ORIGINAL PAPER

Rule transition on the balance scale task: a case study in belief change

Brenda R. J. Jansen · Maartje E. J. Raijmakers · Ingmar Visser

Abstract For various domains in proportional reasoning cognitive development is characterized as a progression through a series of increasingly complex rules. A multiplicative relationship between two task features, such as weight and distance information of blocks placed at both sides of the fulcrum of a balance scale, appears difficult to discover. During development, children change their beliefs about the balance scale several times: from a focus on the weight dimension (Rule I) to occasionally considering the distance dimension (Rule II), guessing (Rule III), and applying multiplication (Rule IV; Siegler, 1981). Because of the detailed empirical findings the balance scale task has become a benchmark task for computational models of proportional reasoning.

In this article, we present a large empirical study ($N = 420$) of which the findings provide a challenge for computational models. The effect of feedback and the effect of individually adapted training items on rule transition were tested for children using Rule I or Rule II. Presenting adapted training items initiates belief revision for Rule I but not for Rule II. The experience of making mistakes (by providing feedback) induces a change for both Rule I and Rule II. However, a delayed posttest shows that these changes are preserved after 2 weeks only for children using Rule I. We conclude that the transition from Rule I to Rule II differs from the transition from Rule II to a more complex rule. Concerning these empirical findings, we will review performance of computational models and the implications for a future belief revision model.

Keywords Balance scale task · Computational models · Cognitive development

B. R. J. Jansen (✉) · M. E. J. Raijmakers · I. Visser
Department of Developmental Psychology, Faculty of Psychology,
University of Amsterdam, Roetersstraat 15,
1018 WB Amsterdam, The Netherlands
e-mail: b.r.j.jansen@uva.nl

It is one Thing, to show a Man that he is in an Error, and another, to put him in possession of Truth.

John Locke

1 Introduction

Young children belief that the amount of lemonade in a glass solely depends on the height of the lemonade in the glass. If we present them a pair of glasses filled with lemonade, a wide one and a narrow one, they will indicate that the glass with the highest level of lemonade contains more lemonade, independent of the width of the glasses. When and how do children come to understand that not just the height determines the volume of liquid in a glass? Do they change their belief as they encounter more and more glasses, of different heights and widths? Or do they change their belief after drinking from a tall, narrow glass, experiencing that the lemonade was finished rather quickly?

The above example on compensation refers to the process of rule transition, which is an important object of study in cognitive development. To our view it is also an interesting domain for belief revision formalisms. For now, beliefs are the set of internal representations a cognitive agent considers reliable for further reasoning and direct action in the world (Paglieri, 2004). In later sections we define more precisely what we mean by rules and rule transition and its relationship to philosophical conceptions of belief change.

During development, children continuously change their internal representations of phenomena in the world around them. We study two possible answers to the empirical question how children change their beliefs. First, children may change their belief because they start to perceive uninspected aspects of the phenomenon at hand (in the example above: children encounter differently shaped glasses and start to notice the width dimension). Second, children may change their belief as they acquire new information that stands in contrast to their old beliefs (in the example above: drinking an unexpectedly small amount of lemonade from a tall, but narrow glass).

In this article, we do not focus on compensation but on the domain of proportional reasoning to study these two possibilities. Proportional reasoning requires understanding of the multiplicative relation between two dimensions of a task. It demands the identification of the relevant dimensions and acknowledgement of the mutual relation between these dimensions. This article focuses on the most famous proportional reasoning task in developmental psychology (Siegler & Chen, 2002), the so-called balance scale task, in which 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. Various configurations of weights and distances are possible, resulting in item types of different levels of complexity. A blocking pin prevents the scale from tipping and children need to indicate the movement of the scale when the blocking pin is removed. The two dimensions that need to be integrated are the numbers of weights on each side of the scale and the distances from the fulcrum at which the weights are placed. Usually, children do not receive feedback at this task. Figure 1 shows a graphical display of a balance scale problem.
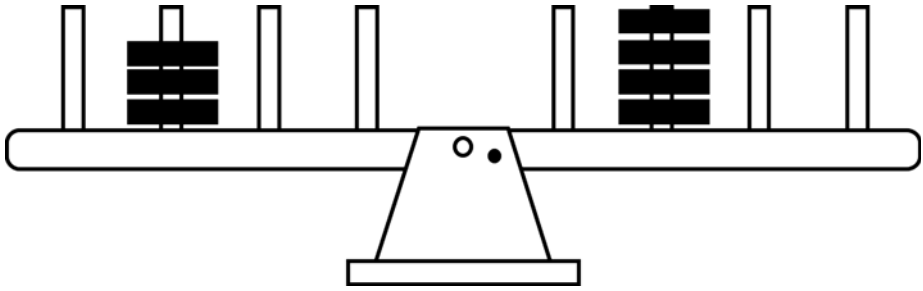
Fig. 1 Graphical display of a balance scale problem

1.1 Rules on the balance scale task

Inhelder and Piaget (1958) formalized a developmental course of performance on the balance scale task. This developmental course is characterized by several qualitatively distinct stages. Siegler (1976, 1981) applied the rule assessment methodology to the Piagetian theory of the balance scale task. The basic assumption underlying this methodology is that cognitive development can be characterized as the acquisition of increasingly complex rules.

Siegler (1981) hypothesized that children use one of four rules when solving proportional reasoning tasks. Children, who use *Rule I*, only consider the dominant dimension of the task. On the balance scale task, they compare the numbers of weights on both sides of the scale, i.e., number of weights is the dominant dimension. These children respond that the scale will remain in balance if the numbers of weights are equal and respond that the scale will tip to the side with the largest number of weights if the numbers differ. The behavior of children who use *Rule II* is more complex. These children also consider the subordinate dimension, distance. However, they only consider this dimension when the values on the dominant dimension are equal. In case of the balance scale task they first consider the numbers of weights. They decide that the scale will tip to the side with the largest number of weights if the numbers of weights are unequal. Only in case of equal numbers of weights, they consider the distances at which the weights are placed. Children decide that the scale will remain in balance when the distances are equal and that the scale will tip to the side with the largest distance when the distances differ. The next complex behavior is referred to as resulting from the use of *Rule III*. The behavior involves considering both dimensions. Subjects resort to guessing when the dimensions conflict. The most complex rule is to calculate and compare the products of weights and distances. Siegler defines this behavior as *Rule IV*.

Siegler (1981) identified these rules after interviewing children about their solution strategies in order to design a nonverbal method of assessing rule use. He designed particular item types that elicit specific responses and make it possible to distinguish the cognitive rules. On *balance items*, the numbers of weights and the distances of the piles of weights from the fulcrum are equal. The numbers of weights differ, but the distances are equal on *weight* items. On *distance* items, the distances from the fulcrum differ, but the numbers of weights are equal. On conflict items, both the numbers of weights and the distances differ and the dimensions conflict. The balance tips to the side with the larger number of weights on *conflict-weight* items, tips to the side with

the larger distance on *conflict-distance* items, and stays in balance on *conflict-balance* items. A balance scale test often consists of several items of each item type, resulting in a total of about 25 items.

Children who use Rule I provide correct responses to balance, weight, and conflict-weight items, whereas children who use Rule II also respond correctly to distance items. Rule III results in correct responses to balance, weight, and distance items and to guessing on conflict items, whereas Rule IV-users respond correctly to all item types.

1.2 Aims and overview

The aim of the empirical research in this article is to investigate the transitions between the rules as insight in rule transitions may be of importance for understanding the processes of belief change. We investigate the influence of emphasizing ignored aspects of stimuli (in order to facilitate the inspection of these aspects) and the influence of presenting contrasting information (by providing the correct response after a mistake is made) in the processes of belief revision.

Before we describe the empirical research, we first clarify our definition of rules and the possibilities of the assessment of rule-guided behavior in Part 2. We consider criteria for rules and discuss the empirical findings on children's performance on the balance scale task in this respect. These findings provide so-called snapshots of development (Siegler & Chen, 1998). The present empirical research, presented in Part 3, focuses on the *process* in between the snapshots. This research is based on an earlier study (Jansen & van der Maas, 2001) that centered on the mechanism of change from Rule I to Rule II. We summarize their findings, as the manipulations in their study are central to the present research. The present empirical research considers both the transition from Rule I to Rule II and the transition from Rule II to more complex rules. We study whether children learn when information that was neglected before is emphasized or whether they learn from experiencing a mistake. The empirical findings of Part 3 give rise to the formulation of additional criteria for computational models of reasoning on the balance scale task. Computational models of balance scale learning have been designed to formalize theories about the mechanisms of balance scale learning. These models test the consistency of informal theories, such as neo-Piagetian ideas about balance scale learning, and may result in new predictions. In Part 4, we mainly focus on performance of a number of computational models on the simulation of rule-like behavior and rule transitions as defined in Part 2. In Part 5, we will discuss whether the rule transitions that we observe in children's reasoning about the balance scale could be considered belief revisions. As yet, unless computational models, belief revision models are not explicitly applied to the domain of balance scale learning. Discussion of the belief revision framework in this context is therefore necessarily more superficial than in the context of computational models.

2 The assessment of rules in empirical research

The concept of a "rule" plays a central role in cognitive science, but it is seldom the subject of debate, and precise definitions are usually not provided. In our view, a rule is a mental procedure that a child follows to solve problems (Jansen & van der Maas, 2002). Rules are typically thought of as being rule-like, abstract beliefs that can

be applied in a wide variety of actual cases, such as the potentially infinite number of balance scale problems. When equating transitions between rules with a process of belief revision, as we propose to do, some justification is needed about the status of psychological rules as beliefs. Here, we consider the criteria for the empirical assessment of rules, proposed by Reese (1989). The criteria are applicable in experimental research to differentiate the ascription of rule-like beliefs from mere rule-complying behavior.

2.1 Consistency of observed behavior with expected behavior

The criterion "Consistency of observed behavior with expected behavior" is operationalized by comparing children's responses to the expectations that follow from the suggested rules.[1] Many researchers observed that children's behavior was consistent with these expectations (e.g., Chletsos, De Lisi, Turner, & McGillicuddy-De Lisi, 1989; Normandeau, Larivée, Roulin, & Longeot, 1989; Siegler, 1976, 1981; Siegler & Chen, 1998). However, a problem of judging the consistency of observed behavior with expected behavior is that consistency will hardly ever be perfect as children sometimes make mistakes when executing the intended response. A non-arbitrary level of sufficient consistency is difficult to define. Rindskopf (1987) proposed to use latent class analysis to solve this problem.

The latent class model is a statistical model that is part of the family of latent structure models (LSM; Lazarsfeld & Henry, 1968). LSMs can describe data that were collected with human subjects as well as data that were generated by computational models, in a statistical model in which a latent variable influences manifest variables. LSMs can describe data whereas computational models mimic behavior of children and generate data. An LSM provides goodness-of-fit measures that provide information on the accuracy of the description of the data. The assumption that children's observed behavior corresponds to the behavior that is expected from the rules can be tested in a confirmatory way. In addition, LSMs can be used in an exploratory way by basing the selection of the best model on the goodness-of-fit measures only, and interpreting the latent categories by inspecting the corresponding responses (as in exploratory factor analysis). LSMs thus enable us to translate theories into formal models that can be tested statistically. The level of sufficient consistency is not determined arbitrarily but can be tested statistically. This solves the problem of behavioral variability: children are allowed to demonstrate a statistically determined level of inconsistencies of their strategy.

Jansen and van der Maas (1997) applied latent class analysis to children's responses to sets of balance scale items. They showed that the individual differences in children's behavior are indeed qualitative and that the types of observed behavior were mainly consistent with the rules that were proposed by Siegler (1981). Jansen and van der Maas (2002) observed the same progression from simple to more complex, distinct beliefs that Siegler demonstrated.

Contrary to Siegler's (1981) findings, Jansen and van der Maas (1997, 2002) observed additional rules for solving balance scale problems. One of the most frequently used rules in children of 11 years and older was the compensation-rule (Halford, Andrews, Dalton, Boag, & Zielinski, 2002). The rule refers to several procedures in which the

---

[1] Note that consistency here does not refer to its usual meaning with respect to belief sets, but to the correspondence between ascribed beliefs and behavior that is generated using those beliefs.

differing values of distance and weight on conflict items are compensated for by each other. Children may either sum the number of weights and the distance on each side of the scale and compare the sums (addition-rule, Normandeau et al., 1989) or they may shift the larger pile of weights away from the fulcrum and take away one weight at each shift until equal numbers of weights or equal distances are obtained (buggy-rule, van Maanen, Been, & Sijtsma, 1989). The complexity of the compensation-rule is comparable to that of Rule III. The age of the children using the rule also supports the idea that it is an alternative for Rule III.

2.2 Regularity

Regularity of behavior is another criterion for rule-guided behavior. Studying whether the same pattern of responses is demonstrated on several parts (e.g., both halves) of a balance scale test operationalizes this criterion. Jansen and van der Maas (1997, 2002) concluded, by applying latent class analysis to several parts of a balance scale test, that regularity is observed for the patterns of responses that are associated with Rule I and the compensation-rule. The use of Rule II was sometimes alternated with Rule III. Jansen and van der Maas (2002) suggested that children who use Rule II are susceptible to rule transitions as the difference between Rule II and Rule III seems quantitative. The distance dimension is noted in both rules, but used on only a few items in Rule II and on all items in Rule III. Jansen and van der Maas (2002) observed consistency for Rule III on balance, weight, and distance items but not on conflict items. This is not surprising as Rule III implies guessing. However, Jansen and van der Maas (2002) also observed that the use of Rule III was alternated with the compensation-rule, a mix of both rules, and Rule II. Rule IV was hardly observed in the studies of Jansen and van der Maas (1997, 2002).

2.3 Discontinuity

The change from one type of behavior to another (more complex) type of behavior is supposed to be discontinuous, in case of rule-guided behavior. Jansen and van der Maas (2001) found evidence for the hypothesis that the change from Rule I to Rule II indeed happens discontinuously. The transitions to Rule III, the compensation-rule, and Rule IV were not studied yet. However, Jansen and van der Maas (2002) suggest that the transition to Rule III may be gradual, as children seem to alternate between the pattern that is associated with Rule II and the pattern that is associated with Rule III. Transitions to the compensation-rule and Rule IV may be discontinuous as the patterns associated with these rules are quite stable, but again, these transitions were not investigated yet. See van der Maas and Molenaar (1992) for precise empirical criteria for testing discontinuity in behavioral transitions.

2.4 Awareness of rule use

A fourth criterion is that subjects should show evidence of awareness of rule use. A verbalization of a response could indicate awareness of rule use if both verbalization and response indicate usage of the same rule. However, verbal methods may underestimate children's knowledge because the interaction between experimenter and child may cause confusion for the child and because linguistic development may lag behind cognitive development. Many researchers (e.g. Siegler & Stern, 1998) found

that more implicit measurements, like response times, reveal use of a rule or strategy at an earlier stage than explicit strategies, like the answers to an interview. Siegler and Stern demonstrated that strategies could be discovered without conscious awareness. See Jansen (2001) for a further discussion about the use of verbal versus nonverbal assessment of rules that children use on Piagetian tasks. Although Siegler (1981) did report correspondence between children's verbal reports and their responses to a balance scale test, we do not consider the criterion of awareness of rule use here.

2.5 More than one kind of behavior is expected and observed

Van der Maas and Jansen (2003) measured both children's responses and their response times and concluded that both measures are in accordance with expectations. They do note however that Rule II-children seemed to make more use of the distance cue than formerly thought. Their inspection times on conflict items (with differing numbers of weights) were unexpectedly long as Rule II only implies a comparison of the numbers of weights in the case of differing numbers. When the definition of Rule II is adjusted accordingly, both accuracy data and response times are regular and predictable. Hence, all rules satisfy the criterion that more than one kind of behavior should be observed and expected.

2.6 The inferred rule generalizes to other behaviors and tasks

Finally, Reese (1989) suggests that the ascribed rule should generalize to other behaviors and tasks. Domain-general studies are scarce in current research. Dolan, Jansen, and van der Maas (2004) apply finite mixtures to the conservation of liquid task and observe, in general, the same types of rules as in the balance scale task, like Rule I. However, more research needs to be done in this area.

Summarized, we conclude that Rule I fulfills all criteria. Rule II fulfills most, although the findings on response times imply that the distance dimension *does* play a more important role in the rule model than formulated by Siegler (1976, 1981). The nature of the difference between Rule II and Rule III may be more quantitative than qualitative, which fits with the suggestion that the transition from Rule II to Rule III is gradual. Behavior that is associated with Rule III is less regular and consistent with expected behavior than the remaining rules. Although both responses and response times can be predicted, the change from Rule II to Rule III seems continuous. The definition of Rule III as a proper rule is doubtful. The available findings on the compensation-rule and Rule IV do support the hypothesis that these are proper rules.

When above criteria are fulfilled, we believe we are justified in ascribing rule-like beliefs to subjects. Consistency, both between observed and expected behavior, and between an individual's responses to items across a balance scale test, and discontinuity between consecutive rules are strong indicators that children move from one belief state to the next. In the following therefore, we use the terms rule and belief interchangeably on the understanding that the rules we infer from behavioral data are indicative of underlying belief states that are amenable to revision (see Gärdenfors & Rott, 1995, for definitions of the kinds of belief states and belief revisions that we have in mind). Note that Siegler's Rules I-IV are abstract beliefs that can be applied to a potentially infinite set of balance scale problems. Hence, when viewing rules as beliefs, the beliefs are not factual beliefs about particular balance scale problems but

[55]

rather abstract procedural beliefs for dealing with balance scale problems in general (see Paglieri, 2004, for a similar distinction). Evidence for rule transitions, given that they are discontinuous, can thus be seen as evidence for belief revision. In the General Discussion section we discuss what kind of belief revision rule transitions could be.

However, the above summary of results does not reveal the underlying processes that determine rule transitions (Siegler & Chen, 1998). Understanding these processes is an essential prerequisite to obtain causal explanations of development and may help uncovering the mechanisms that are of importance in rule transitions and may inspire new research in belief revision.

3 Empirical research

3.1 Inducing rule transition by the facilitation of perception

3.1.1 Theoretical background

The question whether development proceeds continuously or discontinuously is a central and recurrent theme in developmental psychology. When development happens continuously, children constantly refine their beliefs, resulting in quantitative changes. Discontinuous development occurs suddenly and implies a qualitative difference between the behavior before and after the change.

To investigate whether change is truly discontinuous, one requires a formal model for discontinuous development. Van der Maas and Molenaar (1992) noted that the cusp model, which is a formal model of discontinuity, derived from catastrophe theory, could be applied to investigate the nature of development. In catastrophe theory, development is modeled by means of a phase transition.

In the cusp model, the discontinuous change in a variable depends on continuous changes in two independent variables: the normal and the splitting variable. Here, we focus on the relationship between the normal variable and the dependent variable, ignoring the splitting variable. In an application of the cusp model to the balance scale task, the dependent variable corresponds to balance scale performance. Continuous changes in the normal variable cause a discontinuous improvement in balance scale performance. Changes in the normal variable in the opposite direction cause a discontinuous deterioration in balance scale performance. In a true phase transition, the opposite changes occur at different levels of the normal variable. This phenomenon is called "hysteresis" and it is a distinctive feature of a discontinuous development. A more detailed account of catastrophe theory can be found in general introductions to catastrophe theory (e.g., Gilmore, 1981) and in van der Maas and Molenaar (1992, 1996).

The main objective of Jansen and van der Maas (2001) was to find evidence for the existence of hysteresis in the development from Rule I to Rule II. This objective goes hand in hand with the challenges to study development as it occurs *and* to invert development, as hysteresis can only be studied when both improvement (from Rule I to Rule II) and deterioration of balance scale performance (from Rule II to Rule I) are observed. This made it necessary to make assumptions on the independent variable that controls the difference between children who use Rule I and children who use Rule II.

Siegler (1976) suggested that children who employ Rule I do not observe the distance dimension. When trained to notice it, they were more likely to use it in their solution strategies. Jansen and van der Maas (2001) concluded that the independent variable might be the ability to perceive the distance dimension. Emphasizing and abating the distance dimension manipulated the variable. This was done by increasing and decreasing the distance difference: the absolute difference between the distance between the pile placed on the left side of the scale and the fulcrum and the distance between the pile placed on the right side of the scale and the fulcrum. The distance difference in Fig. 1 is one. A wide scale, with six pegs on each side, was used to maximize distance difference. Values on the weight dimension were constant, to increase the salience of the distance dimension. The hypothesis was that children, who focus on the weight dimension only, start to notice the distance dimension when the dimension is more salient and start to use this dimension in their solution strategy. If the development from Rule I to Rule II is discontinuous, children might show a sudden jump, from incorrect responses to correct responses, and back, when distance difference is decreased again. Hysteresis occurs when the jump to a lower level of balance scale performance (incorrect response to distance items; Rule I) occurs at a smaller value of the normal variable (distance difference) than the jump to the higher level of balance scale performance (correct response to distance items; Rule II). When assuming gradual development, children might alternate between correct and incorrect responses.

3.1.2 Method

The manipulations were performed in two experiments but the results of the first experiment suffice for understanding the manipulations in the present empirical research. The first experiment concerned 314 children, from 7 to 10 years old. A paper-and-pencil balance scale test was administered to obtain children's rule use. This pretest was followed by either a "hysteresis" test or a control test. Children's rule use was tested again in a posttest, followed by either the control or the hysteresis test.

The hysteresis test consisted of a series of nine distance items. On all items, two weights were placed on both sides of the fulcrum. The weights on the left were always placed on the first peg from the fulcrum. On the right side, the weights were placed on the second peg from the fulcrum (distance difference = 1). In the next items, the weights on the right were shifted away from the fulcrum, one peg per item. On the fifth item, the weights were on the sixth (i.e. the most distant) peg from the fulcrum (distance difference = 5). From the sixth item on, the weights on the right were shifted back towards the fulcrum (distance difference = 1). Figure 2 helps in understanding this manipulation. The distance difference of the second item in series Facilitation II (explained later) is one. It increases until three in the fourth item. Jansen and van der Maas (2001) were able to increase distance difference until five because of the wider scale. Decreasing the distance difference is not shown in Fig. 2. On the items of the control test, an irrelevant feature of the balance scale, grayness of the weights, was manipulated.

A pattern that corresponds to a hysteresis pattern can be 000111100, with zero and one being an incorrect and a correct response, respectively. The jump to correct responses occurs at a distance difference of four, whereas the jump to incorrect responses occurs at a distance difference of two.
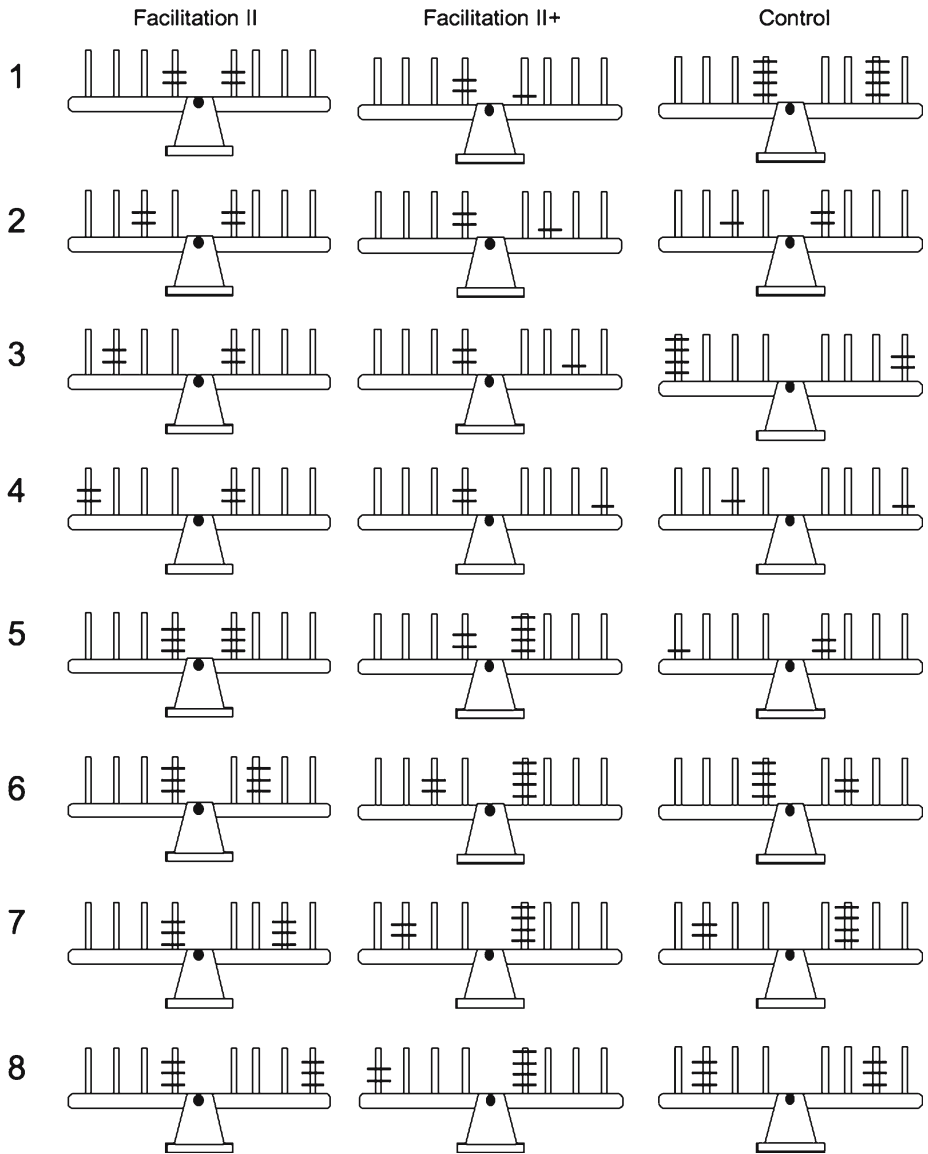
⚫ Springer

Fig. 2 Items of series Facilitation II, Facilitation II+, and the Control Series. The correct responses to these items were presented only in half of the conditions

3.1.3 Results

The responses to both the hysteresis test and the control test were analyzed with a confirmative latent class analysis. The latent class model contained classes associated with all possible hysteresis patterns; all Maxwell hysteresis patterns (jumps forth and back occur at the same value of the normal variable—this pattern confirms to a gradual development but is not incompatible with discontinuous development);

Table 1 Proportions of hysteresis patterns on hysteresis test and control test, experiment 1 of Jansen and van der Maas (2001)

	Hysteresis test	*Control test*
Rule I	.341	.395
Rule II	.350	.427
Delay	.029	.003
Maxwell	.035	.000
Jump	.096	.038
Residual	.092	.073
Missing values	.057	.064

Note. Patterns associated with Delay hysteresis refer to hysteresis as described in the text. Patterns associated with Maxwell hysteresis refer to patterns in which the jumps forth and back occur at the same value of the normal variable. Jumps refer to a series of incorrect responses followed by a series of correct responses

$N = 314$

all possible patterns involving a sudden jump; the use of Rule I; the use of Rule II; and a residual class. Table 1 shows the proportions of the response patterns in both the control and the hysteresis test. The difference between the two frequency distributions was significant (χ^2 (6, $N = 314) = 29.52$, $p = .000$). More children displayed patterns corresponding to hysteresis, and jumps upward on the hysteresis test than on the control test. Responses from children who made the hysteresis test in between the pretest and the posttest were analyzed separately. Most children who employed Rule I on the pretest and who showed a sudden jump on the hysteresis test employed a more complex rule on the posttest. Although jumps to the correct answer were also displayed on the control test, the number of such jumps was significantly smaller than that on the hysteresis test ($\chi^2(1, N = 42) = 7.71, p = .005$). Moreover, patterns corresponding to hysteresis were virtually absent at the control test.

3.1.4 Conclusions regarding the mechanism of change

The evidence for the presence of hysteresis supports the hypothesis that the transition from Rule I to a higher level of balance scale reasoning happens discontinuously. More relevant for the present research is that presenting children with a specific set of items, without providing the correct response, may cause a rule transition. Jansen and van der Maas (2001) demonstrate that children may indeed start to use a more complex rule because they start to observe aspects of the balance scale that they ignored before. It should be noted that a very small number of children also showed a sudden jump when the grayness of the weights was manipulated in the control test. This means that merely presenting children distance items may cause them to use the distance dimension in their solution strategy.

The findings of Jansen and van der Maas (2001) leave unrevealed the transitions to more complex rules. Moreover, the second possible cause of rule transitions, the acquisition of information that stands in contrast to old beliefs, remains unstudied. Studying these possible causes of rule transitions is the objective of the present research.

[59]

3.2 Inducing rule transitions on the balance scale task by facilitating perception and feedback

3.2.1 Material

All children made a pretest, an intervention test, and a posttest. A second posttest was administered 2 weeks later ("delayed posttest"). All balance scale tests were paper-and-pencil tests. The pretest, posttest, and the delayed posttest consisted of twelve balance scale items of different types. Conflict items were constructed such that the compensation-rule resulted in an incorrect response to conflict-distance and conflict-weight items and a correct response to conflict-balance items.

Two series were designed to emphasize information that was ignored before. The first series, "Facilitation Rule II" (depicted in the left panel of Fig. 2) was based on the hysteresis test of Jansen and van der Maas (2001). Distance difference was increased in order to trigger children to take notice of the distance dimension.

The second series, "Facilitation Rule II+" (depicted in the middle panel of Fig. 2), was aimed at facilitating the transition from Rule II to more complex rules, such as Rule III. In this test, children were pointed to a salient distance difference, combined with a difference on the weight dimension. Distance difference was increased in two sets of four items. The first set started with a weight item (two weights placed on the left side of the scale and one weight on the right side of the scale; both piles were placed on the first peg from the fulcrum; $w_l = 2$, $d_l = 1$, $w_r = 1$, $d_r = 1$). The pile on the right was shifted away from the fulcrum in each subsequent item, resulting in a conflict-balance item ($w_l = 2$, $d_l = 1$, $w_r = 1$, $d_r = 2$), and two conflict-distance items ($w_l = 2$, $d_l = 1$, $w_r = 1$, $d_r = 3$; $w_l = 2$, $d_l = 1$, $w_r = 1$, $d_r = 4$). The second set also consisted of, respectively, a weight item, a conflict-balance item, and two conflict-distance items, as the pile of weights on the left side was shifted away from the fulcrum. These sets of items resemble the sets Siegler (1976) used.

Children who use Rule II and receive feedback will notice that their responses become incorrect when the pile of weights is shifted (even though the weight difference remains the same). Hence, they may conclude that a difference on the distance dimension may compensate or even surpass the importance of a difference on the weight dimension.

3.2.2 Design and procedure

The combination of the factor "Facilitation" (consisting of the series "Facilitation Rule II", the series "Facilitation Rule II+", and a control series; see Fig. 2) and the factor "Feedback" (showing the correct response versus no feedback) resulted in six conditions: "FacII_nofb", "FacII_fb", "FacII+_nofb", "FacII+_fb", "C_nofb", "C_fb". School classes were randomly assigned to these six conditions. All children in one class were assigned to the same condition because of the class wise administration of the balance scale test. The dataset consisted of a total of 420 participants. Their ages ranged from 6 to 10 years.

Two experimenters were present in each classroom. One experimenter stood in front of the classroom and demonstrated the workings of a wooden scale. She pointed out the similarity between this scale and the pictures of the scale in the booklets of the children. The experimenter solved four example items together with the children.

After explaining the procedure of response notation and correction, the children were asked to respond to the items of the pretest individually.

Next, the experimenter put the weights on the wooden balance scale, according to the configurations of the items in the intervention test. A blocking pen prevented the scale from tipping. In the feedback conditions, the children were able to see the correct response because the experimenter removed the blocking pen after children filled in their responses. In the conditions without feedback, the experimenter continued with the next item, without removing the blocking pen. Children responded to the items of the posttest individually. The delayed posttest was administered 2 weeks later, in a similar fashion.

3.2.3 Results

A multi-group latent Markov model (Langeheine, 1994) was fitted to the responses to the items of the pretest and the posttest (scored either correct or incorrect). In this type of model, the structure in the data set is disclosed by dividing the response patterns in the dataset in a finite number of classes (for both the pretest and the posttest). Moreover, the transitions (from pre- to posttest) between classes are described in transition matrices. Only a small selection of items was analyzed as computational problems increase and the reliability of fit measures lowers, when the number of possible response patterns increases. The selection consisted of two distance items, one conflict-weight item, and two conflict-distance items, administered at both the pretest and the posttest. Only the responses of children who responded to these ten items, who succeeded at the weight items in the practice part and the pretest, and who responded "balance" to less than half of the items in the pretest were selected. In earlier research, we noted that some children did not understand that they needed to imagine what happened if the blocking pin was removed and consistently responded "balance" to all items because the blocking pin prevented the scale from tipping (van der Maas & Jansen, 2003). A total of 349 children were included in the data set.

In the multi-group latent Markov model, each latent class is associated with a pattern of conditional probabilities. A conditional probability corresponds to the probability of giving a certain response, given membership of a certain latent class. This means that each latent class is associated with conditional probabilities for each item, and each response category (false/correct). In this application, a class is associated with a rule. For example, it is expected that one of the latent classes is associated with Rule I and shows a high conditional probability for answering the conflict-weight item correctly, but low probabilities for answering the remaining items correctly. We restricted the conditional probabilities of the items of the posttest to the same values as the conditional probabilities of the items of the pretest. In this manner, the classes of the pretest and the posttest have the same interpretation.

The six conditions featured as groups in the multi-group latent Markov model. The conditional probabilities were also restricted to be equal across conditions in order to optimize comparison of the classes in the conditions. Hence, the classes have identical interpretations in all conditions.

The proportions of children in each latent class on the pretest were modeled by the unconditional probabilities. These proportions were allowed to differ between conditions. Restricting the unconditional probabilities of the conditions to the same values would imply that the distribution of the rules over the conditions was identical. This is a very strict requirement. Although a random division of participants over

the conditions was the aim, this was hindered by the class wise administration of the balance scale tests. The proportions of children that used the rules on the posttest depended on the transition matrices.

A restricted four-class model was selected. The fit of this model was acceptable (LR = 487.32, pb = .075) and the BIC was low, compared to other four-class models. Figure 3 shows the conditional probabilities for giving a correct response to each of the five items, for the four latent classes.

Figure 3 shows that the behavior of all children was largely interpreted in terms of known rules. The first latent class (lc1) was mainly associated with low probabilities of giving a correct response. Only the conditional probability of answering the conflict-weight item correctly was high (near one). This pattern of probabilities corresponds to the use of Rule I.

The second latent class had high conditional probabilities of answering both distance items and the conflict-weight item correctly. The conditional probabilities of answering the conflict-distance items correctly were low. This pattern of probabilities indicates the use of Rule II.

The third latent class showed high probabilities of answering the distance items correctly, but mixed probabilities of answering the conflict items correctly, which is consistent with using Rule III.

The fourth latent class shows an unexpected pattern of conditional probabilities. All conditional probabilities of giving the correct response were below .5. This pattern of responses remained uninterpreted.

The unconditional probabilities, corresponding to the proportions of children in each class on the pretest, showed the same qualitative pattern in each condition, in spite of some quantitative differences. The first latent class, associated with Rule I, contained the largest proportion of children in all conditions. The average proportion was .56. On average, .33 of the children were in the second latent class, interpreted as Rule II. A proportion of about .09 of the children used Rule III. The uninterpreted class was absent in three conditions (FacII_fb, FacII+_nofb, C_fb). In the remaining
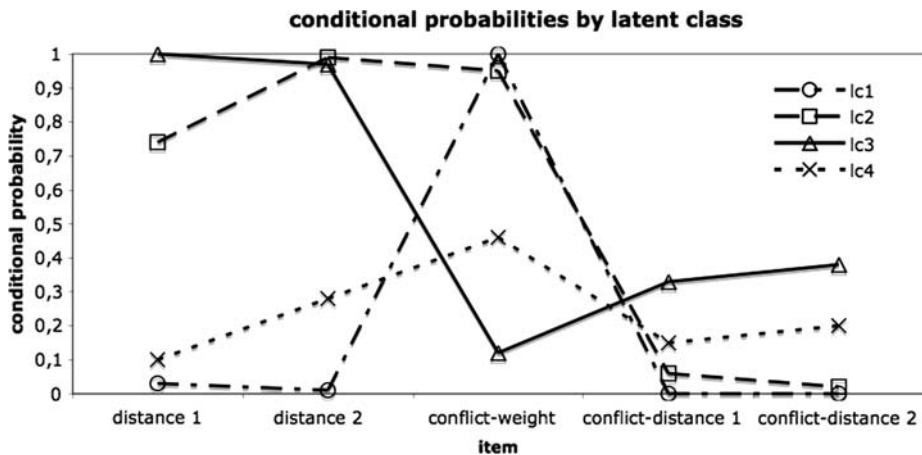

Fig. 3 Estimated conditional probabilities of the four latent classes on the pretest and the posttest, in the four-class multi-group latent Markov model

Table 2 Estimated values of parameters of the transition matrices, indicating the transition from performance on pretest to performance on posttest, by condition

	Rule I	Rule II	Rule III	Unknown
FacII_nofb & FacII_fb				
Rule I ($p = .57/.65$)[a]	.44	.46	.05	.05
Rule II ($p = .32/.25$)[a]	.00	1.00	.00	.00
Rule III ($p = .05/.10$)[a]	.00	.00	1.00	.00
Unknown ($p = .06/.00$)[a]	.61	.00	.00	.39
FacII+_nofb				
Rule I ($p = .63$)[b]	.82	.13	.00	.04
Rule II ($p = .34$)[b]	.04	.75	.12	.09
Rule III ($p = .03$)[b]	.00	.00	1.00	.00
Unknown ($p = .00$)[b]	.00	.00	.00	.00
FacII+_fb				
Rule I ($p = .55$)[b]	.18	.03	.52	.27
Rule II ($p = .28$)[b]	.00	.00	.58	.42
Rule III ($p = .11$)[b]	.00	.00	1.00	.00
Unknown ($p = .07$)[b]	.12	.00	.00	.88
C_nofb				
Rule I ($p = .48$)[b]	.83	.06	.00	.11
Rule II ($p = .35$)[b]	.09	.76	.04	.11
Rule III ($p = .11$)[b]	.14	.00	.86	.00
Unknown ($p = .06$)[b]	.55	.45	.00	.00
C_fb				
Rule I ($p = .45$)[b]	.01	.13	.29	.57
Rule II ($p = .44$)[b]	.04	.09	.58	.28
Rule III ($p = .12$)[b]	.00	.00	1.00	.00
Unknown ($p = .00$)[b]	.00	.00	.00	.00

[a] Unconditional probability of class on pretest; first probability refers to condition "FacII_nofb", the second probability refers to condition "FacII_fb"
[b] Unconditional probability of class on pretest

conditions, this class had an unconditional probability of only .06 or .07 (see also Table 2).

The unconditional probabilities and the conditional probabilities of the selected model replicated findings of earlier studies (Jansen & van der Maas, 1997, 2002). Observed behavior is consistent with the expectations that follow from the rules suggested by Siegler (1976, 1981). Of interest here are the transition matrices of the multi-group latent Markov model. In this application, a transition matrix shows the transition from a certain latent class on the pretest (Rule I, Rule II, Rule III, or the uninterpreted class) to one of the latent classes on the posttest (again: Rule I, Rule II, Rule III, or the uninterpreted class). The transition matrix may be identical to an identity matrix, showing no change, but the transition probabilities may differ from zero and one as well. For each condition, a transition matrix was estimated but the transition matrices for the conditions "FacII_nofb" and "FacII_fb" were restricted to be equal in the selected model, suggesting that feedback does not have an additional

effect on the transition from Rule I to Rule II. Table 2 shows the estimated values of the transition matrices of the selected model.

Children who were using Rule I on the pretest benefited from the items in series "Facilitation II". Almost half of these children progressed to using Rule II on the posttest, whether the correct response was demonstrated or not. Children who were using Rule II or Rule III on the pretest stuck to their solution strategy. The uninterpreted class did show a transition to Rule I, but this class was very small.

The series "Facilitation II+" was successful in generating progression for children using Rule II, but only when feedback was provided. Remarkably, also children using Rule I changed their solution strategy after they were allowed to observe that their responses to the items in the intervention test were wrong.

The transition matrices of the control conditions showed similarities with the transition matrices of conditions "FacII+_nofb" and "FacII+_fb". Both children using Rule I and children using Rule II changed their behavior on the balance scale task, but only after receiving feedback. A considerable proportion of these children used an unknown strategy on the posttest. The results in the feedback conditions match those in Experiment 2 of Siegler (1976), who observed that children using Rule I benefited from watching the movement of a balance scale on which either distance problems or conflict problems were presented.

The responses to the so-called "delayed posttest", administered 2 weeks after the former test, were analyzed with regular latent class analysis. The children in the six conditions were joined in this analysis. The fit measures indicated that only three classes were needed to accurately describe the responses to a selection of items of this test. The response pattern of the largest latent class was expected to stem from the use of Rule II. The proportion of Rule II-users has increased from about .30 on the pretest, to .55 on the delayed posttest. The response pattern observed in the second largest latent class is associated with the use of Rule I. The proportion of this class ($p = .32$) has decreased remarkably, compared to the proportion on the pretest. The smallest latent class was associated with high conditional probabilities for answering the distance items and the conflict-distance items correctly. However, the conditional probability of answering the conflict-weight item correctly was very low. This pattern of responses was not observed in the multi-group latent Markov model of the pretest and the posttest. Possibly the children in this latent class tended to answer that the scale tips to the side with the larger distance, resulting in a correct response to distance and conflict-distance items, but an incorrect response to conflict-weight items. The proportion of this Distance Dominant Rule (DD-rule) was .13, which is large, compared to earlier findings (Jansen and van der Maas, 2002).

Next, we connected rule use on the delayed posttest to rule use on the pre- and posttest. Subjects were assigned to their most likely latent class by means of a posteriori probabilities, i.e., the probability that a given subject belongs to a particular latent class (Rabiner, 1989). Table 3 shows class assignment for the classes on the pretest, the posttest (both were based on the estimated parameters of the multi-group latent Markov model) and the delayed posttest (based on the three-class latent class model), for the children who used either Rule I or Rule II on the pretest.

According to classification by means of a posteriori probabilities, 35 children progressed from using Rule I on the pretest to Rule II on the posttest. On the delayed posttest, 29 of these children were still using Rule II. Only five children were using Rule I again, and only one used the DD-rule. A total of 24 children progressed from using Rule I on the pretest to Rule III on the posttest. On the delayed posttest, 14

Table 3 Cross-tabulation of a posteriori classifications of behavior on pretest, posttest, and delayed posttest, for children using Rule I or Rule II on the pretest

PRETEST	POSTTEST	DELAYED POSTTEST				Total
		Rule I	Rule II	Rule III[a]	Unknown[b]	
Rule I	Rule I	61	24	-	1	86
	Rule II	5	29	-	1	35
	Rule III		14	-	10	24
	Unknown	7	12	-	5	24
Rule II	Rule I	2	2	-		4
	Rule II	7	51	-	5	63
	Rule III		19	-	4	23
	Unknown	4	9	-	3	16

[a] Rule III was not observed in the responses to the delayed posttest

[b] The unknown class in the analysis of the delayed posttest was interpreted as Distance-dominant Rule

of these children were using Rule II, whereas ten of these children were using the DD-rule. A total of 24 children progressed from using Rule I to the unknown rule on the posttest. Half of these children (12) used Rule II on the delayed posttest, but a considerable number of children (seven) regressed to Rule I. Five of the children were using the DD-rule. We conclude that use of the more complex rule lasts if children show a transition from Rule I to a more complex rule. If children start to use an unknown strategy (of which complexity is unclear), there is a fair chance that they regress to using Rule I.

Progression from Rule II on the pretest to Rule III on the posttest occurred for a total of 23 children. Remarkably, Rule III was not observed in the latent class model of the responses to the delayed posttest. Instead, 19 of the children regressed to using Rule II, whereas the remaining four children used the DD-rule. A total of 16 children started to use the unidentified strategy on the posttest, after using Rule II on the pretest. Their behavior was classified in Rule I (four children), Rule II (nine children), or the DD-rule (three children) on the delayed posttest. We conclude that the progression from Rule II to a more complex rule does not last.

3.2.4 Conclusions

Half of the children who used Rule I, and who were presented with distance items, progressed to using Rule II. They continued using this rule 2 weeks later, indicating that the change persisted. The results replicate the results of Jansen and van der Maas (2001), from which we concluded that children might change from Rule I to Rule II when presented with facilitating items, without providing feedback. Facilitating items did not help children using Rule II. Possibly, the wrong stimuli were emphasized, or the wrong items were used. So far, the findings of the present research show that only children who use Rule I can learn without feedback.

Many children who used Rule II on the pretest and who made mistakes (and perhaps experienced they were wrong) on either series Facilitation II+ or the Control series, used Rule III on the posttest. However, the improvement to Rule III did not last and most children regressed to using Rule II on the delayed posttest. This suggests

that Rule III is an unstable rule and that children fly back to Rule II when they are insecure about the correct combination of the two dimensions.

In the next part of this article, Part 4, we discuss the challenges that these findings provide for computational models for balance scale learning. First, computational models should be able to *learn without feedback (*Van Rijn, van Someren, & van der Maas, 2003) but only when learning concerns the transition from Rule I to Rule II. Second, computational models should *not* demonstrate a permanent transition from Rule II to a more complex rule when only negative feedback is presented to the model. We would like to add the *regression* criterion to the criteria suggested by van Rijn et al. The regression criterion states that using Rule II cannot be permanently improved by merely showing children the correct response after making a mistake. Regression to Rule II is expected after feedback is withheld again.

4 Computational models of balance scale learning

Several computational models have been developed to formalize the theories about mechanisms of balance scale learning. The statistical models applied to the empirical data give a reliable description of the performance, but do not explain how changes come about. Computational models are defined to explain how children change their rule-like behavior on the balance scale task. A model of balance scale learning should first capture the developmental phenomena associated with the balance scale task. Our focuses in discussing these models are the criteria about rule-like behavior and rule transitions (Part 2) and the two additional criteria that followed from the findings in Part 3: learning without feedback, and regression. As with empirical data the assessment of rules from the behavior of computational models can be conducted by applying latent class analysis. For models with explicit production rules, however, the inspection of the models' implementation directly is more insightful. Furthermore, in case a model meets most empirical criteria, we evaluate whether predictions that could be tested in empirical studies can be derived from the model. The models of balance scale learning that we discuss, originate from different computational traditions, among which are *connectionist models* (Dawson & Zimmerman, 2003; McClelland, 1989, 1995; Shultz, 2003; Shultz & Schmidt, 1991; Shultz, Mareschal, & Schmidt, 1994; Shultz, Schmidt, Buckingham, & Mareschal, 1995), *production-rule models* (Klahr & Siegler, 1978; Langley, 1987; Sage & Langley, 1983), *decision-tree models* (Schmidt & Ling, 1996) and *hybrid models* (van Rijn et al., 2003). Van Rijn et al. (2003) give a nice overview of these models compared to empirical findings on balance scale learning. So far, none of these models has been able to explain all empirical phenomena, but some of these models give important insights in possible mechanisms of balance scale learning. Because we evaluate models on the basis of empirical criteria, we only evaluate computational models of balance scale learning that are actually described in literature and we will not discuss what computational models of different type could capture in principle.

4.1 Production rule and decision tree models

Rule-like behavior is obvious in computational models that are composed of production rules or decision trees. In all known production rule models (Klahr & Siegler, 1978; Langley, 1987; Sage & Langley, 1983) and decision tree models (Schmidt & Ling,

1996) of balance scale learning, the rules match the rules found in humans. However, the models do not acquire Rule IV, or the compensation-rule, as multiplication and compensation mechanisms are not implemented in these models. The second criterion concerns rule transitions. In these models, learning occurs from negative feedback after an incorrect response. The mechanism of change consists of constructing more complex rules by adding conditions to the lower rule in the hierarchy. The regression criterion does not apply to these models because more complex rules, i.e. Rule III, will not be simplified to Rule II spontaneously, as we observed in the empirical study. Learning without negative feedback also does not occur in these models.

4.2 Connectionist models

Rule-like behavior is less obvious for connectionist models. Connectionist models of balance scale learning (McClelland, 1989, 1995) are often presented as the best example of models that are defined on a low, sub-symbolic level but nevertheless show rule-like behavior (Elman et al., 1996). Connectionist models are constructed of simple units that activate each other according to low-level activation rules. Input units represent a particular balance scale item; output units represent a particular response. Input units are connected to output units via hidden units at which an internal representation is formed during learning. Not only are balance scale rules not defined explicitly, also the learning mechanism of connectionist models is defined on a very low level and is not domain-specific. Learning is implemented by changing the strengths of the connections between the input, hidden and output units. Cascade-correlation models (Shultz & Schmidt, 1991; Shultz, Mareschal, & Schmidt, 1994; Shultz et al., 1995) strongly resemble McClelland's connectionist models but have an additional learning mechanism: hidden nodes are adaptively added during the process of learning. Specific feedback, that is, the correct response, is needed to change the strengths. Conclusions that connectionist models of balance scale learning show rule-like behavior are based on application of the rule assessment methodology (Siegler, 1976, 1981) to the response data. However, the additional application of the statistical technique of latent class analysis shows a very different picture of balance scale learning by connectionist models. Jansen and van der Maas (1997) demonstrated that the behavior of McClelland's (1989) connectionist model could not be classified into rules. Moreover, Raijmakers, van Koten, and Molenaar (1996) showed that this model fails to show qualitative transitions between the rules. In contrast, the cascade-correlation network does show rule-like behavior (Quinlan, van der Maas, Jansen, Booij, & Rendell, in press). However, there were notable differences between the rules found in empirical data and the rules found in the cascade correlation data. Rules I, II, and the addition-rule were recognized in the classes, but the model also performed according to rules that were never found in human performance. The results of latent class analysis did suggest qualitative transitions between several rules, but a proper test of the data (as was done for McClelland's model by Raijmakers et al., 1996) has not yet been performed for the cascade correlation network.

Learning without negative feedback is not possible in these connectionist models. This is mainly due to the fact that these models have only feedforward connections, which implies that activation of units, i.e. behavior, does not depend directly on activation at a former point in time. The regression criterion also does not directly apply to these models for the same reason. That is, if a certain behavior is acquired due to learning by feedback then the behavior can only change again by additional learning.

[67]   ☺ Springer

4.3 Hybrid models: Act-R

Van Rijn et al. (2003) present an *ACT-R* model of reasoning on the balance scale task. ACT-R is a hybrid cognitive architecture in which the use of symbolic knowledge is mediated by associated quantitative values. The general solution strategy of the model of van Rijn et al. is to search for differences between the left and right side of the scale. In each phase, this search is limited by the model's mechanisms, task-specific concepts, and capacity constraints. Task-specific concepts refer to the availability of a particular concept or property, like addition or multiplication. Capacity constraints refer to the amount of information that can be incorporated simultaneously in a particular strategy. Initially, children can only accommodate a difference on one dimension. The general strategy, which first consists of only comparing the numbers of weights, is expanded by the availability of external sources (e.g., perception of distance). Multiplication and addition appear later in development. Capacity increases continuously during development. Moreover, production rules can be merged into a single more complex rule, resulting in more efficient processing, and hence less problems with capacity. The ACT-R model of balance scale learning simulates most available empirical phenomena of balance scale learning. As yet, the model does not simulate the empirical results derived from response time data (van der Maas & Jansen, 2003) although the prediction of response times is possible.

Rule-like behavior is obvious in this model since rules are explicitly represented. The rules correspond to the rules identified by Siegler (Rules I, II, III, and IV) and to the compensation-rule. Moreover, the order of rule transitions matches the empirically observed sequence. Learning without feedback does occur in ACT-R because the application of a rule does not only depend on the suitability of a rule but also on the frequency of retrieval. Hence, extended exposure to initially neglected information may result in a rule transition. Van Rijn et al. (2003) show that the hysteresis patterns, as were found in Jansen and van der Maas (2001), can be simulated by ACT-R. The specific results of the second study presented in Part 3 were not simulated in ACT-R. The absence of learning without negative feedback in Rule II users and the regression criterion pose new challenges to the ACT-R model.

Van Rijn et al. (2003) make some predictions about balance scale learning that have not been tested yet. In general the model predicts an important role for visual saliency. But the effects of visual saliency can only influence the behavior of the model during the less stable periods around transitions.

4.3.1 Discussion

Computational models of balance scale learning, in contrast to the descriptive statistical models, give insight into the possible mechanisms of rule transitions. The existing rule-production and decision tree models of balance scale learning have some important shortcomings in simulating balance scale learning. That is not to say that these approaches are principally ineffective. The connectionist models are powerful learning models because no explicit symbolic rules are hardwired into the system. Unfortunately, despite of several claims in literature, these models do not show proper rule-like behavior for balance scale learning. Because of their attractive set of assumptions about the architecture and learning mechanisms, in our view it is worthwhile to consider connectionist models (that is, neural network models) of balance scale learning in the future. The ACT-R model has a less attractive set of

assumptions because more information is put into the system by hand (for example addition and multiplication operations). Nevertheless, the model gives an important insight in balance scale learning: Learning by searching for differences between the left and the right side of the scale is an effective strategy for finding solutions. Apart from learning from feedback, ACT-R also formalizes a mechanism for *learning without negative feedback*. These mechanisms could generalize to other cognitive domains, because ACT-R is a general-purpose cognitive architecture.

5 Rule transitions: the case for belief revision

Given our results about rules and the factors playing a role in rule transitions, what can we say about rule transitions as a form of belief revision? In order to answer this question, we first need to evaluate the evidence for the existence of distinct belief states in development on the balance scale task.

In the case of Rule I and Rule II, there are a number of reasons to assume that children behaving according to those rules do indeed, more or less explicitly, possess those rules and base their behavior upon them. The latent class models fitted to the data of the experiment, and also on earlier data of Jansen & van der Maas (1997, 2002), show clear evidence of the consistency criterion: behavior of the latent classes concerned hardly deviates from the expected pattern of responses to the different item types for Rule I and Rule II. The application of latent class analysis formalizes the rate of sufficient consistency in terms of having good-fitting models, and the results indicate that children make few errors in applying Rule I and Rule II. Regularity, i.e. consistency over items of the same kind, was not tested explicitly in the present empirical study, but again it can be seen from the latent class analyses that the conditional probabilities in both the class associated with Rule I and the class associated with Rule II are very similar on both distance items and on both conflict-distance items. In sum, both consistency criteria lend credibility to ascribing rule-like beliefs to children by means of latent class analysis. In the literature about belief revision, it is common to assume that the formal belief states that are central to belief revision models, are idealizations of psychological states or cognitive states (see e.g. Gärdenfors, 1988). Reese's criteria, presented in the introductory sections, and the use of latent class models to test those criteria, provide an interesting way of making precise how much idealization is necessary in studying real-life development in terms of belief revision systems.

The results from the current study show that presenting distance items with increasing distances to children can induce the transition from Rule I to Rule II. Including feedback on those items was *not* found to be additionally effective as evidenced by identical transition matrices in the multi-group latent Markov models for both conditions. This is particularly interesting when describing this rule transition as a case of belief revision, in the sense that it provides a precise description of the epistemic inputs (Gärdenfors, 1988) that are necessary for causing the transition. In particular, the information that is apparently necessary to induce a rule transition only consists in increasing awareness that the distance dimension may be important in responding to balance scale items; however, providing feedback, which can be seen as providing information that is inconsistent with the current belief state, proved to be ineffective.

We showed that inducing the transition from Rule II to Rule III involves providing feedback about accuracy on particular types of items of the balance scale task. The

availability of such information constitutes the paradigmatic case of belief *revision*, in which information that is inconsistent with the current belief set induces a change in that belief set (Gärdenfors & Rott, 1995). On the contrary, the transition from Rule I to Rule II constitutes a different case of belief revision because in that case no inconsistent information is provided about the current belief set. Rather, the current belief is expanded with the notion that distance should be involved in deciding about balance scale problems, and the child adopts the simplest (in some to be defined hierarchy of complexity) way in which distance information can be incorporated in his/her strategy. Given our justification above for ascribing Rules I and II as beliefs, combined with evidence of discontinuity in the transition between these rules (Jansen & Van der Maas, 2001), leads us to conclude that this transition may be viewed as a form of belief revision.

Similar arguments could be applied to the transition from Rule II to Rule III, although the instability and lack of consistency of Rule III children makes the case less well founded. In particular, the description of Rule III users as having a firm belief may be questionable. Alternative rules may need to be formulated to accommodate the heterogeneity in this group of children. The main difference in the transitions between Rule I and II on the one hand, and between Rule II and III on the other, lies in the different types of information that is necessary to induce belief revision.

Piagetian developmental theory proposes that development progresses through a number of stages, and that transitions between these stages are discontinuous. Van der Maas and Molenaar (1992) provided detailed statistical criteria for testing discontinuity between developmental stages. The present results and the methods of belief revision combined may prove helpful in coming to grips with cognitive mechanisms that cause such discontinuous development. We are aware that the connections that we suggest between rule-transitions and belief revision are sketchy at best. At the same time, we are convinced that formalization of balance scale learning into the framework of belief revision can result in fruitful future research and cross-fertilization between philosophy and psychology.

6 General discussion

The Piagetian theory of cognitive development provides an excellent empirical domain to study computational models and belief revision models for many reasons. First, the theory is framed in terms of a progression through stages of ever more complex knowledge representations. The progression through these stages is, according to Piaget, governed by the processes of assimilation and accommodation, in which one's representational system gains representational power. This gain in representational power is evidenced by the availability of new concepts that are used to understand one's environment (Boden, 1994). Second, a number of well-studied tasks are available that are indicative of the current belief state of the subject. Centering on a single dimension of a given task, and ignoring other dimensions, is one of the most salient characteristics of the pre-operational stage. During progression to the operational stage, subjects go through a process of de-centering (Siegler, 1996). For the balance scale task, this means that the distance dimension becomes available for processing. Third, because of the well-studied tasks, many factors that influence performance on these tasks are known.

The present study highlights two of these factors. The findings are not only relevant for computational models but also for possible future formal models of belief revision. Such a model is not reported in the literature for the balance scale task but would be very interesting.

Most computational models and formal models of belief revision assume that a belief can be changed after an agent is confronted with information that conflicts his/her belief. The present findings indeed show that both children using Rule I and children using Rule II start to use a more complex rule after experiencing they were making mistakes. However, the influence of negative feedback is subtle as the transition from Rule II to Rule III seemed temporary and feedback did not seem necessary for children using Rule I. Both computational models and a formal model for belief revision on the balance scale task should include a mechanism that explains why children regress from using Rule III to Rule II, after feedback is withheld again. We conclude that additional mechanisms, next to the confrontation with negative feedback, are needed to explain rule transitions in balance scale reasoning.

A second possible mechanism is the perception of information that was ignored before. Both the findings of Jansen and van der Maas (2001) and the present findings demonstrate that emphasizing uninspected information, that may facilitate perception, may indeed initiate the transition from using Rule I to using Rule II (but not the transition from Rule II to a more complex rule).

This finding is of importance for computational models and for formal models of belief revision. Most computational models, except ACT-R, do not demonstrate learning without feedback. Also most formal models of belief revision only consider the possibility that beliefs are changed under influence of conflicting information. An exception is Paglieri's (2004) Data-oriented Belief Revision model. Here, data (information stored in the agent's mind) are differentiated from beliefs (information accepted as reliable). An agent focuses on a relevant subset of data on the basis of "their salience for the agent's purposes" (p. 183). Relevance depends on the goals of the agent and "a heuristic to define which data are to be considered more or less crucial for the achievement of a given goal" (p. 183). Paglieri does not explicitly define this heuristic.

The Data-oriented Belief Revision Model may be an accurate framework for a formal model of belief revision on the balance scale task. In such a model, selection of relevant data may depend on children's possibilities, increasing with age, of evaluating information about the state of the world and, in case of the balance scale task, of perceiving distance information. Emphasizing the distance dimension, as was done in this study, facilitates this process. Paglieri (2004) describes that after selecting the subset of data on basis of their relevance, data may be selected (or rejected) as beliefs on basis of reliability, credibility, importance, and likeability.

In the present research, we found that half of the children that did not use the distance dimension in their rule (i.e., they used Rule I) did use the dimension after observing the facilitating items. It would be interesting to investigate whether these children did consider the distance dimension already before the change. Perhaps they did not consider the data reliable, credible, important, or likeable enough.

A third mechanism that may initiate a change from one rule to a more complex rule may be the search for a difference between the left and the right side of the scale. This search forms the general solution strategy of the ACT-R model, which successfully replicates a number of empirical phenomena of balance scale learning. The search is influenced by the model's capacity of detecting differences and of the

Springer

available task properties. Later in development, more differences can be searched for and more properties become available. Empirical results that support the existence of this search are not available yet. However, it does seem plausible that children become more capable of considering various kinds of data. "Searching for differences" may also be the heuristic for evaluating the relevance of data in a formal model of belief revision.

The three mechanisms above, confrontation with negative feedback, improving perception and expanding search for differences, still do not provide an explanation for the transition from Rule II to Rule III or other complex rules. Rule II seems a quite stable rule. Experiencing they are making mistakes confuses children using Rule II, but the transition to Rule III does not persist. The progression to Rule III is more complex and it seems that children resort to trusting on the dominant dimension when a solution for the combination of variables is not offered. Again, our findings support the hypothesis that Rule III is not a proper rule, which offers children a reliable mental procedure for solving balance scale problems. It is not until they master new combinatory abilities that children progress from using Rule II to a true rule like the compensation-rule, or Rule IV.

It remains unknown how children start to incorporate addition abilities in their solution strategy or start using other compensation abilities. Children in our data set were probably able to sum numbers under ten (e.g., Siegler, 1998). But of course, this ability is only a small part of the compensation-rule. First, the child needs to calculate two sums (one for each side of the scale). Next, during the calculation of the second sum, the child needs to remember the result of the first sum. Finally, the child needs to correctly compare the results of the two sums. The complexity of the performance of the compensation-rule and a growing working memory capacity may explain why Jansen and van der Maas (2002) found that only around 10% of the children of 10 years old used the compensation-rule and that the compensation-rule was hardly observed before the age of 10 years.

Future research should be aimed at the prerequisites of rule transitions from Rule II to more complex rules, and insights from belief revision models could provide valuable starting points for such research. Until then, we concur with the citation of John Locke: it is one thing, to show a child who is using Rule II that he is in an Error, and another, to put him in possession of a more complex rule.

References

Boden, M. A. (1994). *Piaget*. London: Fontana Press.

Chletsos, P. N., De Lisi, R., Turner, G., & McGillicuddy-De Lisi, A. V. (1989). Cognitive assessment of proportional reasoning strategies. *Journal of Research and Development Education, 22*(4), 18–27.

Dawson, R. W., & Zimmerman, C. (2003). Interpreting the internal structure of a connectionist model of the balance scale task. *Brain and Mind, 4*, 129–149.

Dolan, C. V., Jansen, B. R. J., & van der Maas, H. L. J. (2004). Constrained and unconstrained multivariate normal finite mixture modeling of Piagetian data. *Multivariate Behavioral Research, 39*(1), 69–98.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development.* Cambridge: MIT Press.

Gärdenfors, P. (1988). *Knowledge in flux: Modeling the dynamics of epistemic states*. Cambridge, MA: The MIT Press.

Gärdenfors, P., & Rott, H. (1995). Belief Revision. In D. M. Gabbay, C. J. Hogger, & J. A. Robinson, (Eds.), *Handbook of logic in artificial intelligence and logic programming, Vol. 4 Epistemic and temporal reasoning* (pp. 35–132). Oxford: Clarendon Press.

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

Halford, G. S., Andrews, G., Dalton, C., Boag, C., & Zielinski, T. (2002). Young children's performance on the balance scale: The influence of relational complexity. *Journal of Experimental Child Psychology, 81*, 417–445.

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

Jansen, B. R. J. (2001). *Development of reasoning on the balance scale task: Psychometric assessment of cognitive strategies*. Dissertation Series/Faculty of Psychology, University of Amsterdam, The Netherlands.

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*(4), 450–494.

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

Klahr, D., & Siegler, R. S. (1978). The representation of children's knowledge. *Advances in Child Development and Behavior, 12*, 61–116.

Langeheine, R. (1994). Latent variables Markov models. In A. Von Eye & C. C. Clogg (Eds.), *Latent variables analysis*. Thousand Oaks, USA: Sage Publications.

Langley, P. (1987). A general theory of discrimination learning. In D. Klahr, P. Langley, & R. Neches (Eds.), *Production system models of learning and development* (pp. 99–161). Cambridge, MA: MIT Press.

Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. Boston: Houghton Mifflin.

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.

McClelland, J. L. (1995). A connectionist perspective on knowledge and development. In T. J. Simon, & G. S. Halford (Eds.), *Developing cognitive competence: New approaches to process modeling* (pp. 157–204). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Normandeau, S., Larivée, S., Roulin, J.-L., & Longeot, F. (1989). The balance-scale dilemma: Either the subject or the experimenter muddles through. *Journal of Genetic Psychology, 150*, 237–250.

Paglieri, F. (2004). Data-oriented belief revision: Towards a unified theory of epistemic processing. In E. Onaindia, S. Staab, *STAIRS 2004: 2nd Starting AI researchers' symposium* (pp. 179–190). Amsterdam: IOS Press.

Quinlan, P. T., van der Maas, M., Jansen, B. R. J., Booij, O., & Rendell, H. L. J. (in press). Rethinking stages of cognitive development: An appraisal of connectionist models of the balance scale task. *Cognition.* DOI 10.106/j.cognition.2006.02.004

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of IEEE, 77*(2), 267–295.

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*, 101–136.

Reese, H. W. (1989). Rules and rule-governance: Cognitive and behavioristic views. In S. C. Hayes (Ed.), *Rule-governed behavior: Cognition, contingencies, and instructional control* (pp. 3–84). New York: Plenum Press.

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

Sage, S. & Langley, P. (1983). Modeling cognitive development on the balance scale task. In A. Bundy (Ed.), *Proceedings of the eight international joint conference on artificial intelligence* (vol. 1, pp. 94–96). Karlsruhe, Germany.

Schmidt, W. C., & Ling, C. X. (1996). A decision-tree model of balance scale development. *Machine Learning, 24*, 203–230.

Shultz, T. R. (2003). *Computational developmental psychology*. Cambridge, MA: MIT Press.

Shultz, T. R., Mareschal, D., & Schmidt, W. C. (1994). Modeling cognitive development on balance scale phenomena. *Machine Learning, 16*, 57–86.

Shultz, T. R., & Schmidt, W. C. (1991). A cascade-correlation model of balance scale phenomena. *Proceedings of the thirteenth annual conference of the cognitive science society* (pp. 635–640). Hillsdale, NJ: Erlbaum.

Shultz, T. R., Schmidt, W. C., Buckingham, D., & Mareschal, D. (1995). Modeling cognitive development with a generative connectionist algorithm. In T. J. Simon, & G. S. Halford (Eds.), *Developing cognitive competence: New approaches to process modeling* (pp. 205–261). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

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

Siegler, R. S. (1981). Developmental sequences within and between concepts. *Monographs of the Society for Research in Child Development, 46*(2, Serial No. 189).

Siegler, R. S. (1996). Unidimensional thinking, multidimensional thinking, and characteristic tendencies of thought. In A. J. Sameroff, & M. M. Haith (Eds.), *The five to seven year shift: The age of reason and responsibility. The John D. and Catherine T. MacArthur Foundation series on mental heath and development* (pp. 63–84). Chicago, IL, USA: University of Chicago Press.

Siegler, R. S. (1998). *Children's thinking, third edition.* New Jersey, USA: Prentice-Hall, Inc.

Siegler, R. S., & Chen, Z. (1998). Developmental differences in rule learning: A microgenetic analysis. *Cognitive Psychology, 36,* 273–310.

Siegler, R. S., & Chen, Z. (2002). Development of rules and strategies: Balancing the old and the new. *Journal of Experimental Child Psychology, 81*, 446–457.

Siegler, R. S., & Stern, E. (1998). Conscious and unconscious strategy discoveries: A microgenetic analysis. *Journal of Experimental Psychology: General, 127*(4), 377–397.

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 Psychology, 85*, 141–177.

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 der Maas, H. L. J., & Molenaar, P. C. M. (1996). Catastrophe analysis of discontinuous development. In A. von Eye, & C. C. Clogg (Eds.), *Categorical variables in developmental research. Methods of analysis*. San Diego, CA: Academic Press.

Van Maanen, L., Been, P. H., & Sijtsma, K. (1989). The linear logistic test model and heterogeneity of cognitive strategies. In E. E. Roskam (Ed.), *Mathematical psychology in progress* (pp. 267–288). Berlin, Germany: Springer-Verlag.

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.