The Effect of Perceived Difficulty on Perceptual Learning

Kodi Arfer
Allegheny College

Mentor: Paul Schrater
Department of Psychology

Abstract

When trying to account for performance on perceptual tasks such as motion extrapolation, researchers typically postulate specific properties of the perceptual system. The present study investigated whether a counterintuitive finding in a previous motion-extrapolation experiment could be explained solely by subjects' differing beliefs about task difficulty. Nine participants in summer programs at the University of Minnesota completed a computer-based task requiring them to learn the movement patterns of an indefinitely outlined object. As hypothesized, a one-sentence warning of increased difficulty measurably increased learning of object trajectories. But self-reports showed no evidence of increased effort, leaving ambiguous the means of increased learning. These results suggest that high-level attitudes and beliefs can play an influential role in low-level tasks.

When people are asked to guess how an object will move, whether working from an animation, a drawing, or a verbal description of a situation, their responses are many and varied. Sometimes their predictions are blatantly at odds with Newtonian mechanics (Cooke & Breedin, 1994), and sometimes they make theoretically optimal use of limited, competing information (Fulvio, Maloney, & Schrater, 2010). Average accuracy varies widely between tasks, and individual differences can be great (Cooke & Breedin, 1994). The most common approach to making sense of all the data is to postulate mental phenomena specific to the perceptual tasks of interest and then test for their existence. For example, a persistent idea in motion extrapolation is that people make predictions using some implicit theory of mechanics that, however unrealistic and unconscious, is coherent and applied consistently; one such theory is called impetus theory (Kozhevnikov & Hegarty, 2001). A different kind of example is given by the "dynamic mental representations" of Freyd (1987), internal simulations that run in real time and model properties such as momentum and centripetal force (Hubbard, 1996). Yet another approach is to find statistical models that match behavior such that any surprising aspects of the behavior can be explained as properties of the model (Hudson, Maloney, & Landy, 2007).

Purpose-built explanations such as these can of course have great explanatory power, but they come with a price. Their power comes from their specificity, and specificity runs counter to parsimony. If we tried to explain all behavior of interest with specialized hypotheses, we would be left with a gigantic, unenlightening patchwork quilt of theories. Instead, we should as much as possible appeal to more general ideas, ideas that apply to a wide range of human behavior. Indeed, if any such ideas are correct, they probably have ramifications for whatever problems we're interested in and thus may interfere with any specialized explanations. Cooke and Breedin (1994) demonstrated some ramifications of this kind when they found that people's motion predictions were sensitive to contextual cues in the vein of Tversky and Kahneman's (1981) "framing effects" (which have implications for nearly all judgments) rather than remaining consistent with impetus theory.

Can we in fact obviate the need for a specialized explanation by applying a general idea? Consider the case of Schrater and Powell (2010). In this experiment, subjects watched a cloud of rapidly moving dots glide across a computer screen until it went behind an occluder, then indicated where they thought it would emerge from the occluder (see Figure 1). The trajectory of the cloud varied from trial to trial, but it was always a sinusoid whose parameters adhered to certain limits. This consistency, along with the feedback that subjects sometimes received on how accurate their estimates were, permitted subjects to gradually learn about how the cloud moved over the course of hundreds of trials. Thus, subjects' estimates grew more accurate over time. A much more surprising result was that the noisier the cloud of dots was—that is, the wider the individual dots were spread out and the faster they moved within the cloud—the better subjects learned the trajectories. A noisier cloud provided less precise trajectory information, and thus made the task harder; why would subjects perform better on a

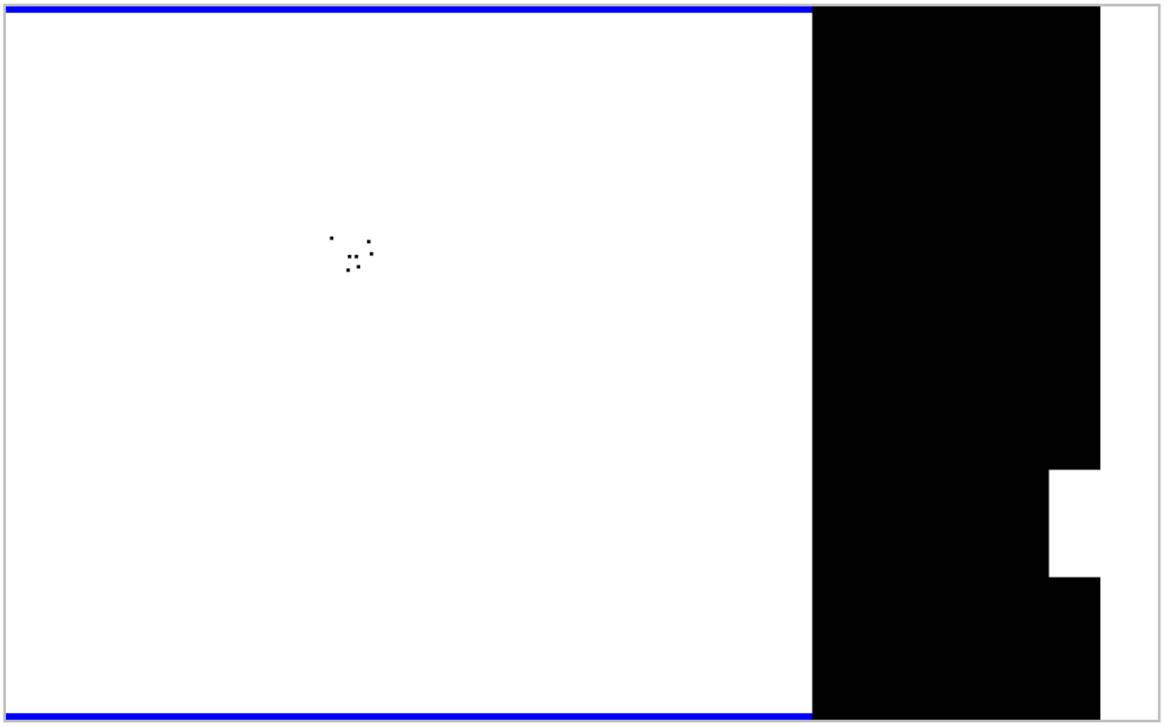

Figure 1. A screenshot of the program used in the present study, which appears much as in Schrater and Powell (2010).

harder task?

Schrater and Powell hypothesize that this difference reflects the adaptive nature of the perceptual system. Subjects had two sources of information they could use to predict trajectories: measurements, or what they saw of the cloud's position at any given moment, and a model, or the idea of the cloud's movement patterns that they'd built up over many trials. According to Schrater and Powell, the perceptual system uses the observed reliability of each of these two input sources to adjust how it combines sources—which one it favors, and to what degree—when making predictions. So noisy measurements would lead participants to use their models more, leading to improved performance, since the exact positions of the dots give no trajectory information.

An alternative explanation is that the difference in learning arose merely from a difference in perceived difficulty. That is, the noisier the cloud of dots, the harder subjects judged the task to be, and the harder the subjects thought the task was, the more they learned. Previous research has established that perceived difficulty is a well-behaved variable, insofar as it's doubly dissociable from perceived degree of control over a difficult situation, and people will rate a task designed to sound hard as more difficult than a task designed to sound easy (Trafimow, Sheeran, Conner, & Finlay, 2002). Brehm, Wright, Solomon, Silka, and Greenberg (1983) found that difficult goals are more attractive than easy goals and thus led to greater mobilization of energy; one can easily imagine that such motivation would increase learning. To directly test whether differences in perceived difficulty could explain Schrater and Powell's (2010) results, I ran a variation of their experiment in which an instructional manipulation intended to affect perceived difficulty varied independently of the task proper. The manipulation, modeled after Moè (2009), was to tell some subjects that the second block of the experiment was harder than the first. I intended thus to send a message that was neither ambiguous (it gave subjects a clear basis for comparison, unlike a message to the effect of "This task is hard.") nor likely to arouse suspicion nor confounded with any other message. I also changed the task slightly between blocks to prevent subjects from doubting the warning on the grounds that the second task was clearly identical to the first.

Method

Participants

Participants were students employed by or enrolled in summer programs at the University of Minnesota, recruited with fliers or in person. Ages ranged from 18 to 33, and all but one participant were female. One subject's data was lost due to a bug in the program used for data collection, leaving $N = 9$ for the purposes of analysis. All subjects gave informed consent and were compensated with $10.

Procedure

Subjects read the following instructions on-screen:

> In this task, your goal is to catch as many of the moving dots as you can in a bucket. The bucket appears as a white rectangle on the right-hand side of the screen; you can move it up and down with the mouse. Beware that the black rectangle will block your view of the dots. You'll also see colored bars at the top and bottom edges of the screen. When the bars are blue, you'll receive feedback on your performance. When the bars are yellow, you'll receive no feedback.

Each subject completed two blocks of 145 trials, taking about 20 minutes in total. The basic procedure was largely the same as in Schrater and Powell (2010); in fact, the computer program used to administer the task was a modified version of Schrater and Powell's. Each trial, a cloud of seven dots emerged from the left edge of the screen and traveled at a constant horizontal speed until it reached the right edge. The cloud's pattern of vertical movement (its *trajectory*) varied depending on the trial. In one of the two 145-trial blocks, trajectories were sinusoids whose wavelength, amplitude, and phase were randomly selected from normal distributions; these were largely identical to the trajectories in Schrater and Powell (2010). In the other block, trajectories were parabolic waves of a constant wavelength equal to twice the width of the screen, with amplitudes and phases selected from uniform distributions. Whether subjects received sinusoidal trajectories in Block 1 and parabolic trajectories in Block 2 or the reverse was determined by random assignment. The positions of the individual dots were random offsets from the center of the cloud, and were determined for each frame of the animation independently of previous frames. So while the cloud as a whole followed a smooth trajectory, the dots appeared to jump around arbitrarily.

When the cloud of dots reached the midline of the occluder, the bucket became immobile, forcing subjects to predict where the dots would emerge in advance. What happened while the bucket was immobile depended on the type of trial. In *testing trials*, indicated by yellow bars at the top and bottom edges of the screen, the occluder suddenly extended rightwards to completely obscure the bucket and wherever the dots actually emerged. Subjects thus had no way of assessing the accuracy of their prediction. In *training trials*, indicated by blue bars, the bucket remained visible, the dots froze as they exited the occluder, and the subject received a brief score report such as:

> This trial, you caught 72% of the dots.
> Across all trials, you've caught 56% of the dots.

The second figure included only trials in the current block. Each block began with 20 testing trials and then used a pattern of 20 training trials followed by 5 testing trials.

One additional experimental manipulation was the instructions subjects received at the beginning of Block 2. All subjects read "Now for the second task, in which the dots will move in a different trajectory.", but for some this was immediately followed by "Note that this task is more difficult than the first."

Random assignment of both between-subjects conditions yielded the following cell sizes:

- 4 subjects saw sinusoidal trajectories first and received no warning of increased difficulty.
- 1 subject saw parabolic trajectories first and received no warning.
- 2 subjects saw sinusoidal trajectories first and received a warning.
- 2 subjects saw parabolic trajectories first and received a warning.

After completing both blocks, subjects answered questionnaires about the task and their experience with video games.

Results

For each trial, participants were assigned a raw score based on how accurately they'd predicted the cloud's trajectory. This was computed as the signed vertical distance from the final mouse position to the point where the center of the cloud emerged from the occluder, such that a positive score indicated that the cursor was above the trajectory. Thus, scores lower in absolute value represented better performance. In general, performance was slightly worse in sinusoidal blocks than parabolic blocks. Figure 2 displays some preliminary data on how trajectory parameters influenced raw scores. Notice that subjects performed better when the occluded portion of the trajectory curved

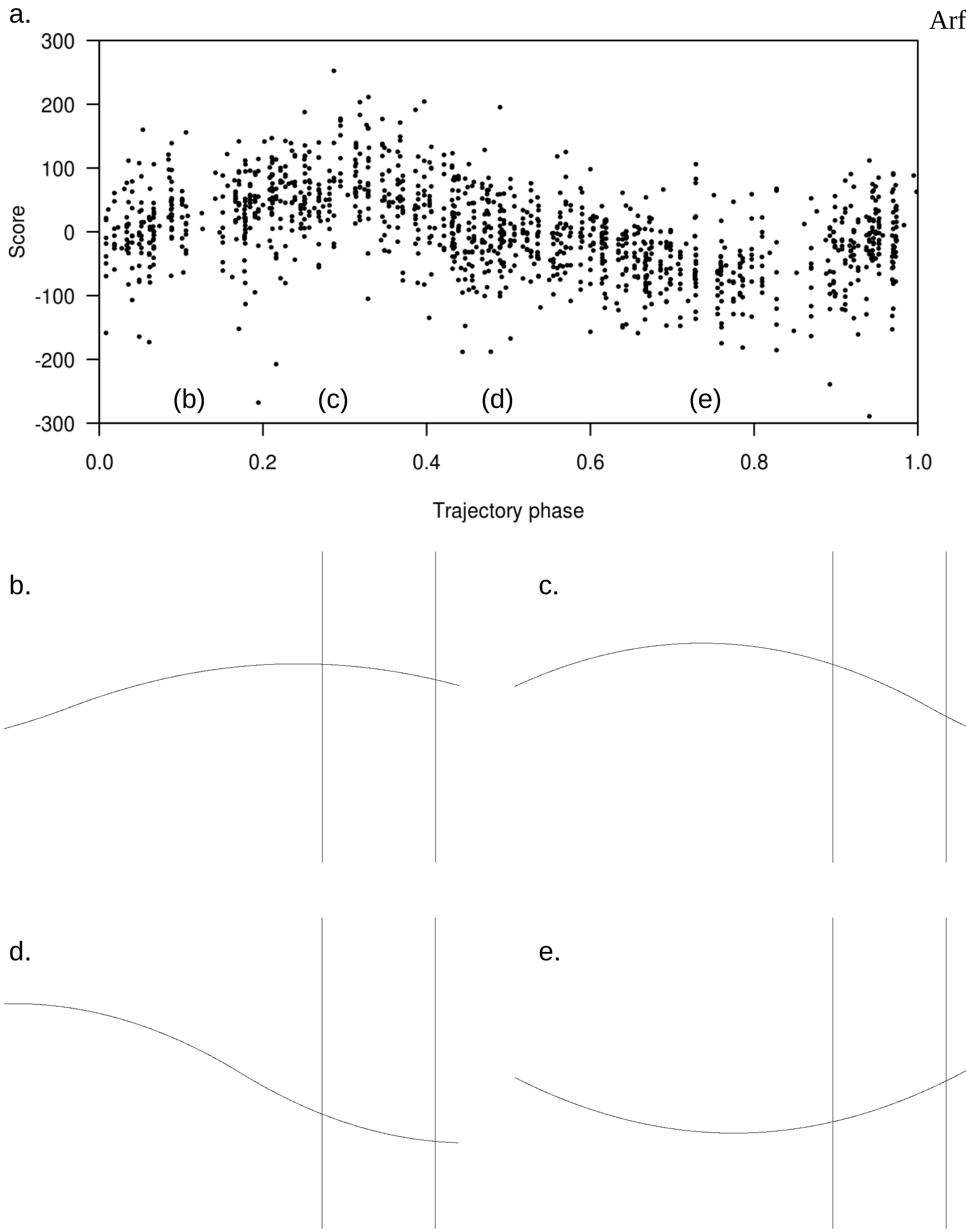

Figure 2. (a) Raw scores (see text) in parabolic trials as a function of trajectory phase, collapsed across all subjects. (b–d) Sample trajectories with correspondingly marked phases in (a). The vertical lines represent the edges of the occluder.

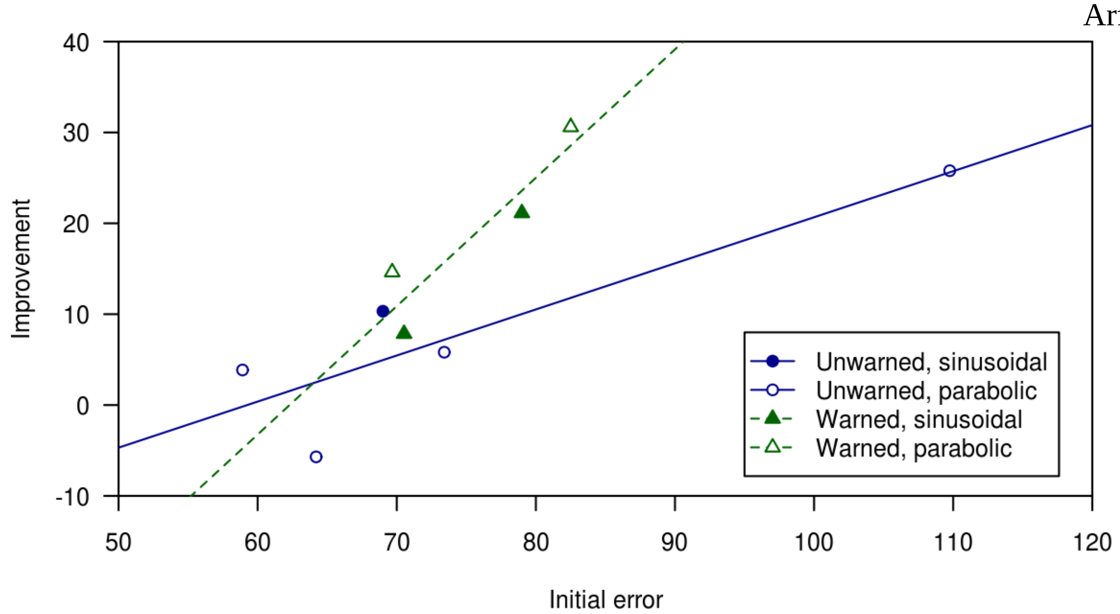

Figure 3. Trajectory learning in Block 2 as a function of initial error. The *x*-axis gives subjects' error scores for the first half of block 2, and the *y*-axis gives the difference between this score and error for the second half, such that positive *y*-values indicate a decrease in error. Note that removing the rightmost point changes the solid line only slightly, and in a way indicating a greater effect of warning—the slope decreases by 0.066.

gently than when it curved sharply; this is consistent with Schrater and Powell's (2010) finding that subjects had a strong bias towards extrapolating linearly.

Raw scores had no clear pattern across trials, oscillating rapidly between high and low. To create a less noisy measure of learning, for each 72-trial half of each block, I fit a least-squares line to raw scores as a function of trajectory amplitude. I defined the *error* for each half-block as the standard deviation of the residuals produced by this regression. Collapsing across all conditions, errors for first half-blocks were significantly greater than errors for second half-blocks, $t(17) = 2.14$, $p = .023$ (one-tailed), indicating that subjects grew better at predicting trajectories over the course of each block.

A comparison of Block 2 error scores between the warned and unwarned conditions can be seen in Figure 3. In both conditions, initial error was strongly correlated with improvement, $r = 0.89$ for the unwarned subjects and $r = 0.92$ for the warned. The slopes of both regression lines are positive, meaning that greater initial error predicted greater improvement, probably because the range of possible performances was finite and thus worse initial performance left more room for improvement. To determine how meaningful the difference in slopes is, I employed bootstrapping, a random-resampling

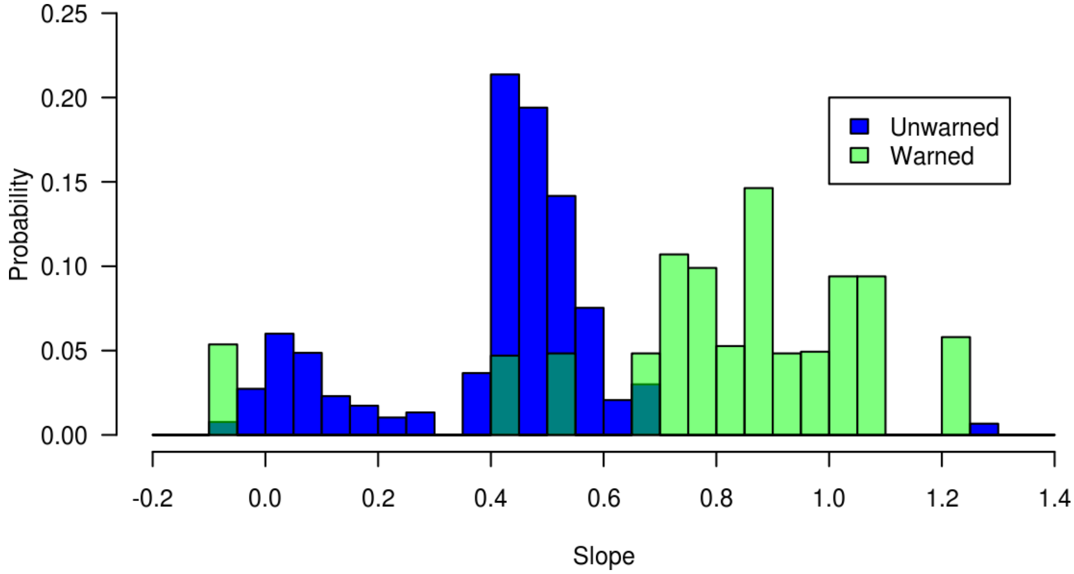

Figure 4. Superimposed histograms of bootstrapping results for the slopes of the two regression lines in Figure 3. To keep the scale of the *x*-axis small, slopes less than −1.5 or greater than +4 (which altogether accounted for less than 7% of slopes) have been omitted.

technique that creates a frequency distribution of a summary statistic from all possible weighings of the available data points. The distributions for the two slopes are shown in Figure 4. At most 10% of each distribution overlapped with the other; since the slope for the warned subjects is steeper, this supports the hypothesis that a warning of increased difficulty can enhance learning. The enhancement to learning seems to have taken the form of a boost to the previously mentioned effect of greater initial error leading to greater improvement.

On the questionnaire, subjects responded to four items, presented once for each block, on 7-point Likert scales. The first question was “How difficult did you find the task?”, with option 1 marked as “very easy” and option 7 as “very difficult”. The mean change in difficulty ratings from Block 1 to Block 2 was −0.20 for unwarned subjects and +1.00 for warned subjects, a difference which is unlikely given the null hypothesis but not to the point of significance, $t(7) = 1.15$, $p = .143$ (one-tailed). Thus, the warning may not have had the intended effect on subjects’ opinions of difficulty; alternatively, the problem may have been the combination of a weak effect size, a small sample size, and additional noise introduced into the data by collapsing across trajectory-order conditions. The second question was “How much effort did you put into the task?”, with the response scale anchored at “no effort” and “great effort”. The mean change in effort ratings between blocks was +0.40 for unwarned subjects and +0.50 for warned subjects, an insignificant difference, $t(7) = 0.12$, $p = .453$ (one-tailed). This gives no evidence that warned subjects felt they exerted extra effort.

Discussion

As expected, subjects grew better at predicting cloud trajectories with practice, and a warning of increased difficulty amplified this effect. These results suggest that the paradoxical improvement in performance found by Schrater and Powell (2010) may be attributable to perceptions of difficulty alone. Less clear are the means by which the warning had its effect.

First of all, these results give little indication of how the warning motivated subjects to learn more. Perhaps some subjects wanted to achieve a certain minimal level of performance from the beginning, and they interpreted the warning to imply that reaching this minimum would require better trajectory knowledge than in Block 1. Or perhaps they figured that the purpose of the warning was to help them adapt to a real increase in difficulty and wanted to cooperate. Conversely, some subjects may have realized that the warning was a manipulation independent of actual difficulty, and, displeased that I was trying to deceive them, sought to falsify the warning by performing well. An intermediate possibility is that subjects were curious as to whether the warning was accurate, leading them to pay more attention in Block 2 (in order to accurately compare it with Block 1) and thus learn more.

If the difference in learning arose from a difference in attention, how was it that attention changed? Did subjects process more information, process the same information more thoroughly, or process less irrelevant information? One way subjects might have learned more without paying more attention is by positioning the mouse more precisely. More precise control of the bucket would’ve permitted more precise interpretation of feedback (i.e., score reports). However, the role of feedback in improving performance may have been mitigated by the difference between what it reported (proportion of dots caught) and the measure of performance I used for analysis (“raw scores”; i.e., the distance between the bucket and the center of the cloud).

The results of the questionnaire raise questions of how conscious and deliberate improvements in learning were. If the warning didn’t affect ratings of effort, subjects either didn’t intend to exert more effort in response to the warning or didn’t feel they succeeded in doing so. Possibly their adaption to the warning was mostly unconscious and involuntary. This poses problems for my simple initial idea of increased perceived difficulty triggering a deliberate increase in effort which in turn was responsible for the increase in learning.

On the whole, however, this attempt to explain the results of a perceptual-learning experiment in terms of a high-level process was a surprising success, in spite of its small sample size. Investigators employing any kind of perceptual task may benefit from considering how the presentation of a task and subjects’ beliefs and attitudes about it could influence behavior. And in combination with Brehm et al. (1983), the present study suggests that teachers may be able to improve student learning not by genuinely challenging students so much as making them *expect* a challenge.

References

Brehm, J. W., Wright, R. A., Solomon, S., Silka, L., & Greenberg, J. (1983). Perceived difficulty, energization, and the magnitude of goal valence. *Journal of Experimental Social Psychology*, 19(1), 21–48.

Cooke, N. J., & Breedin, S. D. (1994). Constructing naive theories of motion on the fly. *Memory & Cognition*, 22(4), 474–493.

Freyd, J. J. (1987). Dynamic mental representations. *Psychological Review*, 94(4), 427–438.

Fulvio, J., Maloney, L., & Schrater, P. (2010). *Model-based extrapolation of circular motion: observers select extrapolation models matched to their visual capabilities.* Unpublished manuscript.

Hubbard, T. L. (1996). Representational momentum, centripetal force, and curvilinear impetus. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(4), 1049–1060.

Hudson, T. E., Maloney, L. T., & Landy, M. S. (2007). Movement planning with probabilistic target information. *Journal of Neurophysiology*, 98, 3034–3046.

Kozhevnikov, M., & Hegarty, M. (2001). Impetus beliefs as default heuristics: Dissociation between explicit and implicit knowledge about motion. *Psychonomic Bulletin & Review*, 8(3), 439–453.

Moè, A. (2009). Are males always better than females in mental rotation? Exploring a gender belief explanation. *Learning and Individual Differences*, 19(1), 21–27.

Schrater, P., & Powell, N. (2010, May). Learning models of motion extrapolation. In a synopsium conducted at the meeting of Vision Sciences Society, Naples, Florida.

Trafimow, D., Sheeran, P, Conner, M., & Finlay, K. A. (2002). Evidence that perceived behavioural control is a multidimensional construct: Perceived control and perceived difficulty. *British Journal of Social Psychology*, 41(1), 101–121.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211, 453–458.