Perceptive Actions in Tetris

David Kirsh — dkirsh@ucsd.edu
Paul Maglio — pmaglio@ucsd.edu

Cognitive Science Department
University of California, San Diego

Using Actions to Help Solve Perceptual Problems

Cognitive organisms have three rather different techniques for intelligently regulating their intake of environmental information. In order of the time needed to uncover information they are:

1. control of attention: *within* an image produced by a given sensor certain elements can be selected for additional processing;
2. control of gaze: the orientation and resolution (center of foveation) of the sensor can be regulated to create a *new* image;
3. control of activity: certain *non-perceptual* actions can be performed to increase the probability of unearthing salient information that currently is unavailable, hard to detect, or hard to compute.

In this note we shall discuss some experiments we have been performing on this last variety of active sensing.

First a word of justification. Motor actions which do not directly control sensor orientation are not usually deemed part of the perceptual process. When a child shakes, bites and throws an object, he or she is performing actions that can serve other functions; they need not be connected to perception. Hence they tend to be classified as shaking, biting or throwing: terms that are logically distinct from perception. But often actions are so well coordinated with the perceptual process that they interpenetrate perception, making it impossible to properly understand the perceptual *system* if they are ignored.

J.J. Gibson was the first psychologist to emphasize this mutuality [Gibson1950, Gibson1966]. His central concern was how the constant activity of an organism created a dynamic stream of information that was richer in action-relevant data than static images. By looking for regularities between actions and the first or second derivatives of the changing optical input created by those actions, he noticed that it was possible to build action control systems that could regulate behaviour without consulting static images of the world. The classical example is the way measurement of looming can be harnessed to control the wing span of gannets, the avoidance of obstacles and so on, see also [Lee and Reddish1981].

Gibson made efforts to extend his analysis to more complex behaviour by introducing the concept of affordance. But the concept has proven too vague and ambiguous to deliver clear hypotheses for testing. The data which we have collected of 30 subjects playing Tetris provides a rigourous computational laboratory from which to study some of Gibson's ideas. It also allows us to study other ideas about action and perception unavailable to Gibson because of his aversion to computation and information processing. One of these is that certain forms of action can substitute for the kind of mental operations that occur in heuristic problem solving: particularly heuristically controlled generate and test methods. In what follow we shall give a brief introduction to some of our data and describe how these might make contact with the notions of affordance and heuristic search.

Tetris as a Domain for Studying Perceptual Skills

Tetris[1] is an interactive video game in which players must choose from three actions: rotate, translate or drop. Tetrazoids enter from the upper boundary of a rectangular playing field at a fixed speed which increases as the game proceeds, leaving the player with less and less time to decide in which column and orientation to place a zoid. We have implemented a computational laboratory that lets us record, at the millisecond level, keystrokes and game situations, as well as allowing us to dynamically create situations.

Three rather curious phenomena we have noted are that players at the intermediate and expert level will occasionally

1. rotate zoids in midstream, seemingly because of confusion about which zoid—the current zoid or its mirror zoid—will fit snugly in an intended location,

[1] *Tetris* is a trademark of AcademySoft/ELORG. We have implemented our own experimental version of this game, *Xtetris*, using the X Window System.

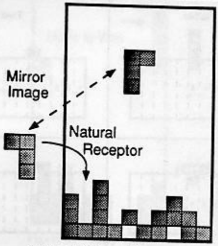

Figure 1: Mirror Receptor

Here we see a rotation in midstream. Along the active contour there is a natural receptor—an ideal fit—for either a left handed L zoid or its mirror image, the right handed L. Because mental rotation both takes time and is subject to error, players tend to adopt the strategy of rotating the piece to avoid incorrectly placing a right handed L in a receptor made for a left handed L.

2. rotate certain zoids as soon as they emerge as if trying to disambiguate the zoid from all others as soon as possible,
3. drop certain zoids after frst perfroming the unneccessary action of translating them to the nearest outer wall and then back again as if to verify the column of placement.[2]

All these phenomena seem to be cases where a particular perceptual-cognitive problem is more quickly or easily solved in the world—or, at least, with the help of external activity—than in the head alone. We regularly use external aids to help us solve purely cognitive problems, such as adding, accounting, composing and so on. Presumably we also perform actions to help us solve perceptual problems.

Consider 1. **Rotate zoid in midstream.** See figure 1 for a statement of the phenomenon. Although we do not yet know exactly how a player selects where to place a piece we can be sure that some matching of piece shape to potential placement location occurs. In the language of generate and test we can say the player at some point must test the fit of the piece against a candidate placement. It is the job of the perceptual system to facilitate this test.

Several strategies for testing fit are possible. One obvious method is to have the player rotate the piece in his or her head and use the resulting masks against the candidate placements. The costs are the time required for mental rotation and the probability of error in rotation. The advantages are that the player may keep his or her gaze directly on the contour while creating the mental mask from the internally rotated image.

As noted, though, the data suggest that players often rotate certain pieces externally. The natural hypothesis is that external rotation replaces mental rotation: that it is cheaper or better to perform the rotation in the world than in the head. But this is not the only conclusion. It is possible that players in all cases mentally rotate pieces, but in instances where a mistake in rotation is easily made, as is the case where pieces have mirror images, the player *verifies* that his or her rotation has been correct by performing the rotation in the world.[3] That is, external rotation supplements mental rotation; it does not replace it.

Yet another explanation for external rotation is that the piece is rotated in order to verify its identity. External rotations, in this case, serve as *perceptual indices* for retrieving the piece type. Once the piece has been correctly identified, one may have access to its shape under all rotations, since it may be stored in this multiple perspective form [Tarr and Pinker1989]. Thus just knowing that a given piece is a left-handed L might suffice to know that it would fit a given candidate placement.

[2] In some implementations, the column of the active zoid is indicated at the bottom by a marker that shows where its base would intersect the lower boundary. So far we have collected data with this feature turned off.

[3] In such cases gaze should be only at the rotating piece, not at the active contour or placement. Gaze may shift to the placement location once the piece has been rotated to the orientation the player thinks is right. But this would not be necessary.

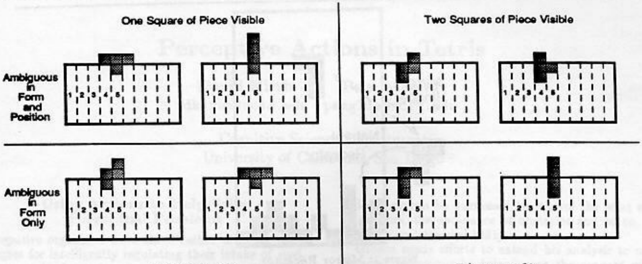

Figure 2a

Figure 2b

In figure 2 we see zoids as they first enter the playing field, in 2a they are one square in, in 2b they are two squares in. The upper portion of both 2a and 2b show zoids that look identical at this stage, both in position and in form. The bottom portions show zoids that look identical in form alone. Careful examination reveals that they are in different columns. Players are not explicitly aware of this column difference. The data show that players do not come out rotating, as we originally thought, but rather have a great burst once they are two rows out. At this point they show considerable sensitivity to column difference. Players have a much greater tendency to rotate zoids ambiguous in both form and position (such as those seen in the upper portion of 2b) than they have of rotating zoids that are ambiguous in form alone. By rotating ambiguous zoids early players are able to make faster identifications thereby either setting up the conditions for testing candidate placements early or setting early constraints on a candidate generator.

This method agrees in a bizarre way with Gibson's theory of affordances. An affordance for Gibson is a property of the environment that facilitates a goal relevant action of the agent. It is an objective fact about the world that chairs afford sitting, that eggs afford breaking, and that right handed L's in particular Tetris games will fit perfectly in certain placements. The problem for the perceptual system is to find (become tuned to) certain agent-environment invariants that reliably correlate with these affordances. If rotating a zoid somehow primes the visual system to be in a state where a composite mask or set of microfeatures are detectable by sweeping the eye over the candidate placement, then the system has a way of detecting affordances on the contour. Thus on this view finding a placement is a matter of setting up the visual system by externally rotating the piece enough so that when the eye is swept across the contour it picks up all the relevant affordances.

Rotating Early. See figure 2 for a statement of the phenomenon. When a zoid first enters the playing field and only a fraction of its total form is visible, the player must rely on subtle clues to disambiguate it. Of course, players may not follow a strategy that requires them to disambiguate zoids as quickly as possible. But in fact we have noted that pieces that are ambiguous in form *and* position are more often rotated early than pieces ambiguous in form alone. The simplest explanation, once again, is that early rotation serves a fact finding purpose. By rotating a partially hidden piece a player un-occludes part of it, thereby flushing out new information. The faster this may be done, the sooner a piece may be unambiguously identified.

The virtue of early identification makes sense on both the generate and test and the affordance accounts. Early identification in a generate and test model may pay off by biasing the generator. As more information about the piece is available, the set of plausible candidate placements ought to be more strongly constrained, and therefore smaller. Fewer candidates make for faster testing.

Early identification on an affordance account also pays off because rotation primes the affordance detector, thereby driving the perception mechanism to see more perfect fits *earlier*. Thus, even if a player is tuned to register *all* the affordances of a contour regardless of current piece and orientation, the order in which affordances are noticed can be primed by actions: Notice what best works with what you have just seen. Thus, whereas the output of the generator was reduced

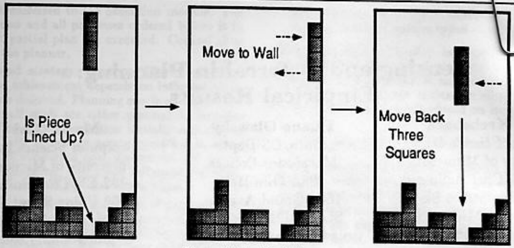

Figure 3: Verify Column

In figure 3 we see an instance of how the bar zoid is regularly translated to the outer right wall and back again before it is dropped. The explanation we prefer is that the subject confirms that the column of the zoid is correct, relative to his or her intended placement, by quickly moving the zoid to the wall and then with eyes on the contour simultaneously tapping and counting out the number of squares to the edge of the intended placement.

in number and its order of generation was unaffected, on the affordance account, the order in which candidate placements are recognized is effected by activity, and so the more activity, the more priming.

Zoid Translation. See figure 3 for a statement of the phenomenon. After finishing the generate and test phase, or alternatively, the choose affordance phase, the player may wish to confirm that he or she has succeeded in moving the piece to the intended column. This further phase is most useful when the piece is still high above the active contour and about to be dropped.

To accomplish this checking phase a player may use several strategies, some taking place entirely in the head, some taking place partly in the head and partly in the world. Here are three possibilities: 1. The player may activate a visual routine [Ullman1983] which compares the horizontal boundaries of the zoid with the horizontal boundaries of the chosen location. This might be done by looking in turn at each of the zoid's boundaries and visually *drawing a line downward.* 2. The player may *count* the number of squares from the nearest wall to the zoid and compare that with the number of squares from the wall (at the active contour) to the intended placement. The cost of this routine (as a function of time and probable error) we expect will vary with the vertical distance between the piece and the contour. Accordingly, for large drops, where the possibility of error and the time needed to guide the eye is significant, we might expect that the player would choose to physically shift the zoid as quickly as possible to the wall and then drop his or her gaze to the contour and count the squares to the intended placement by tapping a keystroke for each square counted (ie, count with their fingers). 3. A third possibility is to physically shift the zoid to the wall and then to count out in keystrokes a number already known. A player may have the numerical distance from the wall to the intended placement already in mind because it is easy to perceive distance from the wall to the intended placement using the already placed pieces for reference.

To date we have insufficient psychological information on eye movement to choose between two and three, and insufficient knowledge of the time required to perform the possible visual routines to adequately justify our conjecture that translating in the world is faster than translating in the head.

References

[Gibson1950] J. J. Gibson. *The perception of the visual world.* Houghton Mifflin, 1950.

[Gibson1966] J. J. Gibson. *The senses considered as perceptual systems.* Houghton Mifflin, 1966.

[Lee and Reddish1981] D. N. Lee and P. E. Reddish. Plummeting gannets: a paradigm of ecological optics. 293:293–294, 1981.

[Tarr and Pinker1989] M. Tarr and S. Pinker. Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, 21:233–282, 1989.

[Ullman1983] S. Ullman. Visual routines. Technical Report Memo 723, MIT AI Lab, 1983.