

Cognitive Dynamics on the Balance-Scale Task

Corinne Zimmerman (czimmer@ilstu.edu)

Steve Croker (s.croker@ilstu.edu)

Department of Psychology, Illinois State University

Campus Box 4620

Normal, IL 61790-4620 USA

Abstract

Our ability to detect patterns and observe regularities is a fundamental part of reasoning and learning. Various theoretical accounts conceptualize induction in different ways. In the current study, we used the balance-scale task and mouse-tracking techniques as a means to explore the cognitive dynamics that underlie inductive pattern recognition. Although we did not replicate an expected interaction between problem difficulty and task instruction because of a procedural constraint, we were still able to examine the implicit cognitive dynamics underlying explicit behavioral responses as a function of problem difficulty. In particular, we found some support that newer, non-algorithmic accounts of cognition on such tasks better characterize performance than rule-based accounts.

Keywords: induction; balance-scale task; mouse tracking; cognitive dynamics

Inductive reasoning is a fundamental component of the scientific process. A key cognitive process underlying scientific reasoning skills (e.g., prediction, hypothesis testing, conceptual change) is the ability to induce patterns in evidence. In order to discover laws and formulate hypotheses, scientists use induction to find regularities in observed phenomena. Sometimes we may search explicitly for rules that govern natural phenomena, but we also use implicit learning processes to discover patterns and rules.

The processes that underlie induction have been conceptualized in different ways. According to information-processing accounts, knowledge structures can be described in terms of symbols, algorithms, and rules. More recent theories in psychology, such as the dynamic systems approach (Thelen & Smith, 1994), have conceptualized learning and change in terms of human behavior being situated and embodied. Dynamic systems theory has provided insight into accounting for change, but largely with respect to perceptual and motor processes. Thus far, it has been difficult for us to determine the best way to apply dynamic systems concepts to higher-level cognitive phenomena such as inductive reasoning. New research techniques and data-analytic methods allow us to apply these theoretical constructs and explanations to higher-level cognition. Although computerized tasks have been used to measure accuracy and reaction time, mouse tracking can be used to examine the temporal, embodied, and dynamic elements of cognition, and to track learning across micro timescales. Moreover, this technique provides a window into the implicit processes that underlie explicit responses.

The Balance-Scale Task

Piaget introduced the balance-scale task as a means of studying cognitive development (Inhelder & Piaget, 1958). Participants make predictions about a two-arm balance scale on which weights are placed on pegs at different distances from the fulcrum. Siegler (1976) demonstrated that changes in performance could be described with respect to different strategies or rules. Several theoretical accounts of change have been proposed. Proponents of classic information-processing theories (e.g., Klahr & Siegler, 1978) argue that we acquire a series of algorithms that we transition through as a function of experience. According to connectionist accounts (e.g., McClelland, 1989), changes in performance are a result of learning statistical relations between problems and correct answers. Zimmerman (1999) derived predictions for performance from an analysis of how neural networks solve these problems and an analysis of the problem space. In particular, these predictions focused on the idea that performance measures (accuracy and response time) should vary as a function of where a problem is located in the problem space (e.g., participants should react faster and more accurately to problems with a high torque difference). Dynamic systems theorists (e.g., van der Maas & Raijmakers, 2009) discuss transitions in terms of patterns of variability and stability; a previously stable behavioral pattern becomes highly variable and undergoes a sudden transition to a new pattern of behavior.

Over the last few decades, many researchers have examined performance on the balance-scale task, typically with the goals of determining (a) what underlying competencies exist at different ages and (b) how performance is affected by factors such as the number of pegs and weights and whether feedback is given immediately following each trial. The balance-scale task has several features that make it ideal for studying cognitive change and learning. It is simple enough that young children understand the task demands, yet the underlying physics principle (torque) is complex enough that adults require much experience with the task before they can induce the underlying physical rules. The problem space has been analyzed in detail (Zimmerman, 1999), and we know much about the context effects that influence performance and facilitate rule discovery (Messer, Pine, & Butler, 2008). Zimmerman and Pretz (2012) examined the effects of instructing participants to make rapid predictions or to think deeply and try to discover the rule underlying the balance

task. Participants who made quick predictions without thinking about potential rules demonstrated superior performance on the most difficult problems, and had better performance on a transfer task.

Kloos and Van Orden (2009) document a large number of context effects, including the number of response options available to participants, the magnitude of the torque discrepancy between the two arms, and whether proprioceptive feedback was given. This strong context dependence, found not just on the balance-scale task, but in most tasks used in psychological research, admits the conclusion that there may be no such thing as a context-free competence that can be uncovered via carefully controlled experiments. Rather, context is “constitutive of cognition” (Riley, Shockley, & Van Orden, 2012, p. 26). Such a conclusion is problematic for traditional cognitive science theories, as they are typically predicated on explaining task behavior in terms of the functioning of specific cognitive components.

However, from a dynamical systems perspective, behavior on the balance-scale task is not the result of participants applying a rule or set rules to each problem, but instead results from the *soft-assembly* (Turvey & Carello, 1981) of the cognitive system, whereby contextual factors constrain the system such that the set of possible responses to a problem reaches a critical state and is collapsed to one response over the time scale of a single trial. On this view, participants are seen as anticipating stimuli rather than just reacting to them, which leads to the prediction that recent trials should affect behavior, a process known as iterativity (Van Geert, 2003). Over the course of the experiment, the cognitive system reorganizes as a result of exposure to multiple trials and feedback on the accuracy of predictions. As the mental representation of the problem changes, participants will change strategies. The emergence of new representations is characterized by an increase in entropy or variability until a critical instability is reached, at which point a rapid decrease in variability will occur as the system undergoes a phase transition and stabilizes into a new organization (Stephen & Dixon, 2009). Each successive organization of the cognitive system produces behavior that can be *described* in terms of the rules described by Siegler (1976), but is not *governed* by these rules.

To examine the predictions made by dynamic systems theory, we need to conduct fine-grained analyses of real-time behavior. Researchers typically use button presses to measure accuracy and reaction times. However, accuracy averaged over a set of trials (e.g., proportion of correct responses) is not a sensitive measure of change. By recording the x , y coordinates of mouse movements we can record participants’ ongoing cognitive dynamics as they decide which response to select.

One aim of the current study is to examine implicit responses on the balance-scale task by examining the temporal dynamics of response choice. The response that a participant selects in any given trial is an *explicit* response. An analysis of the extent to which the trajectory of the

mouse movement deviates towards the distracter informs us of an *implicit* response. A large deviation away from the target response indicates that at least two potential responses are activated prior to the final explicit decision. Smaller deviations indicate that the distracter response may not have become so highly activated on that particular trial. By comparing deviations from the target response on different trials, we can examine whether participants are differentially attracted to the distracter as a function of the difficulty of the problem at hand. We can classify the difficulty of individual balance-scale problems with respect to the torque difference between the two arms and the number of strategies that yield the correct solution. When presented with high torque-difference problems, participants should move towards a single strong *attractor basin* in the state space for that problem (Spivey, 2007). More difficult problems, with low torque difference, can be described in terms of a state space with two strong attractors: both ‘left’ and ‘right’ are strong attractors for the response.

Our second aim was to try to replicate Zimmerman and Pretz’s (2012) finding that the instructions given to participants for how to approach the task interacted with problem difficulty. Participants instructed to try to discover the rule that would allow accurate predictions had an advantage in accuracy for problems that were easy. This advantage was reversed for those instructed to respond quickly without thinking too much; this instruction resulted in greater accuracy on the most difficult problems and significantly better transfer. An analysis of mouse movements will allow us to examine the cognitive dynamics underlying response choices in these two different instructional conditions for problems of varying difficulty at both shorter and longer timescales. Such an analysis can inform a theoretical account of how rapid, unreflective responding differs from more analytic responses, and why there is an advantage for rapid responding for difficult problems.

Method

Stimuli

Stimuli consisted of a set of 144 pictures of balance scales, which were presented in the center of a computer screen (see Figure 1). Each balance scale had 5 pegs on each arm and had between 1 and 5 weights on one of the pegs on each arm. Although there are 625 possible 5-peg, 5-weight balance scales, we used a restricted set of 144 problems for two reasons. First, 64% of the complete problem set contains problems where only weight or distance varies, or where one arm has a greater number of weights at a greater distance. Such problems are very easy to solve, and the latter type are not typically used, even in studies with children. Second, we did not include any scales that balanced (7.8% of the problem space), so that we could restrict the response choices to left or right. The remaining problems constitute a set that allowed a sampling of an approximately equal distribution of problems of three levels of difficulty.

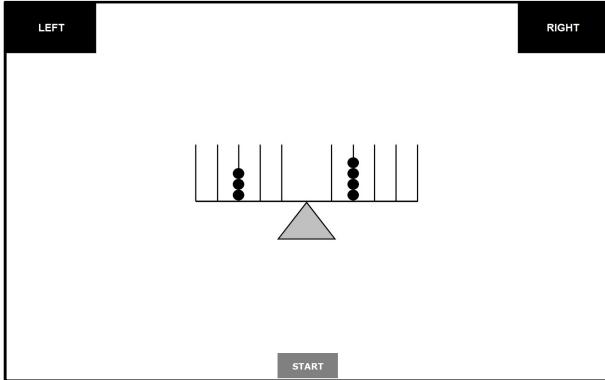


Figure 1: Screenshot of the MouseTracker experiment.

Some examples of balance scales used in the task are illustrated in Figure 2. *Easy* problems are defined as those for which the torque difference between the two arms of the balance scale is between 5 and 15. The example in Figure 2 has a torque difference of 8 (4w x 1d vs. 3w x 4d). *Medium* problems have torque differences less than 5 and can be solved by either adding or multiplying weights and distances. In the example, the torque difference is 2 (4w x 2d vs. 2w x 3d), but a correct prediction also results when adding weights and distances (i.e., 4w+2d vs. 2w+3d). *Hard* problems also have torque differences less than 5, but require multiplication of weight and distance to make accurate predictions. Adding weights and distances will lead to an incorrect prediction.

Procedure

Participants were 154 university students (M age = 20.2; SD = 2.8) from Illinois State University who volunteered to participate for extra course credit. A computerized balance-scale task was used. Participants were asked to make predictions about whether the scale would tip left or tip right. After completing a set of practice trials, participants started each trial by clicking the “start” button at the bottom center of the screen. Immediately after the button click, a pair of response buttons marked “left” and “right” appeared at the upper corners of the screen and an image of a balance scale appeared in the center of the screen. Participants were encouraged to initiate responses as quickly as possible. On trials in which response initiation was over 1000ms, a message was presented to remind participants to start moving the mouse immediately. Stimuli were presented using MouseTracker (Freeman & Ambady, 2010). On trials in which correct predictions were made, there was no feedback (a constraint of the program), and participants proceeded to the next trial. For incorrect predictions, a red X appeared in the center of the screen for 1000ms. Participants completed 4 blocks of 44 trials, and were told they could take a break between blocks. In addition to recording response choices and reaction times, we recorded the streaming x , y coordinates of mouse movements on each trial and normalized each trajectory into 101 time steps in order to compare trials of varying duration.

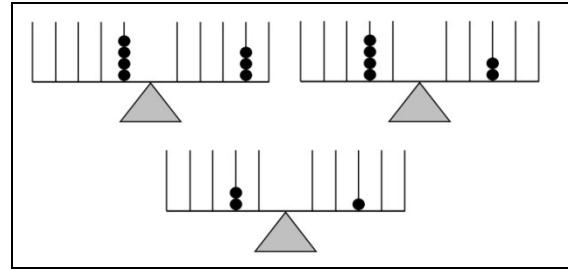


Figure 2: Sample balance scale problems. Top left: easy, top right: medium, and bottom: hard.

Design and Analyses

Each participant was randomly assigned to one of two conditions. In the *prediction* condition ($n = 77$), they were instructed to make predictions only, that is, to make speeded intuitive judgments without thinking too much about each problem. In the *rule-seeking* condition ($n = 77$), they were instructed to make predictions and to attempt to discover the rule that would then allow accurate predictions on every trial.¹ Each block of 44 trials consisted of a randomly presented mix of easy, medium, and hard problems. The design was thus a 2 (instruction condition) \times 3 (difficulty) \times 4 (block) mixed design, with instruction condition as a between-subjects variable and block and problem difficulty as within-subjects variables.

The dependent variables were *accuracy*, *response time*, and *maximum deviation* from an idealized trajectory toward the correct response. The latter is computed by recording the x , y coordinates of mouse movements from the start position to the predictions of “left” or “right” on each trial. This method allows us examine participants’ ongoing cognitive dynamics as they decide which response to select. The extent to which mouse trajectories exhibit curvature away from the correct response is indicative of an evolving response, in which the activations of multiple competing and conflicting implicit responses change over time before resolving into an explicit response. We analyzed accuracy, reaction times, and mouse trajectory curvature using 2 (instruction) \times 3 (problem difficulty) \times 4 (block) mixed ANOVAs. We also conducted a series of linear regression analyses, with torque difference as the predictor. In order to compare the trajectories of correct and incorrect responses, we conducted an additional mixed ANOVA with accuracy (correct vs. incorrect) as a repeated-measures variable.

Results and Discussion

Accuracy

As expected, there was a main effect of problem difficulty on accuracy, $F(2, 304) = 313.92$, $p < .001$, with correct response chosen most frequently for easy problems and least

¹ One potential criticism of this manipulation is that the rule-seeking participants had a higher cognitive load. Our previous research showed equivalent performance for the prediction and prediction-plus-load conditions (Zimmerman, 2010).

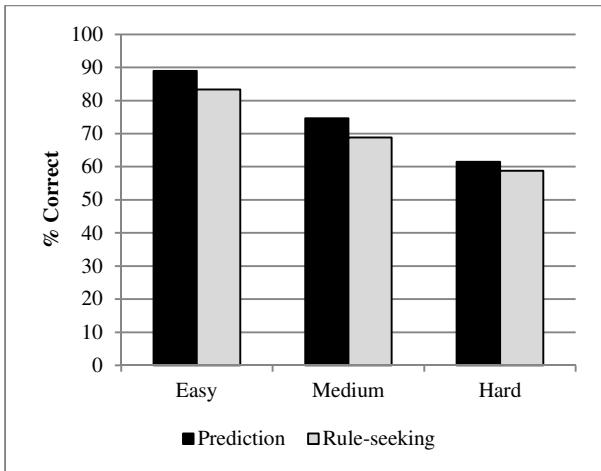


Figure 3: Accuracy of responses (%) by problem difficulty and instruction condition

frequently for hard problems. There was a main effect of instruction, $F(1, 152) = 15.61, p = .01$, with greater accuracy in the prediction condition than in the rule-seeking condition (see Figure 3). A difficulty x block interaction, $F(6, 912) = 2.97, p = .007$, revealed that accuracy for medium problems increased across blocks, but not for easy or hard problems. Torque difference significantly predicted accuracy, $b = 2.16, t(1692) = 23.19, p < .001$, and explained a significant proportion of the variance, $R^2 = .24, F(1, 1692) = 537.85, p < .001$. The expected interaction between instruction and difficulty found by Zimmerman and Pretz (2012; Experiments 1-3) was not replicated. Moreover, the expected advantage for those seeking the rule on easy and medium problems was not evident, with those in the prediction condition having a non-trivial advantage with respect to accuracy for all problem types ($d = .50$).

Response Time

There was an expected main effect of problem difficulty, $F(2, 304) = 22.68, p < .001$, with fastest responses for easy problems (see Figure 4). Response times decreased over time, $F(3, 456) = 20.01, p < .001$. Participants in the prediction condition responded faster than those in the rule-seeking condition, $F(1, 152) = 6.16, p = .014$. A block x instruction interaction, $F(3, 456) = 5.25, p = .001$, revealed a greater decrease in response time for rule-seekers.

The response time pattern for the prediction condition is similar to that found by Zimmerman and Pretz (2012); however, response times here are faster than what they reported for rule-seekers (i.e., means of 3.9 to 4.6 sec across three experiments). The pattern here is consistent with participants who found the task too challenging and gave up looking for the underlying rule (see Zimmerman & Pretz, 2012; Experiment 3). Taken together, this pattern suggests that our instructional manipulation did not have the expected effect. Alternatively, the constraint of needing to start a mouse movement quickly may have made it difficult for rule-seekers to follow our instructions to try to discover

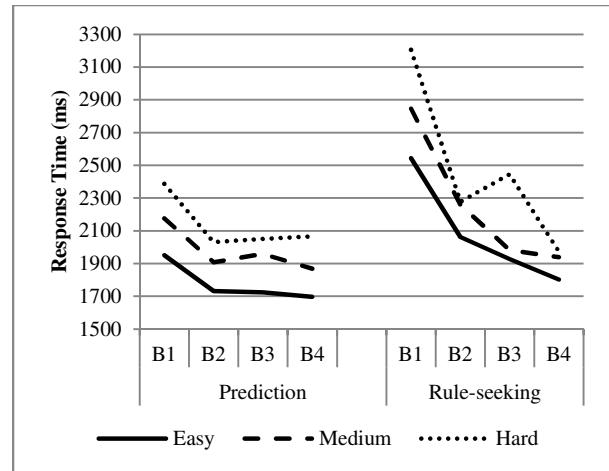


Figure 4: Response times (ms) by problem difficulty, block (B1-B4), and instruction condition.

the rule. Ultimately, this constraint made the task difficult for rule seekers; by the last block, response times are not that different from those making predictions alone.

Maximum Deviation

Consistent with our suspicion that our instructional manipulation did not work as expected, there were no effects of instruction condition or interactions, so further analyses on maximum deviation and trajectory data combine data from the two groups. There were main effects of problem difficulty, $F(2, 304) = 67.70, p < .001$, and block, $F(3, 456) = 17.32, p < .001$. On easy problems there was a smaller deviation towards the distracter than on medium and hard problems, and there was less deviation towards the distracter during the first block than on blocks 2 to 4 (see Figure 5). A regression analysis revealed that torque difference significantly predicted maximum deviation, $b = -.02, t(1691) = -12.59, p < .001$, and explained a significant proportion of the variance, $R^2 = .09, F(1, 1691) = 158.57, p < .001$.

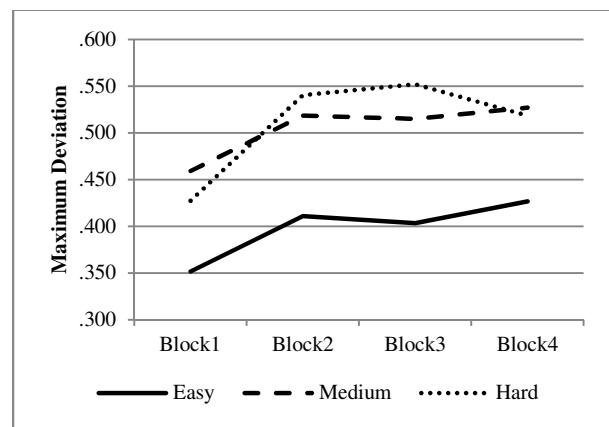


Figure 5: Maximum deviations from an idealized trajectory for the three problem types over the four blocks.

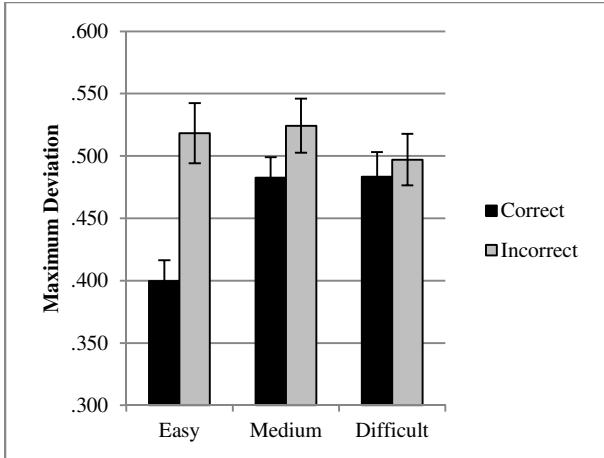


Figure 6: Maximum deviations of the idealized trajectory for the three problem types on correct and incorrect trials.

Comparison of Correct and Incorrect Responses

We compared mouse trajectories and their maximum deviations for correct and incorrect responses using a mixed 2 (accuracy) x 3 (difficulty) x 4 (block) ANOVA. There was a main effect of accuracy, $F(1, 84) = 20.60, p < .001$, and an accuracy x difficulty interaction, $F(2, 168) = 8.57, p < .001$ (see Figure 6). Overall, participants' mouse movements demonstrated greater attraction to the unselected response before making a response on incorrect trials than on correct trials. The correct response thus had greater attraction on trials where the incorrect response ended up being chosen compared to distracter response on correct trials. This effect was exaggerated for easy problems.

We calculated the average x -coordinates for each of the 101 normalized timesteps for each level of problem difficulty on correct and incorrect trials. As there was no effect of block on maximum deviation scores, we combined trials from the four blocks. We subtracted the coordinate values for the incorrect trials from the coordinate values of the correct trials at each timestep to provide a series of difference scores. Figure 7 illustrates the differences between correct and incorrect trials for each of the three problem difficulties. Positive values indicate where mouse positions deviated more towards the unselected response on incorrect trials than on correct trials and negative values show where correct trials deviated more towards the unselected response.

During easy problem trials, there was less deviation towards the unselected response on correct trials than incorrect trials, whereas there was little difference between correct and incorrect trials for hard problems. On medium difficulty trials, there is an early deviation toward the correct response on correct trials, relative to incorrect trials, followed by a deviation toward the distracter response later on, around timestep 68.

The timecourse of mouse trajectories follows a different pattern for each level of problem difficulty. The pattern for easy problems indicates that features of the problem rapidly

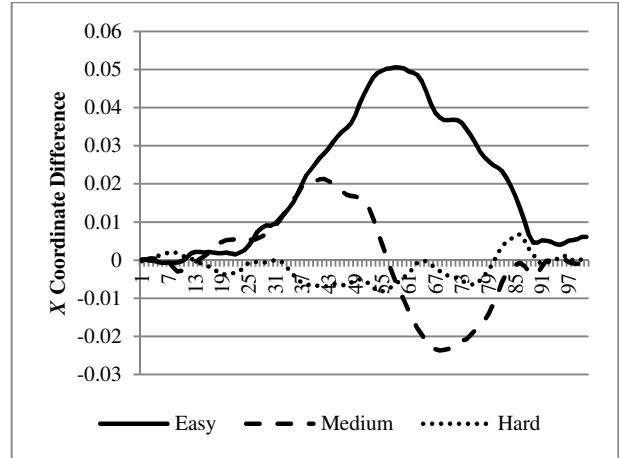


Figure 7: X -coordinate difference between correct and incorrect responses at each timestep. Positive values indicate greater deviation toward the unselected response on incorrect trials.

led to a greater activation of the correct response. The patterns for medium problems look similar to those for easy problems early on, but later in the trial, there is greater relative movement toward the distracter response on correct trials, suggesting that the unselected response became more highly activated at different time points for correct and incorrect responses.

General Discussion

Our goal was to use mouse-tracking techniques to provide insight into the implicit cognitive dynamics underlying explicit response choices in a well-studied task. Zimmerman and Pretz (2012) reported a robust interaction between difficulty and instruction; such context effects make the balance-scale task a strong candidate for examining the ways in which contextual constraints operate at different levels to affect behavior, and how both motor and strategic behaviors evolve over time in response to those constraints.

Unexpectedly, participants in the prediction condition made faster and more accurate responses across the board, relative to those in the rule-seeking condition. The time constraint we placed on participants to ensure that we captured online processing (i.e., to start mouse movements within 1000ms of clicking "start") was consistent across conditions. However, the effect on those seeking the rule was profound. Their accuracy was lower than expected for easy and medium problems, and their response time patterns indicate that they gave up very early in the experiment. In essence, we may have turned every problem into a hard problem for these participants.

Our results show that participants learned something about the medium difficulty problems over the course of the experiment, as evident by the interaction between block and difficulty. However, there was no increase in accuracy for easy or hard problems, and the maximum deviations did not decrease over time; rather, they increased for all problem

types over the first two blocks then remained stable. These data show that, even after 144 trials, participants were not able to reliably use feedback to make accurate predictions (or to discover the torque rule), and that the relative activation strengths (i.e., attraction to the two response choices) remained stable over the experiment.

If participants were using an addition rule, we would expect to see greater differences between the medium and hard problems, and if they used a multiplication (torque) rule, we would expect better (but slower) overall performance. In fact, all dependent measures – accuracy, response time, and maximum deviation – are predicted by the torque difference between the two arms of the balance scale, suggesting that predictions are not made using explicit mathematical rules, but rather by a more implicit weighting of the elements of each problem. However, medium and hard problems have the same low torque differences. Thus, torque difference alone cannot account for the differences in accuracy between the two. According to a dynamic systems account, our pattern of results can be explained in terms of multiple constraints. The exact features of every problem, including torque difference and the way in which that difference is instantiated across the two arms in terms of distance and weight, all contribute to an evolving response.

When we compared x -coordinate differences between correct and incorrect responses at each timestep, we found differential attraction to the distracter response as function of problem difficulty. Relative to incorrect trials, performance on correct trials indicated that easy problems led to a strong initial activation of the correct response; any competition from the distracter was soon inhibited. For medium problems, the patterns were more complex suggesting that competing representations achieved partial activation at various points during a trial.

In conclusion, our data partially support a dynamic systems account of behavior on the balance-scale task. Our data do not support the claim that participants applied rules; response choices evolved over single trials as a function of problem features. However, we did not observe patterns of variability and stability over time; there were no indications of rapid transitions to new states. In future studies, rule discovery can be induced by providing difficult, yet highly diagnostic problems (Zimmerman & Pretz, 2012). This type of learning environment will allow an analysis of the behavioral patterns that precede rule discovery, and inform our understanding of how constraints operating at nested spatiotemporal scales yield a phase transition to a new strategy.

References

Freeman, J. B., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods*, 42, 226-241.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York, NY: Basic Books.

Klahr, D., & Siegler, R. S. (1978). The representation of children's knowledge. In H. W. Reese & L. P. Lipsitt (Eds.), *Advances in child development and behavior* (Vol. 12). New York, NY: Academic Press.

Kloos, H. & Van Orden, G. C. (2009). Soft-assembled mechanisms for the unified theory. In J.P. Spencer, M. Thomas, & J. McClelland (Eds.), *Toward a new grand theory of development: Connectionism and dynamics systems theory reconsidered*. Oxford, UK: Oxford University Press.

McClelland, J. L. (1989). Parallel distributed processing: implications for cognition and development. In R. G. M. Morris (Ed.), *Parallel distributed processing: Implications for psychology and neurobiology*. Oxford, UK: Oxford University Press.

Messer, D. J., Pine, K.J. & Butler, (2008). Children's behaviour and cognitions across different balance tasks. *Learning and Instruction*, 18, 42-53

Riley, M. A., Shockley, K., & Van Orden, G. (2012). Learning from the body about the mind. *Topics in Cognitive Sciences*, 4, 21-34.

Siegler, R. S. (1976). Three aspects of cognitive development. *Cognitive Psychology*, 8, 481-520

Stephen, D. G., & Dixon, J. A. (2009). The self-organization of insight: Entropy and power laws in problem solving. *Journal of Problem Solving*, 2, 72-101.

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: Bradford Books/MIT Press.

Turvey, M. T., & Carello, C. (1981). Cognition: The view from ecological realism. *Cognition*, 10, 313-321.

Van der Maas, H. L. J., & Raijmakers, M. E. J. (2009). Transitions in cognitive development: Prospects and limitations of a neural dynamic approach. In J.P. Spencer, M. Thomas, & J. McClelland (Eds.), *Toward a new grand theory of development: Connectionism and dynamics systems theory reconsidered*. Oxford, UK: Oxford University Press.

Van Geert, P. (2003). Dynamic systems approaches and modeling of developmental processes. In J. Valsiner and K. J. Conolly (Eds.), *Handbook of developmental psychology*. London, UK: Sage.

Zimmerman, C. (2010, June). The interaction of implicit vs. explicit processing and problem complexity in scientific reasoning. Presented at the Second Purdue Symposium on Psychological Sciences (Psychology of Science: Implicit and Explicit Reasoning). West Lafayette, IN.

Zimmerman, C. (1999). *A network interpretation approach to the balance scale task* (Unpublished doctoral dissertation). University of Alberta.

Zimmerman, C., & Pretz, J. (2012). The interaction of implicit versus explicit processing and problem difficulty in a scientific discovery task. In R. Proctor and J. Capaldi (Eds.), *The psychology of science: Implicit and explicit processes*. Oxford, UK: Oxford University Press.