ECOLOGICAL PSYCHOLOGY, *18*(2), 67–91

Mode Transition and Change in Variable Use in Perceptual Learning

Alen Hajnal and Michael Grocki
Center for the Ecological Study of Perception and Action
University of Connecticut

David M. Jacobs and Frank T.J.M. Zaal
Faculty of Human Movement Sciences
Vrije Universiteit, Amsterdam

Claire F. Michaels
Center for the Ecological Study of Perception and Action
University of Connecticut
Faculty of Human Movement Sciences
Vrije Universiteit, Amsterdam

Runeson, Juslin, and Olsson (2000) proposed (a) that perceptual learning entails a transition from an inferential to a direct-perceptual mode of apprehension, and (b) that relative confidence—the difference between estimated and actual performance—indicates whether apprehension is inferential or direct. In 3 experiments participants received feedback on judgments of force; the results replicated Runeson et al.'s observed decrease in overconfidence but showed more overconfidence. Relative confidence depended on how performance was defined. An attempt to manipulate confidence failed, but trait confidence affected relative confidence. It was concluded that overconfidence does not necessarily signal inferential functioning and that a decrease in overconfidence might occur in a direct-perceptual mode. A theory of learning within the direct-perceptual mode, in addition to learning through a mode transition, appears necessary.

The ecological approach in psychology has characterized perceptual learning in terms of differentiation (Gibson & Gibson, 1955) and the education of attention to

---

Correspondence should be addressed to Alen Hajnal, Department of Psychology, U-20, University of Connecticut, 406 Babbidge Road, Storrs, CT 06269. E-mail: alen.hajnal@uconn.edu

information—variables detectable in energy fluxes that specify properties of the perceiver's environment (Gibson, 1966). One elaboration of that theory (Jacobs & Michaels, 2002, 2006) proposes that learning is itself an informationally guided process of traversing an information manifold to a locus that permits the meeting of task demands. The approach assumes that all the while, perception is direct; that is, unmediated and noninferential. A different form of perceptual learning, the mode transition, has been proposed by Runeson, Juslin, and Olsson (2000). These authors argued, "Competence entails the use of advanced kinematic information in a direct-perceptual ('sensory') mode of apprehension, in contrast to beginners' use of simpler cues in an inferential ('cognitive') mode of apprehension" (p. 525). In other words, performance is said to improve because perceivers graduate from an inferential to a direct-perceptual mode of apprehension. In this article, we investigate the mode transition hypothesis and the measures that evidence its occurrence.

Historically, inferential perception and direct perception have been considered to be alternative theoretical positions. Indirect perceptionists have asserted that perception is inferential; ecological perceptionists have asserted that perception is direct. A first difference between inferential and direct perception is that inferential perception is associated with the use of simple variables that are ambiguous with respect to properties to be perceived. The assumption of stimulus ambiguity lays a basis for the claim that properties of interest are inferred from these so-called elementary properties. Adherents of direct perception, on the other hand, expect observers to use analytically complex variables that specify to-be-perceived properties, where specification obviates the need for inference. Runeson et al. (2000) took as one line of evidence in favor of the mode transition hypothesis that, before practice, judgments of novices are often based on variables that correlate only marginally with the to-be-perceived kinetic property but that, after a minimal amount of practice, judgments come to be based on more advanced information. Analogous findings have been reported in a number of other studies (Jacobs, Michaels, & Runeson, 2000; Jacobs, Runeson, & Michaels, 2001; Michaels & de Vries, 1998).

A second line of support for the mode transition hypothesis is provided by the degree of confidence expressed by observers (Runeson et al., 2000). Differences in confidence judgments have historically been shown to be based on whether a task is sensory or cognitive. A typical sensory task, for instance, would be to judge which of two line segments is the longer, and a typical cognitive task would be to estimate whether France or Nigeria has the larger population. Sensory tasks appear to be characterized by underconfidence—perceivers report that they are performing more poorly than they are, in fact, performing (Juslin & Olsson, 1997). Relative confidence is defined as the difference between the percentage of judgments that are actually correct and the percentage of judgments estimated to be correct by the perceiver. In contrast, cognitive tasks appear to be characterized by overconfidence or well calibrated confidence (e.g., Juslin & Olsson, 1997; see also Baranski & Petrusic, 1999, for a contrary view). Runeson et al. reasoned that confidence judgments might, therefore, be a useful index of whether observers perform a task in a direct-perceptual ("sensory") mode or an inferential ("cognitive") mode. In their colliding-balls experiment, Runeson et al. reported that confidence judgments of novices tended to be well calibrated, whereas when the novices became experts, they tended to be underconfident. This supports the mode transition theory because it seems to indicate that performance became less often inferential and more often direct perceptual.

The need for an ecological learning theory has been emphasized (e.g., Michaels & Beek, 1995), and the mode transition hypothesis might be a major breakthrough in that respect. Furthermore, the issue of mode transition is central to debates on direct versus inferential processing (or direct perception versus heuristics; e.g., Gilden & Proffitt, 1989, 1994; Runeson, 1995; Runeson & Vedeler, 1993). As pointed out by Runeson et al. (2000), previous results might have contained a mixture of inferential and direct-perceptual performance, thereby complicating their interpretation. Additional empirical evidence on the nature of the two modes and the transition is therefore desirable (cf. Andersson, Kreegipuu, & Runeson, 2001; Kreegipuu & Runeson, 1999). This contribution aims to provide such further evidence. We exploited a variation on a paradigm used by Michaels and de Vries (1998); participants observed a stick figure executing a bimanual pull (see Figure 1 for a static image from the displays) and were asked to judge the relative pulling force.

The pull-perception task is suited for our purpose for three reasons. First, as is relative mass, relative force is a kinetic (or dynamic) property. Because the optic array comprises only kinematic variables (e.g., velocities and angles), kinematic variables must lay the basis for the perception of kinetic properties. Relative pulling force has been shown to be specified by kinematic patterns (Michaels & de Vries, 1998), which provides perceivers with the opportunity to detect such information

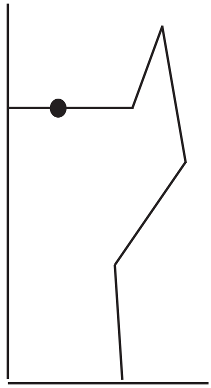

FIGURE 1 The starting and ending configuration of the stick figure puller used in all experiments. See Method sections for details.

and, thus, to apprehend relative pulling force in a direct-perceptual mode. Second, Michaels and de Vries showed that observers often begin with less useful variables but converge on more useful variables if given practice with feedback. Finally, dynamic pulling is easy to simulate; the center-of-mass motion of human pullers can be described well using a model with only a few degrees of freedom (Michaels & Lee, 1996; Michaels, Lee, & Pai, 1993). We used this model to generate center-of-mass motions of the stick figure pullers. Given various constraints (constant elbow and ankle positions), the center-of-mass positions are uniquely related to limb positions, and the center-of-mass motions can easily be transformed into stick figure motions.

Research has shown that both specifying and nonspecifying variables play a role in the perception of pulling force; the implicated nonspecifying variables are maximal displacement and maximal speed of the puller's center of mass (Michaels & de Vries, 1998). The particular variables that are exploited depend on the relative usefulness of variables, which can be measured as the correlations between relative pulling force and the nonspecifying variables. These correlations depend on the characteristics of the particular set of pulls, which depend, in turn, on which model parameters are manipulated to generate the set of pulls. For Experiment 4 of Michaels and de Vries and the experiments reported here, the correlation between a force-specifying invariant and maximal displacement was –.05, and the correlation between the force-specifying invariant and maximal speed was .74 (the intercorrelation of speed and displacement was .62). The rationale for choosing this particular set of stimuli was twofold. First, keeping the correlations among the kinematic variables as low as possible would, we hoped, not inadvertently result in good feedback when the observer is using a nonspecifying variable. Second and relatedly, this set has been shown to encourage perceivers to converge on higher-order specifying variables (cf. Experiments 3 and 4 of Michaels & de Vries, 1998). We expected that this set of displays would be suited to our purpose, and we therefore used it in all experiments and in all phases of the experiments (i.e., pretest, practice, and posttest).[1]

EXPERIMENT 1

As described previously, Runeson et al. (2000) demonstrated changes in variable use and a decrease in relative confidence in a paradigm in which perceivers estimated the relative mass of colliding balls. Convergence on more useful variables was also evident in the perception of pulling force (Michaels & de Vries, 1998). Here we attempt to replicate the decrease in confidence in the latter paradigm. Runeson et al. asked observers to make binary judgments; more precisely, they asked observers to indicate which of two colliding balls was the heavier. We also

---

[1] Experiments 1 and 2 are, in fact, replications of experiments presented in Jacobs (2001, chap. 4); he found little learning in a less constrained set of displays.

asked participants to make binary judgments, first to make the experiment similar to Runeson et al.'s experiment, and second because binary judgments are more appropriate than quantitative judgments if one also solicits confidence judgments. We therefore used the classic psychophysical method of constant stimuli, in which participants are asked to categorize each display with reference to a standard. The standard was not explicitly presented; participants were to judge whether the force on a particular pull was above or below the average of all forces in the collection of displays. This is a commonly used technique in visual psychophysics (e.g., McKee, 1981; Regan & Hamstra, 1993; Regan & Vincent, 1995).

Method

Participants. Eight undergraduate students at the University of Connecticut participated. They received credit in partial fulfillment of their introductory psychology course's experimental requirement.

Displays and design. The center-of-mass motions of the stick figure display were based on the three dynamic parameters of the model of center-of-mass motions in bimanual pulling of Michaels and Lee (1996): slack, stiffness, and torque. The model consists of a mass, *m*, on an inverted pendulum, as depicted in Figure 2. The pendulum is attached to a vertical support by an initially slack elastic cord (Figure 2, Panel A). A constant torque and, after initiation, also a torque caused by gravity accelerate the mass away from its initial upright position until the cord is stretched (Figure 2, Panel B). The stiffness of the cord then decelerates the pendulum until it reaches the position of maximal displacement (Figure 2, Panel C), where the stiffness reverses the motion of the pendulum, eventually leading it back to its initial position. Thus, the motion of a pendulum of a particular mass and length is determined by three model parameters: a constant torque, the initial slack of the cord, and the stiffness of the cord.

We created 36 simulations—of bimanual pulls on a horizontal bar at elbow height—by manipulating two of the model's three parameters. Stiffness and torque

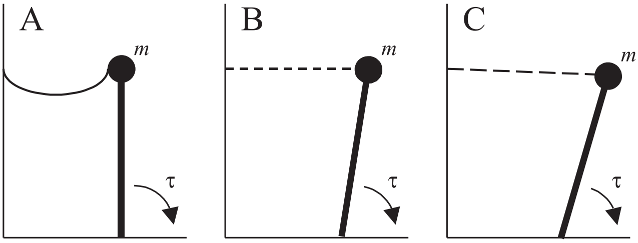

FIGURE 2 The inverted pendulum model for pulling presented by Michaels et al. (1993). See text for explanation.

had six levels each (6, 9, 12, 15, 18, 21 kN/m and moment arms of 2, 4, 6, 8, 10, 12 cm, respectively). Slack was constant at 7 cm. In all trials the length of the pendulum was 90 cm and the mass of the bob was 75 kg. The stick figures on the screen were 7 cm tall, about 25 times smaller than the pullers whose motion was simulated. The stick figures were presented as white lines on a black background, in a simulation window of about 10 × 10 cm (350 × 350 pixels), on a Macintosh CRT monitor at its refresh rate of 75 Hz. Using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) within a Matlab program, we made sure that the animations of the stick figures were tightly synchronized with the buildup of the computer screen. Participants viewed the animations from a distance of approximately 40 cm.

The experiment consisted of a 36-trial pretest, four 36-trial practice blocks, and a 36-trial posttest. Each block comprised the 36 displays in completely randomized order.

Procedure. Participants started a series of trials by placing the cursor and clicking on a *start* button with a computer mouse. After each display generated by a given set of simulation values was presented three times in rapid succession, a response window appeared on the screen. Using the mouse, participants entered two judgments: a binary judgment about whether the peak force exerted by the puller was *above* or *below* the average of all pulls, which was entered by clicking on one of two buttons; and a quantitative judgment, entered on a slider ranging from 50% to 100%: "Placing the slider at the 50% end means that your force judgment is a random choice … placing the slider at the 100% end means that you are sure that your force judgment is correct." Participants were free to change their judgments until they clicked the *ready* button. In the pretest and posttest, the *ready* button initiated a new trial. On practice trials, clicking the *ready* button was followed by feedback: The message *correct* or *incorrect* appeared on the screen, depending on whether or not the participant had correctly judged that the pulling force was above or below average. Each session lasted about 1.5 to 2 hr.

Results and Discussion

On average, 61% of the force judgments were correct in the pretest and 81% of judgments were correct in the posttest. Participants were able to perform the task, and they improved after practice with feedback. A paired *t* test on the percentages of correct judgments in the pretest and posttest indicated that the improvement was significant, $t(7) = 3.03$, $p < .02$, one-tailed. The point biserial correlation between judgment and force increased from .199 on the pretest to .639 on the posttest, on average; a *t* test on these correlations[2] showed a significant improvement, $t(7) = 3.44$, $p < .01$, one-tailed.

---

[2] All significance tests on correlations in this and subsequent experiments were done on the correlations' Z transformations.

TABLE 1
Performance and Relative Confidence for Individuals in Experiment 1

	% Correct		*Relative Confidence*	
Participant	*Pretest*	*Posttest*	*Pretest*	*Posttest*
1	66.7	91.7	4.0	–2.1
2	66.7	91.7	3.7	–11.6
3	75.0	80.6	4.8	12.6
4	42.9	85.7	29.9	–1.6
5	69.4	80.6	11.5	4.1
6	33.3	78.8	48.5	–3.1
7	63.9	75.0	13.6	5.6
8	72.2	63.9	10.4	15.6
M	61.3	81.0	15.8	2.4
SD	14.9	9.1	15.7	8.9

As to the main question of whether relative confidence, the difference between average confidence and percentage correct, decreased with practice, we found that it did so systematically over blocks of trials, from 15.8% in the pretest to 2.4 % in the posttest. An analysis of variance (ANOVA) with Block (1 to 6) as a within-subjects factor showed that the decrease was significant, $F(5, 35) = 2.28$, $p = .035$, again one-tailed. In line with the mode transition hypothesis, we replicated a decrease in relative confidence.

Both the performance and relative confidence scores are presented by individual participants in Table 1. Six out of 8 participants exhibited a decrease in relative confidence. Half of the participants became underconfident in the posttest, whereas the rest remained overconfident. The standard deviation of relative confidence decreased from 15.7% in the pretest to 8.9% in the posttest, suggesting that over participants, confidence judgments became more consistent. The difference between pre- and posttest scores was 13.4% on average, with a large standard deviation of 19.7%. These results reflect large individual differences among participants.

To summarize, this experiment replicated the decrease in relative confidence reported by Runeson et al. (2000). We, however, found more overconfidence: With practice, our participants changed from overconfidence to near well calibrated confidence, whereas participants in the experiments of Runeson et al. changed from well calibrated confidence to underconfidence (more precisely, from 0% to –11% underconfidence). The next experiment addresses an explanation of the decrease in confidence that is invited by the higher level of confidence in our experiment.

EXPERIMENT 2

The fact that we observed overconfidence even in the posttest of Experiment 1, whereas Runeson et al. (2000) observed underconfidence, might have been due to the more frequent use of an inferential mode by our observers which, in turn, might have been due to the use of nonspecifying variables in the case of our observers or to differences in simulation techniques and display qualities (Runeson & Andersson, 2004). More interesting, however, is that the larger overconfidence invites an alternative view of the origin of the decrease in overconfidence. Assume that confidence always reflects the accuracy with which the exploited variable is detected, whatever that variable may be. Imagine further two perceivers who express an equal absolute confidence in their detection of their exploited optical variables. Say that one participant uses a nonspecifying variable; this would affect performance, but not confidence. Such a perceiver would show overconfidence because performance is less accurate than the confidence. The second perceiver, who detects a specifying variable with the same degree of accuracy, would perform better relative to the experimenters' intended quantity. When that better performance is compared with the same confidence, there would be less overconfidence than shown by the first perceiver. Applying this logic to our Experiment 1, most overconfidence was found in the pretest—namely because novices are more likely to use nonspecifying variables. Under this interpretation, *overconfidence would be due to the fact that the performance measure is not a measure of what participants are actually doing.*

If, indeed, overconfidence is due to reliance on nonspecifying variables, rather than to a different kind of cognitive process, an apparent decrease of overconfidence, as observed in Experiment 1, might be due to convergence on specifying variables. This reasoning led to the expectation that overconfidence, as measured here, should increase if perceivers are induced to rely more on nonspecifying variables. To test this, we adopted the methods of Experiment 5 of Michaels and de Vries (1998), which showed that perceivers' reliance on nonspecifying variables could be attained through feedback contingent on such variables. In Experiment 2, we made feedback contingent on often-used nonspecifying variables. For a first group of participants, the displacement group, feedback "on accuracy" was contingent on the maximal displacement of the puller's center of mass. For a second group, the speed group, feedback was contingent on maximal speed. To the extent that categorical feedback is also effective in inducing reattunement (Michaels & de Vries, 1998, had used scalar feedback), we expected that these groups would show an increase in overconfidence. It is more likely that an effect would be observed with displacement because speed was highly correlated with relative pulling force ($r = .74$). Feedback contingent on speed, therefore, may be the same as feedback contingent on force in a large percentage of trials.

In summary, participants who receive feedback based on nonspecifying variables are hypothesized to show an increase in overconfidence. This is in contrast to participants in Experiment 1, who received feedback on force and exhibited a decrease in overconfidence.

Method

As in Experiment 1, participants were instructed to judge peak force and to give confidence estimates. The same pretest-practice-posttest design was used. Also as in Experiment 1, participants were instructed to judge relative pulling force. Feedback, however, was based on the maximal speed of the puller's center of mass for the speed group and on the maximal displacement of the center of mass for the displacement group. This means, for instance, that a response of *below average* was said to be correct if maximal speed or maximal displacement, respectively, was less than average. The speed group and the displacement group each had 8 participants. Perceivers in both groups were instructed to enter their answer and confidence judgment with respect to force. They were told that they would receive feedback on their performance in the training blocks, but the nature of the feedback was not revealed. In all other regards, the experiment was the same as Experiment 1.

Results and Discussion

We divide our results into three sections, each adopting a different definition of performance accuracy. In the first section we define performance as we had in Experiment 1, in terms of the accuracy of judging pulling force. In the second section, performance is defined relative to the feedback conditions that distinguished groups. In the third section we introduce the idea of an intrinsic performance measure. In all three sections, to facilitate comparisons with Experiment 1, those data are included as a third condition, in which force feedback was given.

Force-referential performance. Remember that our hypothesis consisted of two parts: Participants would converge on nonspecifying variables on which they received feedback, and, as a result, overconfidence would increase. We address these issues in turn.

If participants indeed converged on variables on which they received feedback, the force, speed and displacement groups should differ in which variable they exploited in the posttest—the force group (Experiment 1) relying more on a specifying variable, the speed group relying more on maximal speed, and the displacement group relying more on maximal displacement. The r^2 of the correlations relating judgments to each of the kinematic variables are presented by group and blocks of trials in Figure 3.

Inspection of Figure 3 shows again that in the force-feedback condition (i.e., Experiment 1) judgments came to correlate more highly with actual force (top panel) than did the judgments of members of the other groups. A Group (force feedback, speed feedback, and displacement feedback) × Block repeated measures ANOVA on the correlations between judgments and force showed that the difference between feedback groups on the judgment–force *r*s was significant, $F(2, 21)$

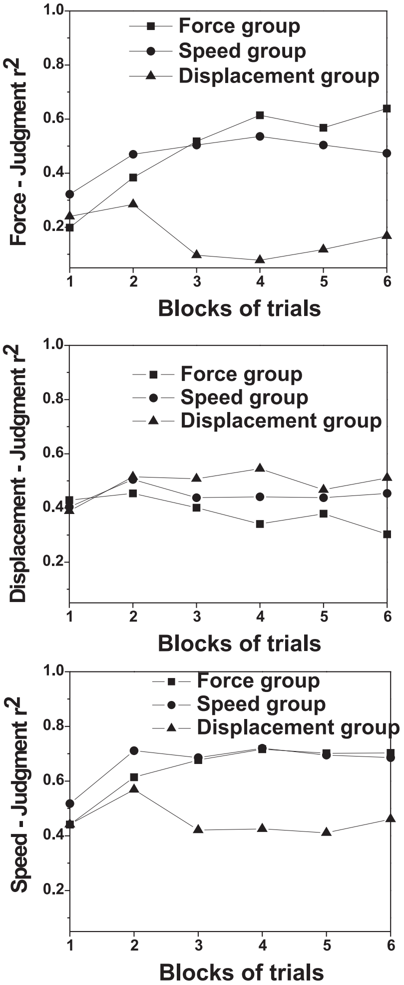

FIGURE 3 Average squares of correlations between candidate variable and binary judgments over blocks for the three feedback groups. In the top panel, the candidate variable is force; in the center, it is displacement; and in the bottom panel, it is speed.

$= 10.02$, $p < .001$, revealing the superiority of the force group. The middle panel hints that a possible increase in the correlation of judgment and displacement occurs only in the group given displacement feedback; however, an ANOVA on displacement–judgment correlations yielded no significant differences among groups. The bottom panel seems to show that both the speed and force groups showed increasing correlation with speed. An ANOVA on the judgment–speed correlation revealed higher correlations for the speed and force groups than for the displacement group, $F(2, 21) = 7.4, p < .004$. That both force and speed groups show high speed–judgment correlations is not surprising given the high correlation between force and speed.

The way the results were plotted in Figure 3, although making the previously discussed characteristics clear, obscures other characteristics. Comparing the values for the force group in the top and bottom panels, for instance, reveals that the correlation between force and judgment was consistently smaller than the correlation between speed and judgment. This suggests that convergence on a specifying variable was perhaps thwarted by the high correlation between force and speed. Second, a comparison of all sets of correlations in the displacement group showed a slight systematic superiority of the displacement–judgment correlations over the velocity–judgment correlations and a huge superiority of the displacement–judgment correlations over the force–judgment correlations. But still the judgment–displacement r^2s in this group were never higher than .6. Thus, it seems that the displacement feedback, although preventing participants from relying on force-specifying information, did not induce systematic reliance on displacement. Again, the benefit of feedback on displacement seemed to spill over onto speed, presumably due to the fairly high intercorrelation of displacement and speed.

In sum, the feedback contingent on the different nonspecifying variables appeared to foster reliance on the different nonspecifying variables, although the effects were neither as strong nor as reliable as those that had been reported when scalar, rather than categorical, feedback was offered (Michaels & de Vries, 1998, Experiment 5).

We now turn to the second part of our hypothesis, which predicted that a convergence on nonspecifying variables in the speed and displacement groups would result in an increase in overconfidence. The left half of Table 2 presents the percentages of trials in which participants correctly judged force in the pretest and posttest, along with their average relative confidence. A Group (force feedback, speed feedback, and displacement feedback) × Block (pretest, posttest) ANOVA on relative confidence revealed a significant Group × Block interaction, $F(2, 21) = 4.27, p = .028$; as expected, the force group showed a decrease in relative confidence, whereas the speed and displacement groups showed an apparent increase

TABLE 2
Average Performance and Relative Confidence by Groups in Experiment 2

	Force-Referential				*Feedback-Referential*			
	% Correct		*Relative Confidence*		*% Correct*		*Relative Confidence*	
Group	*Pretest*	*Posttest*	*Pretest*	*Posttest*	*Pretest*	*Posttest*	*Pretest*	*Posttest*
Force	61.3	81.0	15.8	2.4	61.3	81.0	15.8	2.4
Speed	64.9	69.4	15.5	20.1	72.6	77.8	7.9	11.8
Displacement	60.8	60.2	15.5	25.5	66.1	69.5	10.3	16.1

from pretest to posttest. There was also a marginally significant main effect of group, $F(2, 21) = 3.35, p = .055$; the force condition tended to have lower relative confidence than did the speed and displacement groups.

In short, we found that Experiment 1's decrease in relative confidence was limited to the condition in which feedback was based on force. In conditions in which participants were encouraged to rely on other variables, they did so, and their relative confidence increased. This is in contrast to the results in the colliding-balls paradigm reported by Runeson and Andersson (2004): Although their r^2 values reflect the same kind of influence of feedback as in our study, namely, participants' judgments appear to correlate highly with the variable on which they received feedback, relative confidence decreased in all three feedback conditions. Contradictory empirical results from the pulling-force paradigm and the colliding-balls paradigm undermine the role of relative confidence as an indicator of mode of apprehension.

Feedback-referential performance. Previously, performance, and hence relative confidence, was defined relative to the explicit task demand of judging force. Whereas this is a straightforward choice when force feedback is given, the speed feedback and displacement feedback redefine the task demands. Although participants were instructed to report force, the feedback indicated that force was related to kinematic variables that, in fact, did not correlate highly with the simulated force. In this section, we reconsider performance by defining it with respect to the task demands created by the feedback. Under the new definition, percentage of correct judgments in the speed group is defined with respect to speed, and percentage of correct judgments in the displacement group is defined with respect to displacement. The results so defined are presented in the right half of Table 2.

Feedback-referential performance improved over blocks, as already implied by the changing correlations depicted in Figure 3: A Block (pretest, posttest) × Group (force feedback, speed feedback, and displacement feedback) ANOVA on percentage correct showed a significant effect of block, $F(1, 21) = 12.49, p = .002$; performance improved over blocks. The Block × Group interaction, $F(2, 21) = 3.73, p = .04$, showed that performance in the force group improved more than in the other groups. Clearly, speed and displacement feedback were not as effective as force feedback in drawing participants to rely on these variables. The modest improvement in the speed and displacement groups might be partially attributable to the instructions to report force. Also, given the intercorrelation between the two variables, the categorical feedback would be the same for the two feedback criteria in the majority of trials.

A Block × Group ANOVA on the relative confidence using the feedback-referential definition of performance revealed only a significant interaction, $F(2, 21) = 3.67, p = .04$; the force condition showed a large decrease in relative confidence, whereas the speed and displacement groups continued to show an increase in relative confidence. We had expected the same degree of learning in the three groups and, therefore, the same decrease in relative confidence. Although we did not find that equivalent decrease, it is clear from a comparison of columns 4 and 8 of Table 2 that the overconfidence in the speed and displacement conditions was significantly less than what was obtained through the force-referential measure on these groups, as demonstrated by the significant interaction in a Block (pretest, posttest) × Method (force referential, feedback referential) ANOVA, $F(1, 30) = 5.52$, $p < .026$.

Our analyses of feedback-referential performance led us to conclude that degree of overconfidence depends in part on one's definition of performance. Note, too, that none of our groups demonstrated underconfidence, Runeson et al.'s (2000) hallmark of direct perception. In the next subsection, we examine whether an even more intrinsic performance measure will reveal such underconfidence.

Intrinsic performance. The just-presented analysis supposes that a participant's judgment about confidence is in terms of his or her judgment about performance. Thus, a determination of underconfidence or overconfidence must take into account what the participant is doing. Force-referential performance was defined in terms of explicit task demands ("judge force"), whereas feedback-referential performance was defined relative to implicit task demands. In this section, we go even further in seeking an intrinsic measure of participant performance.

For Table 2 we had defined performance and relative confidence in terms of feedback; Table 3 presents analogous results, but in terms of the variables that appeared to be exploited by a participant on the block in question. So for example, the speed–pretest performance cell of Table 3 is the average of all participants who appeared to base their judgments on speed on that block, as evidenced by a higher percentage correct relative to speed than relative to force or displacement. Similarly, participants in the force–posttest cell had higher percentages correct relative to force than relative to speed or displacement. Put another way, our intrinsic measure of performance is operationalized as the best performance relative to the three variables in question and irrespective of explicit instructions or feedback group membership. Note that a participant can be in one row for the pretest and another row for the posttest.

TABLE 3
Average Performance and Relative Confidence in Experiment 2
Using Exploited Variable as Standard

	% Correct		*Relative Confidence*	
Exploited Variable	*Pretest*	*Posttest*	*Pretest*	*Posttest*
Force	73.6	87.4	7.6	–2.8
Speed	76.2	79.4	3.0	6.7
Displacement	68.7	79.9	6.0	8.5

A Variable (force, speed, displacement) × Block (pretest, posttest) ANOVA on intrinsic percentage correct from pretest to posttest showed a significant increase, $F(1, 42) = 11.71, p < .001$. The same ANOVA on intrinsic relative confidence revealed no significant effects. The lack of significant differences in relative confidence suggests that this measure does not discriminate among groups based on which variable was exploited. More specifically, the lack of an effect suggests that intrinsic relative confidence may not always be diagnostic of the use of a particular kind of variable or mode of apprehension. In our view, the finding that a decrease in confidence can disappear when analyzed with intrinsic measures should be seriously taken into account if one uses confidence measures to distinguish perceptual modes.

Let us turn finally to the performance of individuals. Both the intrinsic performance and relative confidence scores are presented by participants in Table 4; it

TABLE 4
Performance and Relative Confidence for Individuals in Experiment 2
Using Exploited Variable as Standard

Feedback Group/Participant	*Intrinsic % Correct*		*Intrinsic Relative Confidence*	
	Pretest	*Posttest*	*Pretest*	*Posttest*
Force				
1	72.2	91.7	–1.5[a]	–2.1
2	72.2	91.7	–1.8[a]	–11.6
3	86.1	86.1	–6.3[a]	7.1[a]
4	60.0	85.7	12.8[b]	–1.6
5	75.0	80.6	5.9[a]	4.1
6	63.3	84.8	18.5[b]	–9.1[a]
7	75.0	86.1	25.0[a]	–5.5[a]
8	83.3	80.6	–.7[a]	–1.1[a]
Speed				
9	72.2	77.8	.8[a]	10.3[a]
10	69.4	75.0	9.7[b]	20.7[a]
11	66.7	75.0	26.2[a]	18.5[a]
12	77.8	77.8	13.5[a]	9.3[b]
13	75.0	75.0	9.8	15.4[a]
14	77.8	80.6	–12.1[a]	10.8[a]
15	72.2	77.8	2.4[b]	10.9[a]
16	69.4	83.3	6.9[a]	–4.4[a]
Displacement				
17	77.8	88.9	5.8[b]	2.4[b]
18	63.9	75.0	–12.3[a]	–3.3[a]
19	77.1	72.2	–2.9[b]	12.1[b]
20	75.0	57.1	–4.8[a]	21.3[a]
21	58.3	60.0	14.3[a]	–5.1[a]
22	55.6	77.8	–4.6[b]	10.2[b]
23	57.1	52.8	14.0[a]	19.1[a]
24	63.9	72.2	5.5	1.9[a]

[a] Indicates apparent reliance on speed. [b] Indicates reliance on displacement.

conveys a picture somewhat different from that of Table 3, which suggests that underconfidence appears only in the group that received force-based feedback. Table 4 reveals that relative confidence shows large individual differences. Of importance, some participants who exploit variables other than force also exhibit underconfidence. Negative relative confidences that bear the superscipts [a] or [b] in Table 4 are blocks on which underconfidence was seen when the participant's judgment was not based on specifying information. Thus, at the level of the individual, at least, underconfidence is clearly not diagnostic of exploiting specifying information in a perceptual mode.

EXPERIMENT 3

The confidence displayed by participants in a task such as ours is presumably due not only to how well they assess their own performance. Another factor that is likely to influence confidence is the feedback provided by the experimenter. The participant who is told that he or she is correct on 90% of the trials would presumably be more confident than the one who is told of being correct on 60% of trials, irrespective of their actual performance. For example, the overconfidence that participants display when they use the wrong variable in a perceptual judgment is likely to be decreased by feedback that performance is poor. The possible use of confidence judgments as indication of mode of apprehension makes it interesting to know what other factors contribute to confidence judgments.

To test the extent to which confidence judgments depend on the given feedback, we contrived in Experiment 3 a situation in which the feedback on force differed between two groups. In particular, the criterion for what constituted a correct response was modified so as to increase the percentage of "correct" judgments in one group (the encouraged group) and to decrease it in another group (the discouraged group). Encouraging feedback might increase absolute confidence, whereas discouraging feedback might decrease it, both independently of actual performance. Such a change in absolute confidence, in turn, would be reflected in relative confidence measures.

Method

Sixteen new participants were randomly assigned to one of two groups. The 36 visual displays presented were identical to those of Experiment 1, and as in Experiment 1, only force feedback was given. The simulated forces ranged from 329 to 864 N, and participants were again to indicate whether a given force was above or below the average force in the group. However, rather than basing the feedback on the actual mean (583 N), which would have yielded correct feedback for the participants, judgments regarding the forces near the mean force (552–621 N) were always indicated to be *correct* to the encouraged group, and to be *incorrect* to the discouraged group, irrespective of the participants' judgment. This swing set comprised 8 out of 36 displays in each block of trials; thus, if participants in the two conditions were judging at random, the encouraged group would get positive feedback on 61% of trials, whereas the discouraged group would get positive feedback on only 39% of trials. If participants in the two groups always correctly categorized the pulls, the encouraged group would get positive feedback on 100% of trials, whereas the discouraged group would get positive feedback on only 78% of the trials.

Results and Discussion

Before examining the confidence judgments, let us examine performance in the categorization task. Again we include in our analysis the data from Experiment 1, which we now label the accurate feedback condition. We first asked whether participants differed in their force-judgment performance. A Block × Group ANOVA on percentage correct judgments revealed only one significant effect, that of block; the percentage correct categorization increased monotonically over blocks from 63.4% correct to 77.2% correct, $F(5, 105) = 8.87$, $p < .001$. Feedback groups showed no effect on participants' success in judging force. The averages are plotted in Figure 4.

As to relative confidence judgments, a Block × Group ANOVA on the relative confidence in the force-referential definition of performance revealed only a significant decrease over blocks of trials, $F(5, 105) = 5.172$, $p < .001$. The feedback groups did not differ, as is clear from Figure 5. Figure 5 also shows that the decrease in relative confidence did not result in underconfidence. This observation prompted us to consider once again an intrinsic dependent variable as a measure of performance. An ANOVA on the relative confidence intrinsically

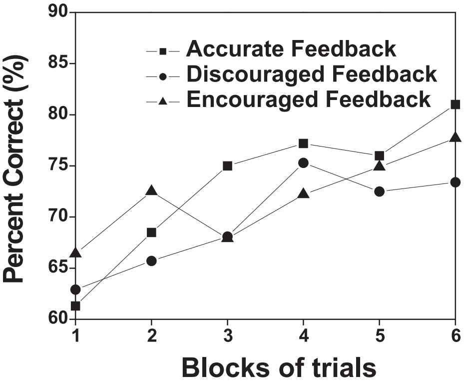

FIGURE 4 Percentage of correct judgments over blocks of trials for the encouraged, discouraged, and accurate feedback groups in Experiment 3.

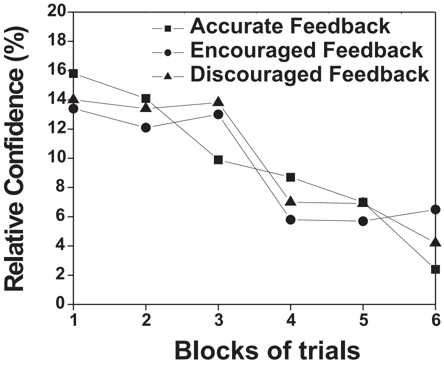

FIGURE 5 Relative confidence judgments over blocks of trials for the encouraged, discouraged, and accurate feedback groups in Experiment 3.

defined also failed to reveal an effect involving feedback group, although predictably, relative confidence was lower.

We conclude that quality of (erroneous) feedback did not have the predicted effect on confidence. Rather than influencing confidence judgments, the feedback may have influenced how they performed the task (e.g., what variable they exploited). We had hoped that giving false feedback only on trials very close to the mean force would have had little effect on performance, but it seems to have been used as genuine feedback, altering variable usage.[3]

Even if erroneous feedback had tended to affect confidence judgments, an examination of the results of individuals (Table 5) suggests that there may be a confidence trait variable whose influence dominates. We see that on Block 1 (the pretest), in which feedback had not yet been given, confidence ranged from 61% to 93%. In the remainder of this section we explore how these individual differences are reflected in relative confidence. Because the feedback manipulation had only a minor effect, we ignore condition in the following analyses.

Absolute confidence on Block 1 might be taken as a crude measure of trait confidence, if confidence can be shown to be unrelated to Block 1 performance. We regressed absolute Block 1 confidence on Block 1 performance and found that they were not correlated, $r(22) = .004$, suggesting that Block 1 confidence is a relatively uncontaminated measure of the trait. We then asked whether this trait affected the relative confidence observed on subsequent trials. We divided participants into three groups (high, medium, and low confidence) based on absolute confidence

---

[3] This conclusion followed primarily from an analysis of intrinsically defined performance, which was significantly lower with false feedback than with accurate feedback, $F(1, 21) = 4.981, p < .037$, and the difference got larger over blocks of trials, $F(5, 105) = 6.111, p < .001$. We interpret this difference as suggesting that the feedback continually encouraged perceivers to look for better information variables.

TABLE 5
Performance and Absolute Confidence for Individuals in Experiment 3

Feedback Group/Participant	*% Correct*		*Absolute Confidence*	
	Pretest	*Posttest*	*Pretest*	*Posttest*
Accurate				
1	66.7	91.7	70.7	89.6
2	66.7	91.7	70.4	80.1
3	75.0	80.6	79.8	93.2
4	42.9	85.7	72.8	84.1
5	69.4	80.6	80.9	84.7
6	33.3	78.8	81.8	75.7
7	63.9	75.0	77.5	80.6
8	72.2	63.9	82.6	79.5
M	61.3	81.0	77.1	83.4
SD	14.9	9.1	5.0	5.7
Encouraged				
9	71.4	85.7	71.2	79.7
10	57.1	65.7	77.0	86.6
11	74.3	82.9	80.2	87.0
12	65.7	77.1	84.7	74.7
13	51.4	71.4	86.3	83.7
14	68.6	68.6	74.3	80.6
15	51.4	71.4	61.0	74.0
16	62.9	64.7	75.2	73.4
M	62.9	73.4	76.2	78.0
SD	8.8	7.8	8.0	5.5
Discouraged				
17	77.1	82.4	82.5	90.4
18	73.5	77.1	77.0	73.4
19	65.7	60.0	69.5	72.2
21	65.7	85.3	78.7	85.8
20	62.9	71.4	88.5	98.6
22	60.0	82.9	92.1	90.4
23	45.7	77.1	83.6	84.3
24	77.1	79.4	79.4	76.5
M	66.0	77.0	81.4	84.0
SD	10.4	8.1	7	9.3

judgments in Block 1. A Block (2–6) × Group ANOVA on relative confidence (using the force-referential performance measure) showed a marginally significant effect of trait, $F(2, 21) = 2.81, p < .083$. We also examined the intrinsic measure of relative confidence. A Block × Group ANOVA revealed a significant main effect of trait confidence, $F(2, 21) = 3.88, p < .037$, suggesting that relative confidence judgments were influenced by individual differences in a systematic way. Averages are depicted in Figure 6. The group that had the lowest absolute confidence in

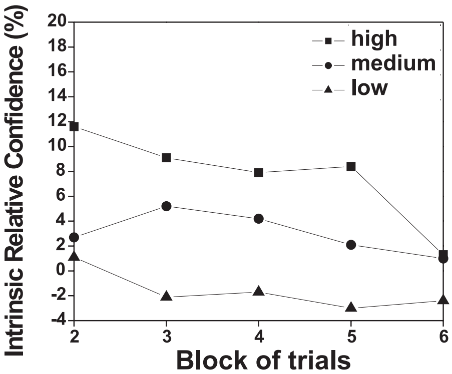

FIGURE 6 Intrinsic relative confidence over blocks of trials for the low-, medium-, and high-trait-confidence groups as operationalized by absolute confidence levels on Block 1 in Experiment 3.

Block 1 was on average underconfident, whereas the more confident groups had an average relative confidence in the overconfidence range.

To summarize, in Experiment 3 we attempted to manipulate confidence by giving false feedback about performance. Our goal was to establish whether factors besides mode of apprehension determine confidence. That manipulation failed, but our attention was called to large individual differences on Block 1, in which no feedback had been given. This absolute confidence was unrelated to performance. We observed that participants who showed high confidence on Block 1 tended to show relative overconfidence on later blocks, whereas participants who showed low confidence on Block 1 tended to show relative underconfidence on later blocks. In short, trait factors, which are independent of both variable usage and mode of apprehension, seem to affect relative confidence.

GENERAL DISCUSSION

In three experiments we explored some characteristics of relative confidence to see its strengths and weaknesses as an index of the type of processing that perceivers engage in when they make judgments about events, as proposed by Runeson et al. (2000). We were particularly interested in the extent to which factors other than mode of apprehension could influence relative confidence. We reasoned that it would be important to take such factors into account if using particular values of relative confidence, and relative confidence in general, as an indicator of whether a perceiver is in an inferential or direct perceptual mode.

The task we used was the judgment of relative pulling force. In Experiment 1, we adapted the Michaels and de Vries (1998) pull-perception paradigm to permit the evaluation of confidence; participants were asked to make categorical judgments (regarding whether the pulling force was greater than or less than average) and to estimate the confidence they had in their judgment. The results replicated two of Runeson et al.'s (2000) basic empirical effects: that feedback improved percentage of correct judgments over blocks of trials and decreased the degree of relative confidence. However, we did not find average underconfidence, the hypothesized hallmark of direct perception, as Runeson et al. saw on later blocks.

The higher overconfidence found in our experiment led us to propose an alternative explanation of the decrease in confidence. If confidence judgments are assumed to veridically reflect the accuracy with which observers detect the optical variable that they use, whatever variable that may be, they might appear overconfident before practice because they tend to use nonspecifying variables, which would not permit accurate performance. Furthermore, overconfidence might then disappear with practice merely because observers change their variable use and come to rely on specifying variables, which permits best performance. This hypothesis was tested in Experiment 2.

If, as argued, the average overconfidence was due to some participants' use of nonspecifying variables, fostering reliance on nonspecifying variables should increase relative confidence. To test this, in Experiment 2 we tried to alter performance level by giving participants feedback based on nonspecifying variables. This manipulation succeeded in steering participants toward relying on the variable they received feedback on, regardless of whether it specified the to-be-judged property, force. More important, the manipulation indeed led to an increase in overconfidence, supporting the hypothesis that the decrease in confidence in Experiment 1 had been due to convergence on a specifying variable. This hypothesis was further confirmed by the finding that no effects regarding relative confidence were observed when we used intrinsic measures, that is, when we measured confidence relative to the variables that observers appeared to use, rather than relative to the explicit task variable.

In Experiment 3, we tried to manipulate absolute confidence directly by giving false feedback about performance, but we did not succeed. This is, of course, encouraging for the use of relative confidence in distinguishing modes of apprehension because the fewer factors that influence relative confidence, the better. Unfortunately, however, we observed large individual differences in absolute confidence on Block 1, and they were not correlated with performance. We then split observers into three groups based on their absolute confidence level in Block 1. A significant effect of group suggested that what seems to be a trait variable heavily influenced the relative confidence.

To summarize, our finding that confidence judgments do not appear to merely reflect feedback is encouraging for the use of confidence judgments in distinguishing the modes of apprehension, whereas the finding of large individual difference and the finding that the decrease in confidence can be due to convergence on specifying information are not. What are the implications of these findings concerning confidence judgments for the mode transition proposed by Runeson et al. (2000)? And, more generally, what are the implications for ecological theories of perceptual learning? Before we turn to these main questions, we want to make a few prior observations concerning task-defined versus intrinsic performance measures, quantitative versus binary feedback, and inference versus perception.

Task-Defined Versus Intrinsic Performance Measures

The skeptic might wonder how seriously one should take conclusions based on arbitrary definitions of performance. Arguably, we asked people to perform a task, then ultimately scored their performance without attention to the explicit task. We believe we can defend this apparently unorthodox move.

First, given that participants were asked to assess their certainty in the accuracy of their judgments, *their* standard was surely whatever they were doing, irrespective of its agreement with how we as experimenters defined success. That is, a participant presumably judges the likelihood of being correct on the basis of the attended-to variable: If it is extreme (i.e., well above or below its mean), the perceiver is likely to be confident; if it is near its mean, the perceiver is likely to be uncertain. If the experimenter chooses a performance measure different from the participant's standard, *relative confidence* becomes a measure of the relative mismatch between what participants are doing and what the experimenter assumes they are doing. In that sense, one might argue that an intrinsic performance measure is less arbitrary than one rooted in explicit task demands.

How to define and measure what perceivers are doing and how well they are doing it should not be considered to be a simple business. This is illustrated by Gibson's (1966) arguments about the meaning of geometric illusions (see also Michaels & Carello, 1981, p. 92). When a perceiver's judgment about the length of a line, say in the Müller-Lyer figure, is different from the scientist's measure, the conventional interpretation is that the perceiver is in error. Gibson (1966, p. 313) remarked, however, that it is the scientist who is not measuring the variable that the perceiver exploits in reporting length. Presumably, perceivers are reliably detecting whatever they detect. Thus, finding poor performance in length estimation in a Müller-Lyer task can be simply the result of bringing to bear an arbitrary standard. Crisp operational definitions of performance may simplify experimental work, but they can obscure more fundamental principles.

Quantitative Versus Binary Feedback

In our attempt to replicate Runeson et al.'s (2000) findings in the pulling-force paradigm, we adapted the pull-perception paradigm to Runeson et al.'s measurement and feedback procedure. Feedback was always categorical, given as "correct" or "incorrect." As such, it offered a restricted range of information that did not allow for much perceptual discrimination. A comparison of the degree of improvement in (extrinsically measured) performance with the results of Michaels and de Vries (1998) indicates that performance benefits much more from continuous than from categorical feedback. The lack of informativeness of the categorical feedback may have slowed the march of learning, and thus possibly the march toward underconfidence. More learning may have led to degrees of underconfidence that were more in line with those reported in the collision paradigm.

Perceiving Versus Inferring

Whereas the Helmholtzian tradition holds that all, or at least most, of perceiving is a matter of inference, ecological psychologists hold that perceiving and inferring are different, and that inferring, if it happens, is subsequent to perception. As outlined in the introduction, one difference that often goes together with the distinction between inferring and directly perceiving is the type of variable that the respective processes are thought to rely on. Inferences are often said to be based on so-called lower-order variables, whereas direct perception is said to be based on analytically complex or higher-order variables. The mode transition hypothesis, however, refers to more than to this difference. Beyond the use of different types of variables, the mode transition refers to a qualitative distinction in the mode of apprehension, perception being characterized by an "immediacy" and directness of the experience of the apprehended property, and inferences being characterized as a "reasoning" process (Runeson et al., 2000, pp. 528, 534).

Although these two distinctions—type of variable and mode of apprehension—are often thought to go together, this need not always be the case. Perceivers could very well detect higher-order information that specifies some property and subsequently engage in thought-like processes. It is equally feasible that observers have a direct impression of, say, speed, on the basis of a variable that specifies speed, which is often said to be a lower-order variable. More extremely, one could claim that observers perceive force, without need for inferential processing, on the basis of speed-specifying variables (e.g., Michaels & de Vries, 1998). We argue for the independence of mode and variable type and thus accept the possible reliance on supposedly different types of variables in each perceptual mode.

One reason to view variable type and mode of apprehension as independent is that, as widely argued, the distinction between lower-order and higher-order variables might not be viable; it might not be possible to rigorously categorize variables as higher order and lower order. This leaves any claims that a particular mode relies on a particular type of variable either vacuous or circular. One could get around this critique by claiming that the mode distinction refers to specifying versus nonspecifying variables rather than to higher- versus lower-order variables. Such a coupling between mode of apprehension and specifying and nonspecifying variables, however, does not fare much better.

First, the all-or-none characterization of perception as related to specifying variables and inference to nonspecifying variables seems to make learning into an all-or-none phenomenon. The catastrophic discovery of specifying information abruptly shifts one from inference to perception. The ability to perceive some property, therefore, cannot improve; it can only come into existence. Although this is possible, our first impression would be that it would have to be an unguided trial-and-error process.

Second, and relatedly, the identification of perception with specifying information renders perception vulnerable to an a posteriori determination that it was merely inference. For example, say one perceives the distance of a falling object by exploiting the specificity of optic-array acceleration to that distance. A falling object accelerates at 9.8 m/s^2, so the optical acceleration of an object falling at eyeheight specifies distance. Surely the discovery that gravity is other than 9.8 m/s^2 in some places should not render any perception of distance that exploited optical acceleration to have been merely inference. Similarly, perception is what it is, and determinations that there are natural or mischievous anomalies with the information do not change the character of perception. Leaving perception ever open to reclassification as inference lets perception slip back into the realm of true (and possibly even false) belief. In short, we argue that changes in variable use, without restrictions to possibly different types of variables, can occur in both modes of apprehension, which leads us back to one of our main questions.

What are the implications of our results for the mode transition hypothesis proposed by Runeson et al. (2000)? One of the difficulties of the mode transition hypothesis is its testability. How can an experimenter tell whether an observer apprehends in an inferential or perceptual mode? With the testability being such a delicate question, probably requiring several converging operations, one would certainly hope that confidence judgments would be one such converging operation. Unfortunately, our results mainly add concern with regard to this use of confidence judgments, thereby indicating that further methods to distinguish inferential and perceptual functioning are needed.

If one assumes that methods to distinguish the modes can be established, then the mode transition theory of learning would have to face a further challenge, namely, to explain the mode transition. That is, however complicated the basis of the observation, it is unsatisfactory just to observe that a transition has occurred. How does a transition occur? Answering this question may prove difficult. Moreover, for theories that entail a mode transition, the challenge is inevitable.

On a positive note, our finding that the decrease in confidence can be explained by mere convergence on specifying variables, with perceivers performing in a direct perceptual mode throughout, suggests that learning results such as those in the experiment presented here can be addressed without resorting to a mode transition. Our expectation is that a lawful account of how people change only in which variables they use might encounter less severe scientific challenges than an account of how a mode transition might occur.

ACKNOWLEDGMENTS

The Netherlands Organization for Scientific Research is gratefully acknowledged for helping to fund this project. David Jacobs was supported by a grant from the Foundation for Behavioural and Educational Sciences of this organization (575-23-011) awarded to Claire Michaels. Preparation of this article was also supported by the National Science Foundation under Grant No. BCS 0339031. The text of this article is based in part on Jacobs, Michaels, Zaal, and Runeson (2001), although all the experiments reported here are new. We thank Isabell Andersson, Peter Juslin, and Till Frank for helpful comments on that version.

REFERENCES

Andersson, I. E. K., Kreegipuu, K., & Runeson, S. (2001, June). *An electrophysiological correlate to the distinction between a direct-perceptual and an inferential mode of apprehension.* Poster session presented at the Eleventh International Conference on Perception and Action, University of Connecticut, Storrs.

Baranski, J. V., & Petrusic, W. M. (1999). Realism of confidence in sensory discrimination. *Perception & Psychophysics, 61,* 1369–1383.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10,* 433–436.

Gibson, J. J. (1966). *The senses considered as perceptual systems.* Boston: Houghton Mifflin.

Gibson, J. J., & Gibson, E. J. (1955). Perceptual learning: Differentiation or enrichment. *Psychological Review, 62,* 32–41.

Gilden, D. L., & Proffitt, D. R. (1989). Understanding collision dynamics. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 372–383.

Gilden, D. L., & Proffitt, D. R. (1994). Heuristic judgment of mass ratio in two-body collisions. *Perception & Psychophysics, 5,* 708–720.

Jacobs, D. M. (2001). *On perceiving, acting, and learning: Toward an ecological approach anchored in convergence.* Utrecht, The Netherlands: Digital Printers Partners.

Jacobs, D. M., & Michaels, C. F. (2002). On the apparent paradox of learning and realism. *Ecological Psychology, 14,* 127–139.

Jacobs, D. M., & Michaels, C. F. (2006). *An ecological theory of information-based perceptual learning.* Manuscript submitted for publication.

Jacobs, D. M., Michaels, C. F., & Runeson, S. (2000). Learning to perceive the relative mass of colliding balls: The effects of ratio-scaling and feedback. *Perception & Psychophysics, 62,* 1332–1340.

Jacobs, D. M., Michaels, C. F., Zaal, F. T. J. M., & Runeson, S. (2001). Developing expertise: Mode change or mere change in variable use? In G. A. Burton & R. C. Schmidt (Eds.), *Studies in Perception and Action VI* (pp. 181–184). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Jacobs, D. M., Runeson, S., & Michaels, C. F. (2001). Learning to visually perceive the relative mass of colliding balls in globally and locally constrained task ecologies. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 1019–1038.

Juslin, P., & Olsson, H. (1997). Thurstonian and Brunswikian origins of uncertainty in judgment: A sampling model of confidence in sensory discrimination. *Psychological Review, 104,* 344–366.

Kreegipuu, K., & Runeson, S. (1999). Becoming competent through the perceiver's eye. In M. A. Grealy & J. A. Thomson (Eds.), *Studies in Perception and Action V: Tenth International Conference on Perception and Action* (pp. 20–23). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

McKee, S. P. (1981). A local mechanism for differential velocity discrimination. *Vision Research, 21,* 491–500.

Michaels, C. F., & Beek, P. J. (1995). On the state of ecological psychology. *Ecological Psychology, 7,* 259–278.

Michaels, C. F., & Carello, C. (1981). *Direct perception*. Englewood Cliffs, NJ: Prentice-Hall.

Michaels, C. F., & de Vries, M. M. (1998). Higher-order and lower-order variables in the visual perception of relative pulling force. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 526–546.

Michaels, C. F., & Lee, W. A. (1996). The organization of multisegmental pulls made by standing humans: II. Submaximal pulls. *Journal of Motor Behavior, 28,* 137–148.

Michaels, C. F., Lee, W. A., & Pai, Y.-C. (1993). The organization of multisegmental pulls made by standing humans: I. Near-maximal pulls. *Journal of Motor Behavior, 25,* 107–124.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442.

Regan, D., & Hamstra, S. J. (1993). Dissociation of discrimination thresholds for time to contact and for rate of angular expansion. *Vision Research, 33,* 447–462.

Regan, D., & Vincent, A. (1995). Visual processing of looming and time to contact throughout the visual field. *Vision Research, 13,* 1845–1857.

Runeson, S. (1995). Support for the cue-heuristic model is based on suboptimal observer performance: Response to Gilden and Proffitt (1994). *Perception & Psychophysics, 5,* 1262–1273.

Runeson, S., & Andersson, I. (2004). On two modes of apprehension. *Ecological Psychology, 14,* 37–44.

Runeson, S., Juslin, P., & Olsson, H. (2000). Visual perception of dynamic properties: Cue heuristics versus direct-perceptual competence. *Psychological Review, 107,* 525–555.

Runeson, S., & Vedeler, D. (1993). The indispensability of precollision kinematics in the visual perception of relative mass. *Perception & Psychophysics, 53,* 617–632.