that J recognized A as the laser cavity with a process of stimulated emission taking place, and that she did not recognize this for B, even though she was actively trying to interpret this figure. Her copy of A is a *"semantic" analog* rather than a close *visual analog*, whereas her copies of B and C are close visual analogs. (Note, the care with which B and C were copied, versus A.) This is supported, as well, by the fact that she refers to A mostly in (laser) meaningful terms, but not so B or C. Interestingly in later sessions, after she has learned that photons are sometimes (visually) represented in a wave-like way she refers to the "springs" of her verbalization above (B) as photons, but still does not interpret the figure as a laser cavity.

From her apparent conceptual understanding of the laser, her failure to "conceptually" interpret conceptual analogs, but the more efficient performance in their reconstruction (over out-of-domain figures), we can argue that much of what seems to be her "conceptual" knowledge has (quasi-) perceptual content. (Always remembering that I include in this *skills* of interpretation and action in the domain.)

Discussion: Cognition Considered as a Perceptual Skill

I have argued on theoretical and empirical grounds that perceptual content underlies our conceptual knowledge of the world. FC is able to "see" a vector-resultant (and the process of vector-addition) in the activity of the BigTrak; Although J seems to "conceptually" understand the operation of the laser, she is unable to interpret figures that were closely analogous to ones that she directly experienced (and which showed precisely the same "conceptual" information), even though she is clearly trying to impose an interpretation on them that would make them sensible in terms of lasers. The paradigms used in the present research may enable us to observe the fine perceptual substructure of what we call conceptual knowledge, and the details of its functioning in learning and interpretation.

A theory of the microstructure of category representation and use that retains content quite close to the perceptual exemplars from which we learn these concepts is not an entirely new theory of concept structure. It echoes exemplar-based theories (e.g., Medin & Schaeffer, 1978) and the dual-coding approach of Pavio (1986, see also Huttenlocher, 1968), however within a significantly more procedural framework.

Our current modelling efforts follow the reasoning in this paper. We have implemented "qualitative" simulation of laser processes which (a) learns about how lasers work using approximately the same information – particularly the figural information – that our experimental subjects have, and (b) which can reason about the lasing process (Shrager, in press). This model contains two "working memories" in different modalities: an iconic (bitmap) memory in which animations take place (ala Funt, 1980), and a "symbolic" (quasi-linguistic) memory in which explicit (rule-based) inference takes place. These are synchronized by inter-modality (inter-memory) "grounding" routines. Learning takes place by introducing routines specific to the application at hand, which serve to *label* the contents of the iconic memory (by making appropriate changes in the symbolic memory), and conversely, to make appropriate changes in the iconic memory whenever inference (or any other change in the symbolic memory) takes place.

Acknowledgments

Thanks go to the reviewers of this paper, and to many of my colleagues for comments on the paper and the research. I especially thank Sharon Lunt for helping me understand the details of quantum chemistry.

References

Chase, W.G. & Simon, H.A. (1973) Perception in Chess & The Mind's Eye in Chess. papers reprinted as chapters 6.4-6.5 (pp. 386-427) in Simon (1979) Models of Thought. Yale University Press; New Haven.

Funt, B. V (1980) Problem-Solving with Diagrammatic Representations. Artificial Intelligence, 13, 201-230.

Huttenlocher, J. (1968) Constructing Spatial Images: A Strategy in Reasoning. Psychological Review, 75(6), 550-560.

Kosslyn, S.M. (1987) Seeing and Imaging in the Cerebral Hemispheres: A Computational Approach. Psychological Review, 94(2), 148-175.

Medin, D.L. & Schaeffer, M.M. (1978) A context theory of classification learning. Psychological Review, 85, 207-238.

Pavio, A. (1986) Mental Representations: A Dual Coding Approach. Oxford.

Schank, R. & Abelson, R. (1977) Scripts, Plans, Goals, & Understanding. Lawrence Erlbaum; Hillsdale, NJ.

Shrager, J. (in press) Commonsense Perception and the Psychology of Theory Formation. in Shrager & Langley (eds.) Computational Models of Discovery and Theory Formation. To appear from Lawrence Erlbaum; Hillsdale, NJ.

Shrager, J. (1987) Theory Change via View Application in Instructionless Learning. Machine Learning, 2: 247-276.

Shrager, J. & Klahr, D. (1986) Instructionless learning about a complex device: the paradigm and observations. Int. J. of Man-Machine Studies, 25, 153-189.

Ullman, S. (1984) Visual Routines. Cognition, 18, 97-159.

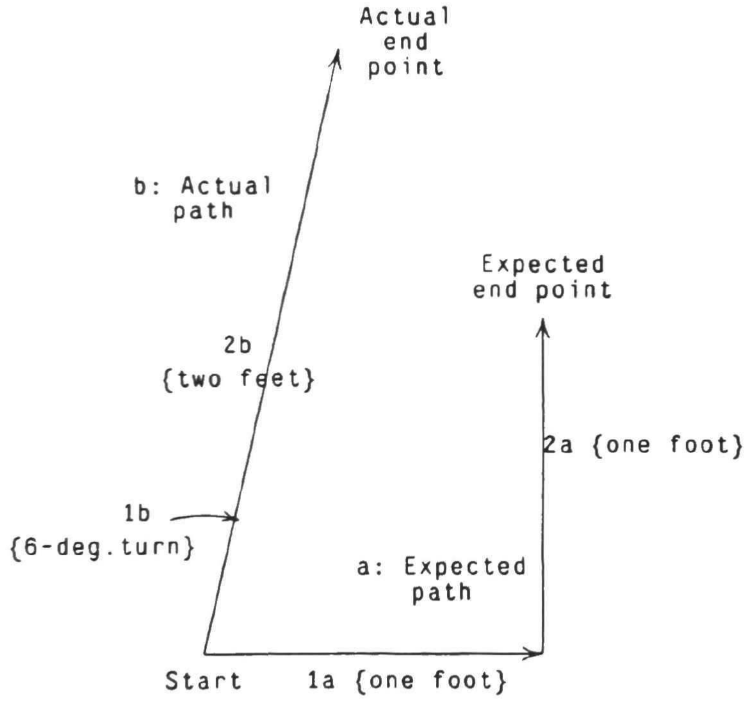
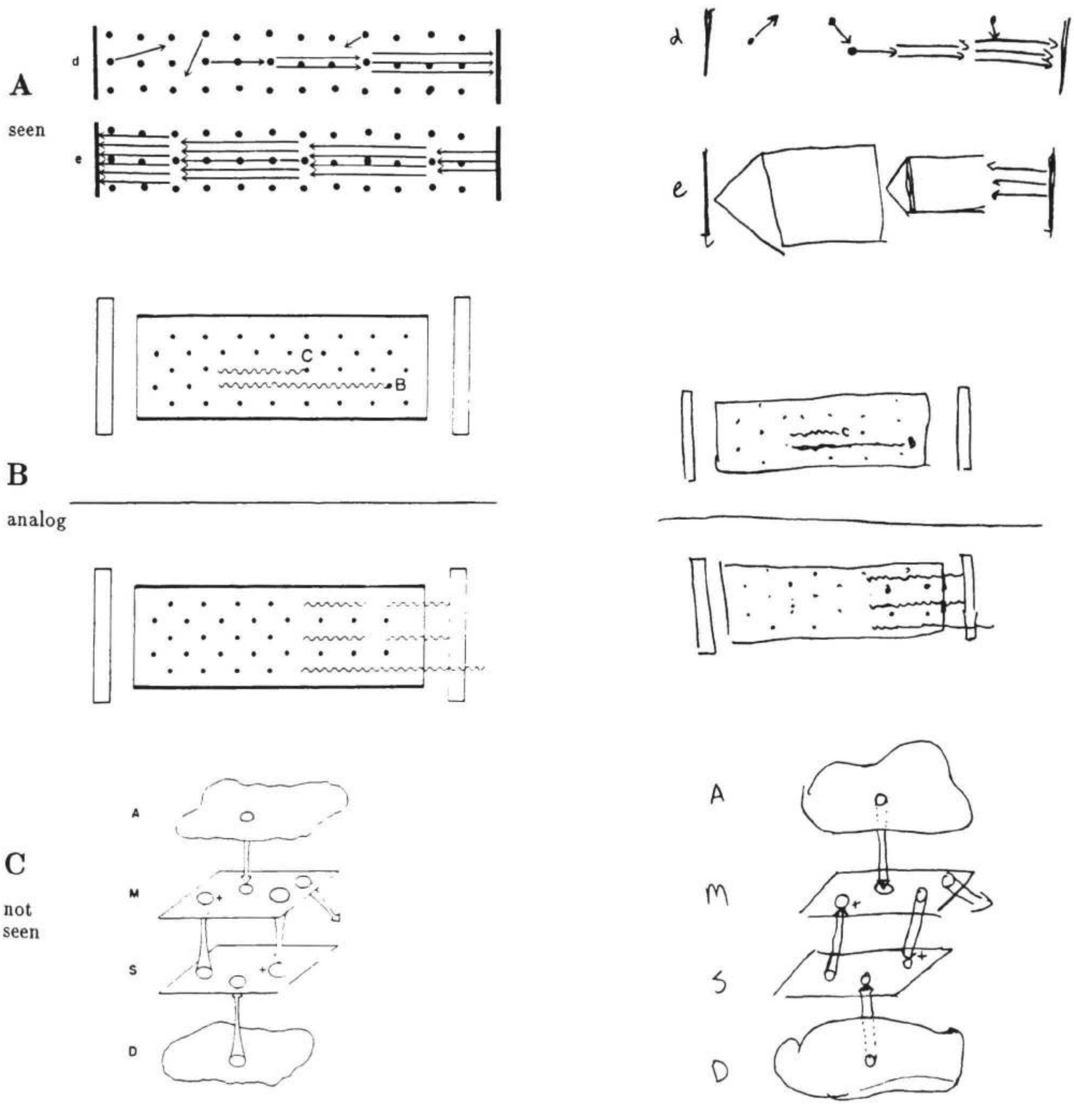


Figure 1: The BigTrak Actions (a) Expected by FC, and (b) That Actually Took Place at FC127.

	In Domain (Lasers)	Not In Domain
Seen	**Exact Figures** Fewest Page Turns Least Exact Images Large "Meaningful" Chunks Time = ~100 s, Turns = ~2	(impossible)
Not Seen	**Analogous Figures** Medium Page Turns Medium Exact Images Medium Image Chunks Time = ~135 s, Turns = ~ 2.5	**Figures from Vision** Most Page Turns Most Exact Images Small Image Chunks Time = ~170 s, Turns = ~5

Figure 2: Design, Predictions and Preliminary Summary Data



1 - Original (Stimuli)

2 - Subject's Reproduction

(retraced for clarity)

Figure 3: Selected Results from J's First Copying Task