Belief revisions in the cognitive development of proportional reasoning: dynamics, statistical models and computational models.

Maartje Raijmakers, Brenda Jansen & Han van der Maas
Department of Psychology, FMG, University of Amsterdam

In the balance scale task, children are asked to predict the movement of a balance scale. Equally heavy weights can be placed on the arms of the scale, at equally spaced distances. Figure 1 shows a graphical display of a balance scale problem. Blocks underneath the arms prevent the scale from tipping. The task assesses proportional reasoning as it requires the understanding of the multiplicative relation between the dimensions.

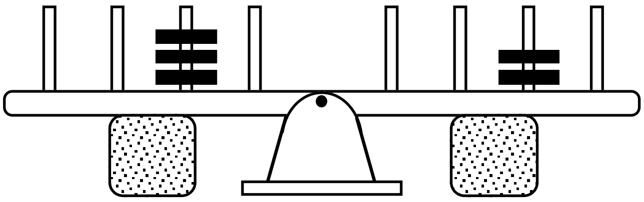

Figure 1. Example of a balance scale item.

Inhelder & Piaget (1958) formulated a developmental course of performance on the task, which is characterized by several qualitatively distinct, increasingly complex, stages. The stages range from considering the number of weights only to comparing the products of weight and distance of both sides of the scale (torque-rule). However, stages can not be observed directly. In this paper we will give a brief overview of children's beliefs about the balance scale, that is, the strategies they apply in solving balance scale problems. Moreover, we will discuss the dynamics of changing beliefs from simple strategies to more complex strategies on a developmental time scale. Finally, we briefly discuss computational models of balance scale development[†].

Rule assessment

Siegler (1981) designed the Rule Assessment Methodology (RAM) to assess children's performance on the balance scale task in a nonverbal way. RAM involves a careful selection of item types that elicit specific response patterns. These patterns can be linked to rules, comparable to the original stages. Complexity of rules increases with development as each rule consists of the steps of the preceding rule, extended with one or more extra steps. Rule I is the simplest rule as it involves only one step, which consists of comparing the numbers of weights. Participants who use Rule I decide that the scale will tip to the side with the largest number of weights when the numbers are unequal and that the scale will remain in balance when the numbers are equal. Participants who use Rule II also compare the numbers of weights on all items and decide that the scale will tip to the side with the larger number of weights when the numbers differ. However, when the numbers are equal, they also compare the distances at which the weights are placed. Rule III is more complex than Rule II because it contains two additional steps. Rule III-users derive the correct response on simple items by comparing the numbers of weights in the first step and comparing the distances at which the weights are placed in the second step. The first additional step is performed if both the weights and the distances are unequal and includes determining whether the dimensions agree (i.e., whether the greater weight is on the same side as the greater distance). If the dimensions conflict, the second additional step is performed. It implies "muddling through" or guessing. Although Rule IV contains the same number of steps as Rule III, Rule IV is more complex because of the

[†] A more elaborated overview is given in Van der Maas, Jansen & Raijmakers (in press).

complexity of the last step. It includes executing the torque-rule on items with conflicting dimensions.

To assess the use of these rules, Siegler designed (1976, 1981) six item types. On simple-balance items ("sb"), both arms of the scale hold the same number of weights, equidistant from the fulcrum. On simple-weight items ("sw"), the arms contain unequal numbers of weights, equidistant from the fulcrum. Simple-distance items ("sd") involve equal numbers of weights, placed at different distances from the fulcrum. On conflict items, one arm contains a greater number of weights, whereas the weights on the other arm are placed at a greater distance. The scale tips to the side with the larger number of weights on conflict-weight items ("cw"), tips to the side with the weights placed at the greater distance on conflict-distance items ("cd") and remains in balance on conflict-balance items ("cb"). Siegler's rule assessment methodology (RAM) consists of matching response patterns on a set of balance scale items of individual subjects with response patterns derived from the expected rules.

Although RAM seems an elegant nonverbal method to assess rules, there is a risk that the rules are an artifact of RAM (see Strauss & Levin in Siegler, 1981). First, RAM does not include any statistical fit measures to evaluate the fit of the rule classification to the empirical data. Second, the criterion of matching responses that is used to classify response patterns (e.g., 20 out of 24 responses correspond to a rule) is chosen arbitrarily and does not lend itself to statistical testing. Finally, with RAM, only the rules that are known a priori can be detected. A statistical approach to the classification of response patterns to rules is required to solve these problems.

Latent class analysis (LCA; e.g., Rindskopf, 1987) is a statistical technique to determine the number of rules, and their response patterns. In these response patterns, children's error processes can be modeled. A child's response may be inconsistent with the rule that the child is using because of carelessness, for instance (Rindskopf, 1987). The deviation from the expected response pattern can be accommodated in a latent class model. Hence, the criterion used to classify children into classes (i.e., to rules) is based on a statistical criterion, rather than an arbitrary criterion. Another advantage is that LCA does not require information on the content of rules but detects clusters of response patterns in the data, which can later be interpreted as (alternative) rules. Most importantly, LCA can falsify the hypothesis concerning rule use, because a latent class model, associated with rule use, can be tested statistically.

LCA divides the population in a finite number of latent classes. Within each class, the manifest variables are assumed to be statistically independent. A latent class model consists of unconditional probabilities, representing the proportions of the classes, and conditional probabilities, representing the probabilities of giving the response "left side down", "right side down", or "balance" on a particular balance scale item, within a given latent class.

As an example, we present the results of the analyses of an empirical data set that features in Jansen and van der Maas (2002), and consists of the responses of 805 subjects to a test of 25 balance scale items. Subjects' ages ranged from 5 to 19 years. Subjects did not receive feedback on their responses. Analyses of individual item types showed that responses to items of one type were homogeneous. It then can be concluded that an item is representative for other items of the type. We present a latent class model of responses to a combination of item types. Only with a combination of item types can all Siegler's rules be detected. We selected the responses to two sd-items, two cw-items and two cb-items.

A seven-class model was selected as this was the simplest, best fitting model according to a formal model selection procedure. Table 1 shows the selected model. Table 1 only contains the conditional probabilities of one item of each type, to save space, and because the probabilities were quite similar. The LCA provides evidence for the application of seven different rules as is indicated in the table.

Table 1. Latent class model of responses to a combination of items

		Conditional probabilities									
		distance 4			conflict-weight 4			conflict-balance 4			
t	*p(t)*	left	balance	right	left	balance	right	left	balance	right	interpretation
1	.27	*.00*	.96	.04	*.96*	.01	.02	.00	*.00*	1.00	Rule I
2	.15	*.70*	.21	.09	*.87*	.13	.00	.00	*.00*	1.00	Rule II
3	.09	*.98*	.02	.00	*.44*	.23	.34	.66	*.07*	.27	Rule III
4	.11	*1.00*	.00	.00	*.82*	.11	.07	.00	*1.00*	.00	Rule IV
5	.18	*.97*	.01	.02	*.03*	.95	.02	.10	*.90*	.00	compensation-rule
6	.02	*.17*	.17	.67	*.92*	.08	.00	.00	*.08*	.92	SDD
7	.17	*.93*	.07	.00	*.24*	.66	.10	.04	*.39*	.57	Rule III/comp

Note. *t* = latent class, *p(t)* = proportion of latent class *t*. The probabilities for the correct response are printed in italics. For each type, item 4 is shown. For the conflict-balance items, the right side has the larger number of weights. Rule III/comp is a combination of Rule III and the compensation-rule. SDD is the Smallest Distance Down Rule. Adapted from Jansen & van der Maas (2002).

Conclusion The results of LCA suggest that children actively employ rules in solving balance scale problems. An advantage of the application of LCA is that we were able to detect alternative rules, like the compensation-rule. Contrary to the rule classifications of RAM, the latent classes reflect the structure in the observed data and arise independently from the rules postulated in a theory.

LCA models were also applied to response data generated by a connectionist PDP model of balance scale learning (McClelland, 1989; Jansen & vanderMaas, 1997). The difficulties of interpreting the LCA models for the PDP data set were striking. Although fitting latent class models was possible, the results were often unstable (i.e., dependent on starting values). Furthermore, the response patterns of the classes did not show consistency within item types, and hardly matched Siegler's rules or any plausible alternative rule. Probably, no rules underlie the response patterns of this connectionist model, in spite of the results that were obtained by applying RAM (McClelland, 1989). These results are in accordance with the results of Raijmakers, van Koten, and Molenaar (1996). It seems that the RAM is too liberal and may falsely suggest the presence of rules. Strauss and Levin's (in Siegler, 1981) criticism that rules are an artifact of the methodology seems correct. However, application of the statistical technique of LCA allows for the falsification of the rule theory. We conclude that it is possible to reveal the rules children use by combining RAM with LCA, even without asking children to clarify their responses.

Reaction time predictions

The above presented rule assessment methodology is solely based on the judgements of balance scale items. However, in addition to this categorical data also response times (RTs) could provide information about rules children apply. Each rule is hypothesized to consist of consecutively executed steps. Each step is indicated with a parameter that indicates the duration of executing the step. We hypothesized that the time involved with solving a balance scale item equals the sum of the duration of the steps that are completed. This hypothesis leads to a number of predictions and findings (van der Maas & Jansen, 2003) of which we present two below. The predictions concerning RTs are tested by fitting regression models to RT data. A complete analysis of these data can be found in van der Maas and Jansen (2003).

Rule II. Siegler's original rule model predicts that Rule II users compare the distances at which the weights are placed only when the numbers are equal. However, from the RT models it was concluded that Rule II-users compared distances at any item, whether the numbers of weight were equal or not. Although the response patterns showed that Rule II-users always answered that the scale would tip to the side with the larger number of weights, their RTs showed that they did consider the distance dimension when the numbers of weights differed.

Weight-distance items The last 6 items of the balance scale test in the study reported in (vanderMaas & Jansen, 2003) were so-called weight-distance items. Weight-distance items are items in which the larger number of weights is on the side of the larger distance. All subjects should solve these simple items easily. Users of Rule I and II are expected to take into consideration the weight dimension only, whereas users of higher rules are expected to look at the distance dimension too, and even check whether the larger weight is on the side with the larger distance. This results in the provoking prediction that older and more advanced subjects are expected to respond more slowly to these easy items. This prediction was confirmed by the data. Mean RT for Rule I and II was 2.63 s, whereas the mean RT for Rule III, the compensation-rule and Rule IV was 3.96 s.

Transitions from Rule I to Rule II

Having established a clear idea of children's strategies for solving balance scale problems, it is time to study the dynamics of changes in balance scale behavior, that is belief revisions. A central and recurrent issue in developmental psychology is whether the transitions between the stages proceed either continuously or discontinuously. For a long time, criteria to distinguish the two types of development were lacking (Brainerd, 1978). Recently, van der Maas and Molenaar (1992) proposed to use catastrophe theory to test for discontinuities. Catastrophe theory (Thom, 1975) is a general mathematical theory of transitions, which are defined as large sudden jumps as function of small continuous changes in independent variables. The most popular model of catastrophe theory is the so-called cusp model.

Cusp-like processes are characterized by a number of phenomena, so-called catastrophe flags, among which are the sudden jump, bimodality, inaccessible region, critical slowing down, hysteresis and divergence (Gilmore, 1981). The last two are convincing indicators of discontinuity. Hysteresis occurs when the sudden jump position depends on the direction of change in the normal variable. Here, we will discuss evidence of two catastrophe flags in the transition from Rule I to Rule II.

Bimodality Bimodality is a necessary, although not sufficient, indicator of transitions. Bimodality is clearly shown in the distribution of sum scores of sets of distance items. The distribution shows a strong distinction between two modes of responses: all incorrect (Rule I) or all correct (Rule II or a more complex rule) (Raijmakers, van Koten, & Molenaar,1996; Jansen & van der Maas, 1997, 2001; see also Section 2).

Hysteresis According to Siegler and Chen (1998), the main difference between Rules I and II is the ability to encode distance. We emphasized the saliency of the distance dimension to increase the awareness of the distance dimension of Rule I-users. Increase of awareness may result in encoding the dimension and, eventually, in a switch to using Rule II. Saliency was emphasized by increasing the distance difference between the two sides of the scale in distance items. Hence, we chose the difference in distance between the weights on the left and the distance of the weights on the right side of the scale as the independent variable that may induce hysteresis. In an empirical study with Rule I users and Rule II users, we showed, in various ways, that a small number of subjects show hysteresis in their responses, which could not be attributed to chance (Jansen & van der Maas, 2001).

Rule development

Consistency of responding during the administration of a balance scale test was analyzed to study whether children who use Rule II, Rule III, Rule IV, and the compensation-rule do so as consistently as children who use Rule I. It appeared that both the use of Rule I and Rule IV was quite consistent during the test, but there seemed to be much switching between Rule II, Rule III, the compensation-rule, and the combination of Rule III and the compensation-rule. The progress of children who start by using Rule II to using Rule III or the compensation-rule can be explained by spontaneous learning. Children who use Rule II already know that the distance dimension can be important. Their RTs indicate that they consider it at any item, but their responses indicate that they only incorporate it in their strategy when the weights are equal. Merely presenting these children with balance scale items may sensitize them to the distance dimension and may convince them of the importance of the distance dimension. However, they do not know how to combine the two dimensions yet. This is consistent with the interpretation of Rule III. A different mechanism than learning must be responsible for regressing from Rule III or the compensation-rule to Rule II. Finally, many children switched between Rule III, the combination of Rule III and the compensation-rule, and the compensation-rule. This finding supports the hypothesis that children who use Rule III sample from an ensemble of strategies and that switching between these strategies is inherent to Rule III.

Computational models of proportional reasoning

Several computational models have been developed to capture the developmental phenomena associated with the balance scale task. These models, which originate in different computational traditions, attempt to explain phenomena of development in proportional reasoning. So far, none of these models has been able to explain all empirical data. Above, we already mentioned McClellands (1989) feedforward neural network with back-propagation learning. Applying the RAM, McClelland showed that the model captures Siegler's sequence of increasingly complex rules. However, statistical tests of rule use by LCA shows no convincing evidence that the behavior of the model can be described as using rules (Jansen & van der Maas, 1997). Moreover, the transition from Rule I to Rule II shows indications of discontinuity, such as sudden jumps and bimodality (Raijmakers et al., 1996).

Recently, van Rijn et al. (2003) proposed a computational model that is implemented in ACT-R (Anderson & Lebiere, 1998). One off-spin of such an attempt is that one is forced to be very explicit about the mechanisms, constraints and assumptions underlying the model. The model consists of ACT-R mechanisms, general knowledge, assumptions about task specific properties and the role of memory capacity constraints. We were able to explain the main empirical phenomena with this model. As yet, RTs are not evaluated in the ACT-R model.

Conclusion

Research on the development of balance scale reasoning involves the assessment of beliefs and belief revisions in children on a developmental time scale. Verbal justifications of responses on balance scale items involve all sort of methodological problems (Jansen, 2001). Nonverbal assessment of children's strategies with the statistical technique of Latent class analysis shows reliably which strategies are applied at different ages. Reaction time analysis provides additional information about the cognitive processes that occur in different rule users.

Belief revisions are investigated by studying transitions between rules, which children make during development. Not all rule use is stable, even when observed within one session without any feedback on the given responses. Rule I, Rule II, and Rule IV are relatively stable and development seems to occur with a sudden transition between Rule I and Rule II. In contrast, children easily switch between Rule III and the compensation rules. The rich set of empirical results makes the balance scale task an interesting benchmark task for computational models of cognitive development.

References

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Brainerd, C.J. (1978). The Stage Question in Cognitive-Developmental Theory. *The Behavioral and Brain Sciences, 2*, 173-213.

Gilmore, R. (1981). *Catastrophe theory for scientists and engineers*. New York: Wiley.

Jansen, B. R. J. (2001). Development of reasoning on the balance scale task: psychometric assessment of cognitive strategies. PhD Thesis at the department of Psychology of the University of Amsterdam.

Jansen, B. R. J., & Van der Maas, H. L. J. (1997). Statistical test of the rule assessment methodology by latent class analysis. *Developmental Review*, *17*, 321–357.

Jansen, B. R. J., & Van der Maas, H. L. J. (2001). Evidence for the phase transition from Rule I to Rule II on the balance scale task. *Developmental Review*, *21*, 450-494.

Jansen, B. R. J., & Van der Maas, H. L. J. (2002). Development of children's rule use on the balance scale task. *Journal of Experimental Child Psychology*, *81*(4), 383–416.

McClelland, J. L. (1989). Parallel distributed processing: Implications for cognition and development. In M. G. M. Morris (Ed.), *Parallel distributed processing, implications for psychology and neurobiology*, pp. 8-45. Oxford: Clarendon Press.

Raijmakers, M. E. J., Van Koten, S., & Molenaar, P. C. M. (1996). On the validity of simulating stagewise development by means of PDP networks: Application of catastrophe analysis and an experimental test of rule-like network performance. *Cognitive Science*, *20*(1), 101–136.

Rindskopf, D. (1987). Using latent class analysis to test developmental models. *Developmental Review, 7*, 66-85.

Siegler, R. S. (1981). Developmental sequences within and between concepts. *Monographs of the Society for Research in Child Development*, *46*(2), 1–74. (With commentary and reply from Sidney Strauss and Iris Levin)

Thom, R. (1975). *Structural stability and morphogenesis: An outline of a general theory of models*. Reading, MA: Benjamin.

Van der Maas, H. L. J. & Jansen, B. R. J. (2003). What response times tell of children's behavior on the balance scale task. *Journal of Experimental Child Psycholog, 85, 141-177*

van der Maas, H. L. J., Jansen, B. R. J. & Raijmakers, M. E. J. (in press). Developmental patterns in proportional reasoning. In A. Demetriou & A. Raftopoulos (Eds.), Cognitive Developmental Change Theories, Models and Measurement. Cambridge University Press.

Van der Maas, H. L. J., & Molenaar, P. C. M. (1992). Stagewise cognitive development: An application of catastrophe theory. *Psychological Review, 99***,** 395–417.

Van Rijn, H., van Someren, M., & van der Maas, H. L. J. (2003). Modeling developmental transitions on the balance scale task. *Cognitive Science, 27*(2), 227-257.