EXTERNAL REPRESENTATIONS AND

THE ADVANTAGES OF EXTERNALIZING ONE'S THOUGHTS

DANIEL REISBERG

REED COLLEGE

We consider some of the functional differences between internal and external representations. In particular, we argue that different knowledge and skills can be brought to bear on external representations than on internal ones; that external representations, as non-intensional entities, are open to reinterpretation in ways that internal representations are not; and that one can discover omissions from external representations that cannot be found in internal representations, suggesting that attention is deployed differently in these two cases. We consider several implications of this view, including the advantages to be gained by externalizing one's thoughts, i.e. by writing out the content of mental propositions or by sketching the content of mental images.

A vast amount of cognitive science is concerned with the formation and use of mental representations. These representations may be "distributed" or "symbolic;" they may be "frames" or "prototypes" and so on. However we conceive them, though, internal representations indisputably play an essential role in cognition. In the present paper, I will argue that there is also an immensely important function served by external representations. This function is not limited to the obvious role that external representations play in communication: To convey my thoughts to you, I must translate them into words or pictures; you must decode these to understand me. Beyond this, though, external representations also serve an important non-communicative function, in the examination and development of one's own ideas. As we will see, the function of external representations derives from the particular nature of internal representations, and, in fact, from the limits of internal representations.

External representations are often used an an aide memoire or as an aide pensèe. The scientist writes the hypothesized equations on the blackboard; the artist sketches the lay-out in a notebook. Or, to take a rather different case, I spent some months contemplating the contents of this paper before sitting down to write it. Yet much of the discovery and development happened only when I began to write. I believe this is a common experience -- that translating one's thoughts into words can be extremely instructive; indeed, I will try to argue that some of the instruction is only possible once the thoughts are externalized.

Before proceeding, it is worth noting the diversity of what we are calling external representations. The examples just listed include pictures, equations on a blackboard, and long papers; some of these are symbolic, some are depictive. These are different from each other in many regards, and, for many purposes, can not be grouped together. Nonetheless, I will treat them as interchangeable for present purposes. Indeed, much of the evidence which informs the present discussion concerns a single paradigm case -- the contrast between mental images as internal representations, and pictures as external ones. The issue of whether we can generalize from this case will be a central theme in what follows.

There are of course some trivial reasons why it is helpful to externalize a thought. Working memory is limited in capacity, setting boundaries on how much information can be kept ready-at-hand. Thus one can mentally multiply 2-digit numbers; for 3- and 4-digit numbers, one uses an external aid in retaining the interim results. This is obviously important when contemplating the long series of steps in a complex argument, or the many facets of a complicated problem. Additionally, this is exacerbated when the mere maintenance of the internal representation itself requires effort. Kosslyn (1980, 1983), for example, has argued that maintaining a mental image requires constant effort, as parts of the image "fade" and must be refreshed. This potentially sets limits on both the maximum complexity of mental images and the kinds of effortful operations which can be applied to an image.

These considerations of memory load and of effort are no doubt correct, but there is no reason to believe they exhaust the limitations of internal

representations. The following sections describe three other, more interesting, ways that external representations can enhance cognitive performance. To anticipate the argument, I consider some of the ways in which external representations are functionally different from internal ones, by virtue of being open to different operations, and by tapping different skills. Said differently, we will consider the functional limitations of internal representations, and so, by contrast, the advantages of external representations. While our stress will be on the consequent benefits of externalizing one's thoughts, this perspective also has many other implications, and I discuss some of these in the final section.

Perceptual knowledge

There is a growing body of evidence that the knowledge to which one has conscious access may be interestingly different from the knowledge used in perceiving or identifying objects in the world. To start with a trivial case, a baseball outfielder's perceptual anticipation of a flyball's trajectory may (implicitly) require some speedy calculus; the player's explicit, conscious grasp of calculus may be far less sophisticated. Figure 1 illustrates the converse case.

AS SMART AS HE WAS, ALBERT EINSTEIN COULD NOT FIGURE OUT HOW TO HANDLE THOSE TRICKY BOUNCES AT THIRD BASE.

This important distinction between implicitly available and explicitly available knowledge has recently surfaced in discussions of "intuitive physics" (e.g. McCloskey et al., 1980). College students' explicit judgments about inertia and momentum and the like seem systematically incorrect. Even after a year of training in physics, subjects seem to hold an "impetus theory" similar to the one rejected by modern science over 400 years ago. Remarkably, subjects make predictions which are clearly contrary to their two decades of interaction with the world, predicting (for example) anomalous trajectories different from any they have seen in their lifetimes.

Subjects fare far better, though, if we request an implicit judgment rather than an explicit one: Rather than asking subjects to predict a motion, Proffitt and Kaiser (1986) showed subjects videotapes of actual motions and (simulated) anomalous motions, and asked subjects which looked correct. In this mode of test, subjects' intuitive grasp of physics seems quite good, as they reliably reject anomalous motions in this perceptual task. Apparently, then, subjects have perceptual knowledge about motions which exceeds their conscious knowledge: they cannot answer simple problems, but they do know what "looks right."

In a different vein, Jacoby and others have repeatedly shown that conscious memories can easily be dissociated from perceptual learning. For example, Jacoby and Dallas (1981) had subjects study a word-list. After a 24-hour delay, subjects' conscious recognition of these words was at near-chance levels. However, in a tachistoscopic-recognition paradigm, subjects were twice as likely to identify the previously viewed words (compared to control words). More, different variables seem to influence conscious recognition and perceptual identification, with depth of processing (for example) being critical for the former but irrelevant to the latter.

In describing these data, Jacoby argues that prior exposure leads to "perceptual fluency." The fluency by itself does not ensure that the item will be "recognized" as familiar. The recognition will occur only if the perceiver attributes the fluency to the relevant prior encounter. In the absence of this attribution, the item will not be acknowledged as (say) being from the previously viewed list, but, even in this case, the perceptual fluency still has

demonstrable effects: the item will be recognized more readily (Jacoby and Dallas, 1981), may be judged as more attractive in appearance (Seamon, 1981), may be judged as being more "prominent" (e.g. in a judgment of relative "fame" - Jacoby, 1985) and so on. Thus we are influenced by perceptual fluency -- things "look right" or "look familiar" -- even when we cannot recall what is right or attribute the familiarity. It remains an open question whether this perceptual knowledge simply provides a more sensitive assessment of what we know, or whether, more strongly, it is a different species of knowledge. One way or the other, though, these identification processes seem to be privileged with regard to various procedures, and it is therefore to one's advantage to be able to tap this knowledge.

The proposal, then, is that one benefit of external representations is that they give us access to knowledge and skills which are otherwise unavailable to us. Subjects in the "intuitive physics" experiments could, at least potentially, improve their performance by actualizing the motions they are attempting to judge; by creating the requisite input for accessing perceptual knowledge, that knowledge could be employed. Subjects in Jacoby and Dallas' memory experiments could likewise employ a species of the "generate and test" strategy -- generating candidate memory items, and then responding on the basis of the fluency with which these are processed. If these cases seem far-fetched, consider a more familiar case: In trying to spell a rarely encountered word, one writes the two possible spellings out on a piece of paper, then judges which "looks right." The suggestion is that this is a sensible strategy, tapping the same skills in perceptual identification which are evident in Jacoby's procedures. It is interesting that one seems unable to derive the same benefit by merely thinking about or imaging the alternate spellings; in these latter cases, one has no access to the perceptual skill. One gains access by externalizing the representation, by creating the relevant input for the perceptual process.

This kind of claim is entirely consistent with (although does not rest on) current claims about the "modularity" of cognitive functioning. The thesis of modularity is that separate and independent processes are employed by various input and output modules; these are inaccessible to and impenetrable by the central processes. Perceptual processes are one clear candidate for "modular" status; the impenetrability, for example, is evident in the fact that illusions

remain effective even when one knows one is looking at an illusion. The inaccessibility of these processes is evident in the "unconscious," automatic status of much perceptual "inference." The notion here, however, is that one can in a sense defeat the modularity by externalizing one's thoughts. To put this concretely, the perception module "knows" the correct spelling of a word, but one cannot directly access this knowledge. However, one can offer the perception module a candidate spelling and so to speak learn how the module reacts. In this way, by creating inputs for the module, and by monitoring the module's output, one indirectly gains access to the specific knowledge or skills inhering in the module.

The non-intensionality of external representations

It is extremely important that external representations are both representations and also things in themselves -- ink marks on paper, acoustic signals in the air. Thus the representations exist after we no longer are thinking about them, and they are open to interpretation by others, even if the others do not know the intent with which the representation was created. Indeed, even the creator of the representation has the option of setting his or her understanding of it aside, returning so to speak to the "raw material," and interpreting it anew. This opens the possibility that a different interpretation may be reached, rendering the external representation at least potentially ambiguous.

This potential for ambiguity is not present for internal representations, providing an important functional contrast between these and external representations. This point can be illustrated with an example drawn from our work with mental images (Chambers and Reisberg, 1985; Reisberg and Chambers, 1987). Many have argued that mental images are picture-like depictions, inherently ambiguous, and in need of interpretation (presumably via processes related to perception). In contrast, our claim has been that mental images, as mental representations, are inherently meaningful, representing some particular thing or state of affairs. Mental images are thus intensional entities, embedded in a particular context of understanding, and so necessarily unambiguous.

To test this claim, Chambers and Reisberg led subjects to encode some of the classical "ambiguous figures" (e.g. the duck/rabbit, the Necker cube, etc.).

Subjects were then asked to form mental images of these figures, and to examine the image for an alternative construal of the form. Subjects had previously received practice and instruction with other ambiguous figures, but, critically, were naïve to the test figures. In addition, subjects were not given enough time at encoding to find both construals of the test figures. In these ways, any reconstruals of the image will be bonafide "discoveries," indicating that images can support reinterpretation and so are ambiguous.

Across a variety of procedures, exactly zero subjects succeeded in reconstruing their images. An important control, though, makes clear the contrast between internal and external representations. After subjects had tried (and failed) to reconstrue their image, they were asked to draw a picture from the image, and then to attempt reconstrual of their own drawing. In sharp contrast to the imagery data, 100% of the subjects were able to reinterpret their own drawings. (These drawings were in fact ambiguous, as a new group of subjects was also able to find both construals in them.) Once there was a stimulus on the scene (even a self-created one), subjects could set aside the understanding they had in mind in creating the stimulus, and interpret it anew. In imagery, the understanding is inherent in the representation, so that there simply is no representation separate from the understanding. With no freestanding icon to interpret, no reinterpretation is possible. Consistent with this, our broad pattern of imagery data repeatedly show clear boundaries on what subjects can learn from or discover in their mental images. Discoveries easily occur when these are consistent with the way an image is understood. Discoveries which require a change in the image's understanding seem rigidly impossible. When subjects draw pictures of their images, though, there is a stimulus to be interpreted, and reinterpretations are routinely possible.

While the empirical content of this work has focused on imagery, I believe that the lesson is a general one. As long as ideas are internally represented, they exist only via a certain context of understanding, so there can be neither doubt nor ambiguity about what is intended. If I believe I am thinking about (say) crocodiles, then I am. It may turn out that I know very little about crocodiles, or I may later on discover that my knowledge about crocodiles is instead true of alligators. But, at the moment of having the thought, if I understand it as being about crocodiles, then it is about crocodiles, because it

is only through my understanding of it that the thought has any definition at all. This is obviously of functional importance, since it entails (among other things) that "lapses of communication" cannot happen within thought. At the same time, though, it may place limits on thought, especially if one is looking for a novel interpretation of a prior conception.

I believe that this is what is at stake when one tries to evaluate one's own ideas by "distancing" or "detaching" oneself from them. One needs to learn how well the idea fares outside of the context in which it was created, away from the implicit assumptions which may have accompanied it, divorced from the specific construals or emphases one initially had in mind. One likewise needs to discover if understandings which accompanied the idea were essential corollaries, or gratuitous associations. To achieve all this, one needs to preserve what is central in the idea while changing as much as possible about the mental context. One important way to achieve this is by externalizing the idea, so that there is something to be interpreted or reinterpreted once one's perspective has changed.

I do not wish to imply that this reinterpretation will be easily achieved. It is often difficult to change one's initial understanding of an external representation. In the case of ambiguous geometric figures, for example, reinterpretation requires foreknowledge that the figure is in fact ambiguous (Girgus et al., 1977); reinterpretations are delayed if one is prevented from concentrating on the figure (Reisberg, 1983; Reisberg and O'Shaughnessy, 1984). Reisberg and O'Shaughnessy argue that these difficulties lie, not in finding a new construal, but in ousting the old, in subverting the processes which maintain the stimulus's interpretation. Be that as it may, interpretation is possible: stimuli can be ambiguous even if mental representations cannot. Hence, one can learn by turning one's representations into stimuli -- i.e. by externalizing them.

Discovering omissions

Our final distinction between internal and external representations is the one, I believe, with broadest applicability. Internal representations often strike us as being rich and detailed, but the detail is uneven: Aspects of the representation which are attended will be elaborated; aspects which are not attended will be undetailed and vague. It is unclear whether the attention

creates the detail or vice versa, but, in either case, this co-occurrence often causes us to be surprised when the areas of vagueness are discovered. As long as they are unattended, the omissions from the representation can remain unnoticed.

Once again, our principal example comes from imagery. Even the most vivid of images omits myriad details, often details that seem central. There has been considerable discussion in the philosophy literature of the fact that the imaged tiger may have an indeterminate number of stripes (e.g. Dennett, 1981; Fodor, 1981). One can, if one chooses, image three or eight or however many stripes. If, however, one simply images a tiger, the number of stripes may be unspecified, and this indeterminacy can easily be overlooked.

This omission from the image can be discovered in various ways. As a simple extreme, one can ask the imager to count the tiger's stripes; the imager will typically register surprise at his/her inability to do this. As a more sophisticated variation of this, one can challenge subjects to form an image of a word (e.g. "pumpkin") and then to read off the letters of this word backward (n-i-k-p...). (Cf. Morris and Hampson, 1983, pp. 221-223.) To their surprise, subjects find this reasonably difficult, even when the imaged word was subjectively quite clear and vivid. Apparently, then, the image is appreciably less clear than the corresponding picture (from which the letters can easily be read), and, importantly, appreciably less complete than the subjects themselves had believed.

A final example brings us back to our theme: It is a common experience that even if one has a very clear image in mind, it is nonetheless difficult to "copy" the image onto a drawing. In part, it seems difficult simply to reproduce the contours "seen clearly" by the mind's eye. More important, one discovers in the process of drawing that some aspects were simply absent from the image. For example, one first images a horse, then attempts to draw this image. For most of us, it is very hard to draw the horse's knees; to use John Kennedy's description, the pictured horse often ends up looking as if it were wearing pajamas, and the picture looks clearly anomalous. The initial image, though, did not look at all anomalous -- one may have omitted the knees from the image (if, for example, one has no idea about how these look), but, as long as attention is elsewhere, the omission is consistently overlooked.

There are probably several reasons why externalizing the image leads to the discovery of omissions. After all, one could in principle apply the same selective filter to the drawing as to the image (attend only to the nose of the pictured horse, not its knees), and so discover nothing new. One reason this does not occur is the production process itself: At some point, one's pen must near the horse's knees, and then the contour must be dealt with, not ignored. At this point, one will immediately realize the absence of any clear idea about just how the horse's knees should look. In addition, it seems possible that the control processes for attention to stimuli may not be the same as those for mental representations. In the former case, one can easily allocate attention spatially: I will attend to the drawing's lower-right corner, or to the global contour, etc. This may be more difficult or perhaps just less likely with internal representations. In this case, attention seems governed by something like a content-addressing system, so that one focuses more easily on specified targets or aspects, rather than on whatever-is-in-region-X. This certainly would fit with our claim above that images are intensional entities, and not depictions to be effortlessly reparsed into arbitrary spatial regions.

As before, while our illustrations for this point have been drawn from imagery, I believe the phenomenon to be a general one. In thinking through a verbal argument, for example, one attends to what is included, and not to the gaps; thus the gaps are overlooked. It is only in explicitly and externally spelling out the argument that the gaps are discovered and the vaguenesses unveiled.

Conclusions and implications

Our prior data have pointed to clear differences between images and pictures; we have discussed some of the consequent advantages of externalizing images. The attempt to generalize beyond these data, at least for now, obviously rests more on conjecture than on hard evidence. One point of this paper, then, has been to argue that this evidence is worth gathering. In addition, this research is obviously relevant to the psychology of instruction and the psychology of discovery. Given current interest in these topics (e.g. Glaser and Takanishi, 1986; Gruber, 1974; Tweney et al., 1981), this by itself motivates this kind of inquiry. Finally, this attempt to contrast internal and external representations will inevitably sharpen our understanding of each,

again an aim of some importance within cognitive science.

For the moment, therefore, the worth of this perspective rises or falls on the richness of its set of implications, and not on the strength of its empirical base. In addition to the implications we have mentioned, consider the fact that much work in artificial intelligence has sought to segregate the problems of intelligence and machine vision. This is obviously consistent with current claims about modularity, but, if I am right, it is in some ways an unwise strategy. To take just one salient case, we know that Darwin sketched a great deal in his notebooks (cf. Gruber, 1974); our suggestion is that these sketches may have been causes of his insights as much as expressions of them. Current "unseeing" models of discovery, therefore, (e.g. Bradshaw et al., 1983) might for this reason exemplify the wrong way of conceiving this achievement.

Or, to indulge in one last conjecture, we know that humans have created decorative arts since the beginnings of our evolution as a species. Many have speculated about the cultural or psychological causes which may underlie this urge to decorate (e.g. Arnheim, 1971; Festinger, 1984; Gombrich, 1960). Our perspective suggests at least one possible factor: Humans may have discovered early on the advantages of drawing rather than imaging, of thinking out loud, so to speak, rather than thinking to oneself. These advantages of externalizing for various forms of problem-solving may be one element which encouraged and maintained the development of decorative art.

Much of this paper has aimed at motivating questions about external representations as aids to thought. It seems important to close, therefore, with one of the critical and early questions that has to be asked, namely, what <u>is</u> an external representation? By this I do not mean to ask what is a representation at all. Instead, our concern lies with what it is that distinguishes internal from external representations. I have employed straightforward examples here (e.g. mental images vs. pictures) but there are many less clear cases. An example will illustrate this point: Reisberg et al. (1987) asked subjects either to listen to or (in other conditions) to image an ambiguous auditory stimulus. That is, the stimulus (rapid repetitions of the word "stress") could be parsed as either repetitions of "stress" or of "dress" or of "rest" or "tress." When subjects <u>heard</u> this stimulus, they easily discovered the various construals. When subjects <u>imaged</u> the soundstream, the results depended on whether additional steps were taken to block subvocalization.

Without subvocalization (e.g. if subjects imaged while chewing a large piece of candy), the results follow the pattern of our data with visual imagery: subjects reliably fail to reconstrue the auditory image. With subvocalization (e.g. no candy), though, subjects do discover the alternative interpretations of the image. Reisberg et al. argue that this is because subvocalization creates a stimulus, in this case an articulatory event. This articulatory stimulus, as an "external" representation, is then subject to interpretation and will support reinterpretation. Note, therefore, the possibility of covert external representations. In fact, while I have used the term "external representation" throughout this paper, the relevant, non-intensional, stimulus event need not be literally external. I am nonetheless inclined to retain the "internal"/"external" terminology (for expository ease), but now the problem of definition is in plain view. In addition, in this paper I have suggested several empirical properties of external representation, with regard to perception, or to ambiguity, or to controlling attention, but these properties may not be co-extensive. Given all this, I am reluctant to stipulate a definition. Instead, these properties of representations need to be assessed empirically. This is exactly what we have tried to do in our work with imagery; a major purpose of this paper has been to recommend a broadening of this effort.

References

Arnheim, R. (1971) Art and visual perception. Berkeley: University of California Press.

Bradshaw, G, Langley, P, and Simon, H. (1983) Studying scientific discovery by computer simulation. Science, 222, 971-975.

Chambers, D. and Reisberg, D. (1985) Can mental images be ambiguous? Journal of Experimental Psychology: Human Perception and Performance. 11, 317-328.

Dennett, D. (1981) The nature of images and the introspective trap. In N. Block (Ed.), Imagery (pp. 51-61). Cambridge, MA: MIT Press.

Festinger, L. (1984) Reflections on Human Nature. New York: Columbia University Press.

Fodor, J. (1981) Imagistic representation. In N. Block (Ed.), Imagery (pp. 63-86). Cambridge, MA: MIT Press.

Girgus, J., Rock, I. and Egatz, R. (1977) The effect of knowledge of reversibility on the reversibility of ambiguous figures. Perception and Psychophysics, 22, 550-556.

Glaser, R. and Takanishi, R. (Eds.) (1986) Psychological science and education. American Psychologist, 41, 1025-1168.

Gombrich, E. (1960) Art and illusion. New York: Pantheon.

Gruber, H. (1974) Darwin on Man. London: Wildwood House.

Jacoby, L. (1985) The relationship between learning and recollection: Memory attributes vs. Memory attributions. Paper presented at the Annual Meeting of the Psychonomic Society, Boston, Mass.

Jacoby, L. and Dallas, M. (1981) On the relationship between autobiographical memory and perceptual learning. Journal of Experimental Psychology: General, 3, 306-340.

Kosslyn, S. (1980) Image and mind. Cambridge, MA: Harvard University Press.

Kosslyn, S. (1983) Ghosts in the mind's machine: Creating and using images in the brain. N.Y.: Norton.

McLoskey, M., Caramazza, A., and Green, B. (1980) Curvilinear motion in the absence of external forces: Naïve beliefs about the motion of objects. Science, 210, 1139-1141.

Morris, P. and Hampson, P. (1983) Imagery and consciousness. New York: Academic Press.

Proffitt, D. and Kaiser, M. (1986) Why you cannot see what holds a gyroscope up. Paper presented at the Annual Meetings of the Psychonomic Society, New Orleans.

Reisberg, D. (1983) General mental resources and perceptual judgments. Journal of Experimental Psychology: Human Perception and Performance, 9, 966-979.

Reisberg, D. and Chambers, D. (1987) Neither pictures nor propositions: What can we learn from a mental image? Manuscript under review.

Reisberg, D. and O'Shaughnessy, M. (1984) Diverting subjects' attention slows figural reversals. Perception, 13, 461-468.

Reisberg, D., Smith, J.D. and Sonenshine, M. (1987) "Enacted" auditory images are ambiguous; "pure" auditory images are not. Manuscript under review.

Tweney, R., Doherty, M. and Mynatt, C. (1981) On scientific thinking. N.Y.: Columbia University.