Detecting Strategies in Developmental Psychology

Helen Steingroever¹ · Marieke Jepma¹ · Michael D. Lee² · Brenda R. J. Jansen¹ · Hilde M. Huizenga¹

Published online: 28 January 2019

© The Author(s) 2019

Abstract

Differential strategy use is a topic of intense investigation in developmental psychology. Questions under study are as follows: How do strategies change with age, how can individual differences in strategy use be explained, and which interventions promote shifts from suboptimal to optimal strategies? In order to detect such differential strategy use, developmental psychology currently relies on two approaches—the rule assessment methodology and latent class analysis—each having their own strengths and weaknesses. In order to optimize strategy detection, a new approach that combines the strengths of both existing methods and avoids their weaknesses was recently developed: model-based latent-mixture analysis using Bayesian inference. We performed a simulation study to test the ability of this new approach to detect differential strategy use. Next, we illustrate the benefits of this approach by a re-analysis of decision making data from 210 children and adolescents. We conclude that the new approach yields highly informative results, and provides an adequate account of the observed data. To facilitate the application of this new approach in other studies, we provide open access documented code, and a step-by-step tutorial of its usage.

Keywords Qualitative development · Strategies · Decision making · Bayesian inference · Model based latent mixture analysis

Introduction

The concept of differential strategy use is ubiquitous in developmental psychology. A first line of developmental research focuses on whether strategy use changes with age. For example, a young child having to add three and four may count to three and then count four further, whereas an older child might retrieve the answer from memory (Ashcraft and Fierman 1982). There is thus a developmental shift in strategy use: from a counting strategy to a retrieval strategy. Such developmental shifts in strategy use have been reported in a variety of domains, for example in decision making (Aïte et al. 2012; Bereby-meyer et al. 2004; Betsch and Lang 2013; Huizenga et al. 2007; Jacobs and Potenza 1991; Jansen et al. 2012; Kwak et al. 2015; Lang

and Betsch 2018; Reyna 2008; Schlottmann 2001, 2000; Schlottmann and Anderson 1994), reasoning (Bouwmeester and Sijtsma 2007; Bouwmeester et al. 2004; Houdé et al. 2011; Jansen and van der Maas 2001, 2002; Siegler 1987, 2007; Siegler et al. 1981; van der Maas and Molenaar 1992; Wilkening 1981), reinforcement learning (Andersen et al. 2014; Decker et al. 2016; Palminteri et al. 2016; Potter et al. 2017; Schmittmann et al. 2012; Schmittmann et al. 2006), mathematics (Ashcraft and Fierman 1982; Bjorklund and Rosenblum 2001; Cho et al. 2011; Torbeyns et al. 2009), and categorization (Rabi et al. 2015; Raijmakers et al. 2004). A second line of research focuses on individual differences in strategy use within age groups (Siegler 1988), which have been shown to depend on variables such as intelligence (Bexkens et al. 2016) and the capacity to inhibit (Borst et al. 2012; Poirel et al. 2012). Finally, a third line of developmental research examines the effects of interventions aimed at changing children's or adolescents' sub-optimal strategies into optimal ones (Alibali 1999; Felton 2004; Perry et al. 2010; Stevenson et al. 2016). For these three types of developmental questions, it is important to detect differential strategy use adequately. Hence, the goal of this paper is to describe an adequate method to detect differential strategy use, which can easily be used by developmental psychologists.

An example of a paradigm in which strategy use is crucial is the balance scale task (e.g., Siegler and Chen 2002). In this task, participants have to decide whether a balance will tip, and if so, to which side (cf. Fig. 1). Many strategies can be

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s42113-019-0024-x>) contains supplementary material, which is available to authorized users.

✉ Marieke Jepma
m.jepma@uva.nl

¹ Department of Psychology, University of Amsterdam, PO box 15916, 1001 NK Amsterdam, The Netherlands

² Department of Cognitive Sciences, University of California Irvine, Irvine, CA, USA

Fig. 1 In the balance scale task, participants have to decide whether the balance will tip, and if so to which side, after the supporting blocks have been removed

used to solve this task. For example, some participants, using the optimal strategy, decide given the number of weights multiplied by their distance to the fulcrum. Other participants, using a suboptimal strategy, decide given the number of weights only. Still, other participants use another suboptimal strategy, in which they decide given the number of weights if weight differences are present, and decide given distance if weight differences are absent. Note that the particular item in Fig. 1 does not differentiate between the latter two suboptimal strategies, as they will give rise to the same answer (“right”). Therefore, studies investigating strategy use typically administer not only a single item, but a set of items. In a fictitious eight item example, the response pattern of a single participant may be “right, balance, balance, right, left, right, left, right.” The response patterns of multiple participants are then input to analyses aimed at detecting strategy use.

In developmental psychology, one approach to detect strategy use is the rule-assessment methodology (Siegler 1976; Siegler et al. 1981). According to this method, a participant is assigned to a particular strategy (“rule”), if the proportion of observed responses consistent with the predicted responses of that strategy exceeds a cut-off value. In the balance scale example, the observed response pattern (“right, balance, balance, right, left, right, left, right”) and predicted response pattern (e.g., “left, balance, balance, right, right, right, left, right”) match six out of eight (75%) responses. If the cut-off value was set at 70%, it thus is decided that this participant uses this particular strategy.

An advantage of the rule assessment methodology is that it is a one-step approach: Strategy use is inferred immediately (Table 1). Another advantage is that it can be used for small samples; even if only one child is tested, their strategy can, in principle, be inferred. However, there are also disadvantages. First, it does not account for error responses. That is, participants may not always make the response predicted by their strategy but may sometimes deviate from it, for example due to lapses of attention or premature

responding. This can make strategy assignment difficult. Second, the approach only allows for discrete strategy assignment: A participant is assigned to one strategy and not to others, which might be problematic if the fit to several strategies is approximately equal. Last but not least, the cut-off value (e.g., 70%) is arbitrary; hence, strategy assessment may change if the cut-off value is changed. For these reasons, it has been argued that this methodology may result in spurious detection of strategies (Jansen and van der Maas 2002; Thomas and Horton 1997; van der Maas and Straatemeier 2008; Wilkening 1988).

An alternative approach is based on latent-class analysis (also called latent-mixture analysis; Bouwmeester and Sijsma 2007; Bouwmeester and Verkoeijen 2012; Hickendorff et al. 2018; Huizenga et al. 2007; Jansen and van der Maas 2002; Nylund et al. 2007a). Although latent class analysis can be used in a confirmatory way, the exploratory way is more common. In this type of latent class analysis, participants are classified into several latent groups, based on similarity in observed response patterns. The number of latent groups required to describe the data is typically determined using penalized information criteria, such as the Bayesian Information Criterion (Schwartz 1978). Once the latent groups have been determined, the researcher evaluates the average response pattern in each latent group, comes up with a strategy that could explain this pattern, and assumes that all participants in that group used that strategy. If different possible strategies are specified beforehand, researchers can also compare the average response pattern in each latent group to the response pattern predicted by each of the pre-specified strategies, for example by computing the Euclidean distance between observed and predicted response patterns. The strategy that best fits the average response pattern of a latent group is then assigned to all participants in that group (Jansen et al. 2012).

The latent-class analysis approach has several advantages. A first advantage is that strategies are identified for average

Table 1 Characteristics of existing and new approaches to detect strategy use

	One step	Small samples	Accounts for errors	Continuous assignment	Non-arbitrary	Complexity correction
Rule assessment	+	+	–	–	–	–
Latent class	–	–	–	+	+	–
Latent mixture	+	+	+	+	+	+

response patterns within subgroups of participants. As average response patterns are affected less by error responses than individual ones, this may benefit adequate strategy assessment. A second advantage is that—in contrast to the rule-assessment methodology—strategy assignment does not depend on an arbitrary criterion. Third, although participants are generally assigned to their most likely latent group, a posteriori probabilities offer the opportunity to identify each individual's likelihood that they belong to each of the latent groups. Finally, as latent-class analysis does not require the specification of possible strategies beforehand, it is an ideal method when studying phenomena in which the number and nature of potential strategies are largely unknown, as in exploratory studies. There are also, however, disadvantages. First, latent-class analysis is a two-step procedure: Instead of directly defining subgroups in terms of common strategy use, it defines subgroups in terms of homogeneous response patterns, and the researcher subsequently assigns the most likely strategy to each subgroup. Second, it uses model selection criteria that have been criticized for not adequately accounting for model complexity (Burnham and Anderson 2002). Third, it does not account for individual differences in error rate. Fourth, it requires large sample sizes to guarantee stable solutions (Jaki et al. *in press*; Nylund et al. 2007b).

The goal of the current article is to introduce a recently developed alternative approach that is based on the advantages of both the rule assessment and the latent-class analysis methodologies while avoiding their disadvantages: a latent-mixture model implemented in a Bayesian framework (Lee and Sarnecka 2011; Lee 2016; Lee 2018a). Like the rule assessment method, latent-mixture models directly define groups in terms of common strategy use and can be used for any sample size. Furthermore, similar to latent-class analysis, this approach infers the likelihood that each participant uses each of the assumed strategies. Thus, it can simultaneously infer the number and size of strategy groups, each individual participant's group membership, and the certainty of each participant's group membership (Bröder and Schiffer 2003; Lee 2016). In addition, latent-mixture models provide an estimate of how accurately participants follow their inferred strategy, by explicitly formalizing individual differences in error rate. Finally, all conclusions are obtained using Bayesian inference, which quantifies uncertainty about parameters in terms of probability distributions, controls for all aspects of model complexity, and handles missing data in a comprehensive model-based way (e.g., Andrews and Baguley 2013; Bayarri et al. 2016; Lee and Wagenmakers 2005; Wagenmakers 2007; Wetzels et al. 2011, for an introduction to the use and advantages of Bayesian methods in the psychological sciences, see Lee 2018b; Lee and Wagenmakers 2013; Vandekerckhove et al. 2018; Wagenmakers et al. 2016).

Here, we apply the latent-mixture approach, implemented in a Bayesian framework, to infer differential strategy use on a

previously established decision-making task, the “Gambling-Machine Task” (Jansen et al. 2012). In the remainder of this article, we first describe the Gambling-Machine Task and various strategies that can be used to perform this task. Next, we describe the latent-mixture model and its implementation in a Bayesian framework, and assess its capacity to recover strategy-use parameters by applying it to generated data. Then, we apply the latent mixture model to Gambling-Machine Task data from 210 children and adolescents (Jansen et al. 2012). In the final section, we discuss our results and their ramifications.

The Gambling-Machine Task

The Gambling-Machine Task (Jansen et al. 2012) is a paper-and-pencil decision-making task in which participants choose between two gambling machines (i.e., a specific Machine A and a specific Machine B that together represent one of the 28 items of the task; an example trial is shown in Fig. 2). Each machine contains 10 balls. Each ball is associated with a specific reward that is printed on the left corner of each machine (i.e., +2 for Machine A and +4 for Machine B in Fig. 2). In addition, some balls, the gray shaded so-called loss balls, are associated with an additional loss—the exact amount is printed on the balls (i.e., −10 for both machines in Fig. 2). The two machines of each item differ in at least one of three overtly presented cues: the reward associated with each ball (certain gain; CG), the proportion of loss balls (frequency of loss; FL; 0.1 for Machine A and 0.5 for Machine B in Fig. 2), and the amount of loss associated with each loss ball (AL). On each trial, participants have to indicate whether Machine A or Machine B is more profitable, or whether it does not matter. The correct answer is the option with the highest expected value (with expected value = $CG + AL \times FL$).

Fig. 2 Example trial from the Gambling-Machine Task. In this trial, there is a conflict between frequency of loss and certain gain. Here, “Machine A” is the correct answer. Figure taken from Jansen et al. (2012)

The task is based on seven item types, each of which has four versions. The seven item types consist of three simple item types in which the machines differ in only one cue, and four complex item types in which the machines differ in at least two cues that produce a conflict. The different versions of each item type are obtained by slightly adapting the specific amounts of CG, AL, or FL, and by changing the position of Machine A and Machine B.

Decision Strategies

We assume that participants use one of 17 strategies to complete the Gambling-Machine Task (Table 2, first two columns). These strategies cover a guessing strategy, an integrative strategy focusing on the expected value of the machines, and 15 non-compensatory strategies based on take-the-best strategies (Bröder 2000). The guessing strategy assumes that participants randomly choose one of the two machines without considering any of their cues, such that on each trial, the answers “Machine A” and “Machine B” are equally likely. It is assumed that, when guessing, participants will never consider the response “does not matter.” The integrative strategy assumes that participants correctly integrate all relevant cues by computing the expected value of each machine and subsequently choose the machine with the highest expected value; if both machines have the same expected value, the option “Does not matter” is chosen. Thus, when following the integrative strategy, the correct answer is always given.

Non-compensatory strategies, such as take-the-best strategies, on the other hand, are sequential strategies that differ in the number of cues that are considered (i.e., the complexity) and the order in which the cues are considered (i.e., the saliency of the cues). According to one-dimensional take-the-best strategies, participants compare the options on only one cue, say FL. If the machines differ on FL, the machine with the lowest FL is chosen; otherwise, the answer “Does not matter” is given. According to two-dimensional take-the-best strategies, the machines are sequentially compared on the two most salient cues, say FL and AL. The second, less salient, cue is only considered if the machines do not differ on the first cue. If the machines do not differ on either of the two cues, the answer “Does not matter” is given. The three-dimensional strategies work analogously. We obtain 15 possible take-the-best strategies as the result of all combinations of the three cues, while allowing to sequentially consider either one, two, or three cues (Table 2).

These 17 strategies predict distinctive response patterns that are deterministic for all but the guessing strategy. Thus, a participant’s answers to the Gambling-Machine Task items can be used to infer which of the 17 strategies the participant used. To accomplish this, we use a latent-mixture model implemented in a Bayesian framework.

The Latent-Mixture Model

The latent-mixture approach assumes that each individual uses one of the 17 decision strategies, such that the overall data set

Table 2 Overview of 17 possible decision strategies for completion of the Gambling-Machine Task, and percentage of participants assigned to each strategy according to our latent-mixture model, the latent-class analysis used by Jansen et al. (2012), and the rule-assessment method using two different cut-off values

Model	TTB order	Latent-mixture model	LCA	RAM (70%)	RAM (90%)
Guessing		0.48	5.71		
Integrative		31.43	27.61	30.08	15.71
TTB ₁	FL > AL > CG	27.62	25.24	24.09	9.76
TTB ₂	FL > CG > AL	14.29	20.95	13.93	5.95
TTB ₃	AL > FL > CG	13.33	9.52	14.29	4.29
TTB ₄	AL > CG > FL	0.95	0	1.11	0.48
TTB ₅	CG > FL > AL	0.48	0	0.48	0
TTB ₆	CG > AL > FL	0	0	0	0
TTB ₇	FL > AL	4.29	10.95	5.52	2.38
TTB ₈	FL > CG	0.95	0	0.71	0
TTB ₉	AL > FL	1.9	0	1.35	0.95
TTB ₁₀	AL > CG	0.95	0	0.95	0
TTB ₁₁	CG > FL	0	0	0	0
TTB ₁₂	CG > AL	0.48	0	0.24	0
TTB ₁₃	FL	0.95	0	1.07	0.48
TTB ₁₄	AL	1.43	0	0.48	0.48
TTB ₁₅	CG	0.48	0	0.48	0.48

Note that some participants were not assigned to any of the strategies according to the rule-assessment method with a 90% cut-off value

TTB take-the-best, FL frequency of loss, AL amount of loss, CG certain gain, LCA latent-class analysis, RAM rule-assessment method

reflects a mixture of these specific strategies. Figure 3 shows a graphical representation of our model. The nodes stand for model variables and data, and the arrows indicate how the model variables are assumed to generate the data. Square nodes represent discrete variables and round nodes represent continuous variables. Nodes are shaded when the corresponding variables are observed but are unshaded when the variables are latent or unknown. Nodes with double borders indicate deterministic variables. The plates show independent repetitions in the graph structure across all items and participants.

The data $y_{i,q}$ are participants' answers to the Gambling-Machine Task items, with i referring to a specific participant and q to one of the 28 items; $y_{i,q}$ can be either "A" (Machine A), "B" (Machine B), or "DM" (Does not matter). The data are represented by a shaded, square node because they are observed, discrete choices.

The model assumes that each participant's data is generated by one of the 17 strategies. The specific strategy used by participant i is indicated by the latent categorical variable z_i , which can take one out of 17 values. Note that z_i is represented by an unshaded, square node because it is latent and discrete. All of the strategies, except for the guessing strategy, generate deterministic predictions of participants' answers to each item. These predictions are denoted by $t_{i,q}$. The predictions depend on the cues of item q (i.e., the machines' certain gain, amount of loss, and frequency of loss) which are labeled as s_q . To account for possible deviations from the deterministic predictions, the model assumes that each participant has an individual error rate ϵ_i (Rieskamp 2008) that is used to convert the predictions into answer probabilities $p_{i,q}$. Specifically, each participant chooses the option predicted by his/her strategy with probability $1-2\epsilon_i$, and each of the two remaining options with a probability of ϵ_i . Thus, the higher the error rate, the more likely decisions are to deviate from the deterministic

predictions. We did not implement an error rate for participants who use the guessing strategy but assumed that these participants are equally likely to choose both machines on each trial (i.e., $p_{A,i,q} = p_{B,i,q} = 1/2$, $p_{DM,i,q} = 0$). Finally, the obtained probabilities $p_{i,q}$ specify the categorical distribution that generates the observed data, that is, $y_{i,q} \sim \text{Categorical}(p_{A,i,q}, p_{B,i,q}, p_{DM,i,q})$.

Bayesian Implementation

We inferred the posterior distributions for each participant's model parameters, z_i and ϵ_i (Lee and Wagenmakers 2013; Vandekerckhove et al. 2018), using the following priors. We assumed that all 17 strategies are a priori equally likely; hence, we specified a categorical prior on z_i that can take values 1 to 17 to indicate which strategy participant i used for all items. For the individual error rate ϵ_i , we chose a Uniform(0, 0.25) prior. An error rate of zero indicates that strategies are followed faultlessly, and an error rate of 0.25 indicates that there is a 50% chance of choosing the option predicted by the strategy and a 25% chance of choosing either of the remaining two options.

Applying the model to the observed data yields, for each participant, the posterior probabilities that each of the 17 strategies was used. In addition, for each participant, we can infer how closely the assigned strategy is followed.

Methods

Simulation Study

To examine our model's ability to recover data-generating parameters, we first applied the model to generated data and compared the inferred parameters to the data-generating

Fig. 3 A graphical representation of the Bayesian latent-mixture model to infer which of 17 decision-making strategies each participant uses to complete the Gambling-Machine Task

$$z_i \sim \text{Categorical}\left(\frac{1}{17}, \dots, \frac{1}{17}\right)$$

$$t_{i,q} = \begin{cases} \text{EV}(s_q) & \text{if } z_i = 2 \\ \text{TTB}_1(s_q) & \text{if } z_i = 3 \\ \vdots & \vdots \\ \text{TTB}_{15}(s_q) & \text{if } z_i = 17 \end{cases}$$

$$\epsilon_i \sim \text{Uniform}(0, 0.25)$$

$$p_{A,i,q} \sim \begin{cases} 1/2 & \text{if } z_i = 1 \\ 1-2\epsilon_i & \text{if } t_{i,q} = A \\ \epsilon_i & \text{otherwise} \end{cases}$$

$$p_{B,i,q} \sim \begin{cases} 1/2 & \text{if } z_i = 1 \\ 1-2\epsilon_i & \text{if } t_{i,q} = B \\ \epsilon_i & \text{otherwise} \end{cases}$$

$$p_{DM,i,q} \sim \begin{cases} 0 & \text{if } z_i = 1 \\ 1-2\epsilon_i & \text{if } t_{i,q} = DM \\ \epsilon_i & \text{otherwise} \end{cases}$$

$$y_{i,q} \sim \text{Categorical}(p_{A,i,q}, p_{B,i,q}, p_{DM,i,q})$$

values (see “Results” section for the results). For each of the 17 decision strategies, we generated data for 15 synthetic participants, resulting in 255 synthetic participants. For all participants, except for those using the guessing strategy, we generated an error rate ϵ_i by drawing from a Uniform(0, 0.25) distribution. Each synthetic participant completed all 28 items of the Gambling-Machine Task.

Real Data Set

We then reanalyzed the Gambling-Machine Task data of 230 participants previously published by Jansen et al. (2012).¹ The data set comprises four age groups: 8–11 ($n = 48$), 11–12 ($n = 51$), 12–15 ($n = 67$), and 14–17 years ($n = 64$; see Jansen et al. 2012, for more details). After having removed 20 participants because of missing data entries ($n = 9, 5, 5, 1$ in age groups 1, 2, 3, 4, respectively), we applied the model to 210 participants. We did not obtain Ethics Committee approval because we reanalyzed previously published data.

Model Fitting Procedure

We inferred posterior distributions for each parameter for each participant using Markov chain Monte Carlo (MCMC) sampling, which directly samples sequences of values from the posterior distribution for each model parameter (Gilks et al. 1996; Lunn et al. 2012), as implemented in JAGS (Plummer 2003) and R (R Development Core Team 2008). We ran 10 independent MCMC chains and initialized all chains with random values. For the real dataset, we collected 45,000 samples per chain and discarded the first 5000 samples as burn-in. In addition, we only used every fifth iteration to remove autocorrelation. Consequently, we obtained 80,000 representative samples per parameter (8000 per chain) per participant. In the simulation study, we collected 20,000 samples per chain and discarded the first 10,000 as burn-in. As for the real dataset, we used a thin rate of 5, resulting in a total of 20,000 representative samples per parameter per synthetic participant.

All chains showed convergence (see the [online supplementary material](#) for an assessment of the convergence and stability of our results). R and JAGS code, and the data, can be downloaded from our Open Science Framework repository ([goo.gl/tkh4gE](https://github.com/tk4gE)). In addition, to facilitate the use of this method by others, all steps of the modeling analysis are summarized in the [online supplementary material](#).

Strategy Inference

We computed the mode of the posterior distribution of z_i to infer, for each participant, which strategy they most likely

used. In addition, to infer the certainty of the strategy assignment, we computed the proportion of z_i samples that equal the mode of the posterior distribution of z_i . We computed the mode of the posterior distribution of ϵ_i to infer each participant’s error-rate parameter.

Assessment of Model Fit

To assess the descriptive adequacy of the model—that is, its ability to “fit” the observed choice data—we conducted a standard Bayesian posterior predictive check. Specifically, we generated data for each participant using the mode of the posterior distributions of their z_i and ϵ_i parameters, and subsequently compared the generated and observed data.

Results

Parameter Recovery

Figure 4 shows the strategy assigned to each synthetic participant in our simulation study. The synthetic participants are ordered in such a way that the data from the first 15 participants were generated with the guessing strategy, the data for the second 15 with the integrative strategy (expected value; EV), etc. We can speak of perfect recovery if the first 15 bars are all of height “guessing,” the second 15 bars all of height “EV,” etc. Figure 4 clearly shows this expected step-wise pattern. Specifically, 81.6% of the synthetic participants were classified correctly. Note that recovery is perfect if we use a small error rate to generate data (e.g., $\epsilon_i = .05$ for all participants), clearly indicating the correctness of our implementation of the model. However, we prefer discussing a more realistic scenario with different error rates for each participant that vary between 0 and .25.

Figure S1 in the online supplementary material shows the posterior distributions of the error-rate parameter as well as the true data-generating value for 15 randomly selected synthetic participants. The mode of the posterior distribution is close to the data-generating value for some participants, indicating an adequate recovery. However, for other participants, there is a noticeable difference between the posterior mode and the data-generating value, indicating inadequate recovery. There was no systematic bias in the recovery of the individual error rates, nor a systematic relation between the bias and the standard deviation of the posterior distribution of the individual error rates (see online supplementary material; Fig. S2). Since there is no systematic bias, but a noticeable difference between the true value and the recovered value of the individual error rate for many synthetic participants, we recommend interpreting the estimated error rate at a group-level and considering individual error rates with caution.

¹ Jansen et al. (2012) reported one additional participant due to accidentally having included a participant with completely missing data.

Fig. 4 Strategy assignment for the 255 synthetic participants. The first 15 participants were generated according to the guessing strategy, the second 15 participants according to the integrative (EV) strategy, etc. With ideal recovery, the y value would increase one unit after every 15 participants (i.e., after every color change).

Real Data Set: Inferred Strategy Use

The third column of Table 2 reports the percentage of the 210 real participants assigned to each of the 17 strategies according to our model. The majority of the participants used the integrative strategy, correctly integrating all three relevant cues to compute the expected value of the machines, followed by the $FL > AL > CG$ and $FL > CG > AL$ take-the-best strategies. In general, among the take-the-best strategies, FL seems to be a very important cue: 70.1% of the used take-the-best strategies consider FL as the first cue. It is also evident that none of the participants used the $CG > AL > FL$ and $CG > FL$ strategies.

Figure 5 shows the posterior inferences for strategy use, organized by the most likely strategy, for the six largest groups; the remaining 11 strategies were used by at most three participants. The header of each sub-panel contains the most likely strategy and the number of participants assigned to that strategy. The area of the circles represents the certainty of the assignment (i.e., the proportion of z_i samples that equal the mode of the z_i posterior): The larger the circle, the higher the certainty of the corresponding strategy assignment. It is evident that most participants were assigned with high certainty, but for some participants, related strategies are also plausible. For example, for the first participant that was assigned to the $FL > AL$ group, the FL strategy is also very plausible.

Figure 6 shows the developmental trajectory of the six most often used strategies. With increasing age, integrative strategy use decreased, and the $FL > AL > CG$ strategy use increased. Clear developmental changes with respect to the four remaining strategies are not evident.

Inferred error rates

The average inferred error rate was 0.08 (SD = 0.06), suggesting that, on average, participants chose the option predicted by their specific strategy on 84% of the trials and chose each of the two remaining options on 8% of the trials. Inferred error rates for individual participants varied between 0.002 and

0.23. However, as our simulation study showed that many individual error rates were not recovered accurately, these should be considered with caution.

Posterior Predictive Check

Figure 7 shows the posterior predictive check for four randomly selected participants.² In this figure, observed choices are represented by solid lines and generated choices by dotted lines. Filled dots indicate that the generated and observed choice coincide. The header of each subplot shows the assigned strategy and the certainty of this assignment.

To summarize the descriptive adequacy of the model across all participants, we computed for each participant the prediction error (i.e., the proportion of items for which the generated and observed choice differ). Figure 8 shows the distribution of the prediction error across all participants. The mean prediction error is 0.135, and the prediction error is below 0.20 for most participants. This suggests that, in general, the model adequately accounts for the data and that we can draw meaningful conclusions from the model's strategy indicator parameter.

Comparison of Results from the Bayesian Latent-Mixture Model and Existing Methods

The fourth, fifth, and sixth columns of Table 2 show the percentage of participants assigned to each strategy by the latent-class analysis reported by Jansen et al. (2012) and by the rule-assessment method using two different cut-off values. The two largest inferred strategy groups are the same for all three methods. However, the results obtained by the different methods differ in some crucial ways as well. First of all, we detected 15 strategy groups, whereas Jansen et al.'s (2012) latent-class analysis detected only six groups (i.e., a model with six latent classes provided the best fit to the data, according to the Bayesian Information Criterion), and the number of

² Figures for all participants are available at goo.gl/tkh4gE

Fig. 5 Posterior inferences for strategy use for the majority of the participants ($n = 195$) obtained from Jansen et al. (2012, $n = 210$). Results are only presented for the six largest groups; the remaining nine groups contain one, two, or three participants. The area of each circle corresponds to the posterior probability that a participant used a particular strategy

groups detected by the rule-assessment method depends on the cut-off value that is used. Second, the percentage of participants assigned to each group differs across the three approaches. The differences between our method and the latent-class analysis must be caused by the differences of these two modeling approaches. These differences include our use of the Bayesian rather than frequentist framework, the explicit as opposed to no explicit formalization of individual differences in error rates, and the assumption that each participant uses one out of 17 decision strategies and deviates from that strategy by an individual error rate compared to the assumption that participants with the same response pattern use the same strategy. Finally, our model outperforms the latent-class analysis on the posterior predictive check: Our model yields posterior predictions that differ from the observed data for on average 13.5% of the items, compared to 16.7% of the items

for the model of Jansen et al. (2012).³ This suggests that in addition to yielding more informative results, the Bayesian latent-mixture model is more descriptively accurate.

Discussion

In this paper, we presented a Bayesian latent-mixture model approach to infer strategy use during a decision-making task. Our approach circumvents shortcomings of the traditional rule-assessment methodology (Siegler 1976; Siegler et al. 1981) and the latent-class analysis approach (Clogg 1995; Dolan et al.

³ We were able to generate posterior predictions based on the inferences of Jansen et al. (2012) because the authors kindly provided us with the data and the classification of each participant.

Fig. 6 Proportion of participants who used specific strategies as a function of age group. Results are only shown for the six most often used strategies

2004; Heinen 1996; Huizenga et al. 2007; Jansen and van der Maas 2002; Jansen et al. 2012; McCutcheon 1987; Rindskopf 1983, 1987) by allowing for the simultaneous inference of the number and size of strategy groups, as well as the strategy that each individual participant most likely used. In addition, our approach includes individual error rates that systematically account for individual inconsistencies in following the assigned decision strategy. Finally, it does not require large sample sizes.

We illustrated our latent-mixture model approach by applying it to a data set comprising 210 children and adolescents completing the Gambling-Machine Task (Jansen et al. 2012). Our latent-mixture model combined 17 different accounts of decision making (i.e., guessing strategy, integrative strategy, and 15 different take-the-best strategies) and allowed us to infer which of these 17 strategies every participant used. We found that most participants used the integrative strategy, followed by three-dimensional take-the-best strategies that have frequency of loss (FL) and amount of loss (AL) as most salient cues. In addition, our results suggest that with increasing age, integrative strategy use decreases, but the use of the FL > AL > CG

strategy increases. These findings are consistent with the results of the original analysis from Jansen et al. (2012), and seemingly different from typical findings of proportional-reasoning tasks with two attributes, such as the balance task. As discussed by Jansen et al. (2012), the decrease in integrative strategy use with age supports the fuzzy-trace theory (Reyna and Ellis 1994) which suggests a developmental increase in the dominance of gist (fuzzy) over verbatim (detailed) processing.

A major advantage of our latent-mixture model approach is that it is very flexible. The online provided code can be modified in several ways. For example, the reader can include additional decision strategies (e.g., the weighted additive model; Gigerenzer et al. 1999; Rieskamp and Otto 2006) or exclude some of the strategies. It is also straightforward to modify the prior distributions and test the impact of such modifications as a form of sensitivity analysis. Other plausible modifications are adaptations of the model to other developmental domains and tasks, for example focusing on differential strategy use when solving categorization, logical-reasoning or mathematical problems. In addition, the model could be applied in longitudinal studies to investigate developmental changes in strategy use within individual participants.

More challenging modifications and suggestions for future research involve incorporating a more theoretically motivated model of probabilistic responding given deterministic choice rules than the simple error-rate mechanism used here (Regenwetter et al. 2011). In future work, the model could also be fitted hierarchically, such that the model parameters of each participant are sampled from a group-level distribution (Farrell and Lewandowsky 2018; Lee and Wagenmakers 2013; Rouder and Lu 2005; Shiffrin et al. 2008). In a Bayesian framework, this entails that each individual-level parameter is assigned a group-level prior distribution, whose parameters (hyperparameters) are estimated as well. One advantage of a hierarchical Bayesian implementation would be that it allows for the inclusion of covariates in the model (e.g., age, gender, intelligence) such that the effects of relevant individual-difference variables on each parameter can be directly tested. In addition, it is desirable to modify our model to account for the fact that children might learn during the task; for example, just by considering the different test items, children may discover new features and therefore change their strategy (e.g., Boncoddio et al. 2010; Jansen and van der Maas 2002). This could be done by using standard Bayesian learning applied to the valence of each cue (Mistry et al. 2016). We also note that a generative assumption of our model is that each participant uses the same strategy throughout the task. This assumption is violated if strategy use depends on specific item characteristics. The method can, however, easily be extended to include such item-specific strategy use.

Finally, it should be noted that despite the advantages of the proposed mixture model and the straightforward adaptations of the code to other application areas, caution is needed when

Fig. 7 Posterior predictive check for four randomly selected participants. The dotted and solid line represent the generated and observed choices, respectively. Filled dots indicate that the generated and observed choices are identical. The header indicates the most likely strategy and the posterior probability that the participant used that strategy. DM stands for the answer option “Does not matter”

Fig. 8 Summary of the posterior predictive check of all participants. The figure shows the prediction error of all participants. The prediction error is defined as the proportion of items for which the predictions differ from the observed choices

inferences are drawn. It is important to verify that the model provides a sufficient account of the data and that the results are stable, and in case of doubts, this should be made explicit. For example, participants' responses to all items are weighted equally in the model, whereas in actual datasets, it is possible that some task items are more important to distinguish strategies than others. At the same time, the reader should be aware that a latent-mixture approach will always yield a mixture of partially overlapping distributions. Therefore, although each participant is assigned to the most likely group, in some cases, another group is nearly as likely. For these reasons, a mixture analysis should always be performed and interpreted with caution (Huizenga et al. 2007).

Developmental change in strategy use is a popular topic in the developmental sciences. Inferences have been based on application of the rule-assessment methodology proposed by Siegler (Siegler 1976; Siegler et al. 1981) or latent-class analysis (Clogg 1995; Heinen 1996; McCutcheon 1987;

Rindskopf 1983, 1987). Here, we proposed a model-based approach using latent-mixture models and Bayesian inference. Given the advantages of our approach—its highly informative inference and flexibility—we hope it will be applied in future studies to advance research areas covering the development of decision making, categorization, logical reasoning, and mathematical concepts.

Funding Information HS, MJ, and HMH were supported by a VICI grant of the Netherlands Organization for Scientific Research (number 453-12-005).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Aïte, A., Cassotti, M., Rossi, S., Poirel, N., Lubin, A., Houdé, O., & Moutier, S. (2012). Is human decision making under ambiguity guided by loss frequency regardless of the costs? A developmental study using the Soochow Gambling Task. *Journal of Experimental Child Psychology*, 113, 286–294. <https://doi.org/10.1016/j.jecp.2012.05.008>.
- Alibali, M. W. (1999). How children change their minds: Strategy change can be gradual or abrupt. *Developmental Psychology*, 35(1), 127–145. <https://doi.org/10.1037/0012-1649.35.1.127>.
- Andersen, L. M., Visser, I., Crone, E. A., Koolschijn, P. C. M. P., & Raijmakers, M. E. J. (2014). Cognitive strategy use as an index of developmental differences in neural responses to feedback. *Developmental Psychology*, 50(12), 2686–2696. <https://doi.org/10.1037/a0038106>.
- Andrews, M., & Baguley, T. (2013). Prior approval: The growth of Bayesian methods in psychology. *British Journal of Mathematical and Statistical Psychology*, 66, 1–7. <https://doi.org/10.1111/bmsp.12004>.
- Ashcraft, M. H., & Fierman, B. A. (1982). Mental addition in third, fourth, and sixth graders. *Journal of Experimental Child Psychology*, 33(2), 216–234. [https://doi.org/10.1016/0022-0965\(82\)90017-0](https://doi.org/10.1016/0022-0965(82)90017-0).
- Bayarri, M. J., Benjamin, D. J., Berger, J. O., & Sellke, T. M. (2016). Rejection odds and rejection ratios: A proposal for statistical practice in testing hypotheses. *Journal of Mathematical Psychology*, 72, 90–103. <https://doi.org/10.1016/j.jmp.2015.12.007>.
- Bereby-meyer, Y., Assor, A., & Katz, I. (2004). Children's choice strategies: The effects of age and task demands. *Cognitive Development*, 19, 127–146. <https://doi.org/10.1016/j.cogdev.2003.11.003>.
- Betsch, T., & Lang, A. (2013). Utilization of probabilistic cues in the presence of irrelevant information: A comparison of risky choice in children and adults. *Journal of Experimental Child Psychology*, 115, 108–125. <https://doi.org/10.1016/j.jecp.2012.11.003>.
- Bexkens, A., Jansen, B. R. J., van der Molen, M. W., & Huizenga, H. M. (2016). Cool decision-making in adolescents with behavior disorder and/or mild-to-borderline intellectual disability. *Journal of Abnormal Child Psychology*, 44(2), 357–367. <https://doi.org/10.1007/s10802-015-9996-8>.
- Bjorklund, D. F., & Rosenblum, K. E. (2001). Children's use of multiple and variable addition strategies in a game context. *Developmental Science*, 4, 184–194. <https://doi.org/10.1111/1467-7687.00164>.
- Boncoddo, R., Dixon, J. A., & Kelley, E. (2010). The emergence of a novel representation from action: Evidence from preschoolers. *Developmental Science*, 13, 370–377. <https://doi.org/10.1111/j.1467-7687.2009.00905.x>.
- Borst, G., Poirel, N., Pineau, A., Cassotti, M., & Houdé, O. (2012). Inhibitory control in number-conservation and class-inclusion tasks: A neo-Piagetian inter-task priming study. *Cognitive Development*, 27(3), 283–298. <https://doi.org/10.1016/j.cogdev.2012.02.004>.
- Bouwmeester, S., & Sijsma, K. (2007). Latent class modeling of phases in the development of transitive reasoning. *Multivariate Behavioral Research*, 42(3), 457–480. <https://doi.org/10.1080/00273170701384324>.
- Bouwmeester, S., & Verkoeijen, P. P. (2012). Multiple representations in number line estimation: A developmental shift or classes of representations? *Cognition and Instruction*, 30(3), 246–260.
- Bouwmeester, S., Sijsma, K., & Vermunt, J. K. (2004). Latent class regression analysis for describing cognitive developmental phenomena: An application to transitive reasoning. *European Journal of Developmental Psychology*, 1(1), 67–86. <https://doi.org/10.1080/17405620344000031>.
- Bröder, A. (2000). Assessing the empirical validity of the “take-the-best” heuristic as a model of human probabilistic inference. *Journal of Experimental Psychology Learning, Memory and Cognition*, 26(5), 1332–1346.
- Bröder, A., & Schiffer, S. (2003). Bayesian strategy assessment in multi-attribute decision making. *Journal of Behavioral Decision Making*, 16(3), 193–213. <https://doi.org/10.1002/bdm.442>.
- Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. New York: Springer.
- Cho, S., Ryali, S., Geary, D. C., & Menon, V. (2011). How does a child solve 7+8? Decoding brain activity patterns associated with counting and retrieval strategies. *Developmental Science*, 14(5), 989–1001. <https://doi.org/10.1111/j.1467-7687.2011.01055.x>.
- Clogg, C. C. (1995). Latent class models. In G. Arminger, C. C. Clogg, & M. E. Sobel (Eds.), *Handbook of statistical modeling for the social and behavioral sciences*. New York: Plenum.
- Decker, J. H., Otto, A. R., Daw, N. D., & Hartley, C. A. (2016). From creatures of habit to goal-directed learners: Tracking the developmental emergence of model-based reinforcement learning. *Psychological Science*, 27(6), 848–858. <https://doi.org/10.1177/0956797616639301>.
- van der Maas, H. L. J., & Molenaar, P. C. M. (1992). Stages of cognitive development: An application of catastrophe theory. *Psychological Review*, 99, 395–417. <https://doi.org/10.1037/0033-295X.99.3.395>.
- van der Maas, H. L., & Straatemeier, M. (2008). How to detect cognitive strategies: Commentary on ‘differentiation and integration: Guiding principles for analyzing cognitive change’. *Developmental Science*, 11(4), 449–453.
- Dolan, C. V., Jansen, B. R., & van der Maas, H. L. (2004). Constrained and unconstrained multivariate normal finite mixture modeling of Piagetian data. *Multivariate Behavioral Research*, 39, 69–98.
- Farrell, S., & Lewandowsky, S. (2018). *Computational modeling of cognition and behavior*. Cambridge: Cambridge University Press.
- Felton, M. K. (2004). The development of discourse strategies in adolescent argumentation. *Cognitive Development*, 19(1), 35–52. <https://doi.org/10.1016/j.cogdev.2003.09.001>.
- Gigerenzer, G., Todd, P. M., & ABC Research Group. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.
- Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (Eds.). (1996). *Markov chain Monte Carlo in practice*. Boca Raton, FL: Chapman & Hall/CRC Press.
- Heinen, T. (1996). *Latent class and discrete latent trait models: Similarities and differences*. Thousand Oaks, CA: Sage.

- Hickendorff, M., Edelsbrunner, P. A., McMullen, J., Schneider, M., & Trezise, K. (2018). Informative tools for characterizing individual differences in learning: Latent class, latent profile, and latent transition analysis. *Learning and Individual Differences*, 66, 4–15.
- Houdé, O., Pineau, A., Leroux, G., Poirel, N., Perchey, G., Lanoë, C., Lubin, A., Turbelin, M. R., Rossi, S., Simon, G., Delcroix, N., Lamberton, F., Vigneau, M., Wisniewski, G., Vicet, J. R., & Mazoyer, B. (2011). Functional magnetic resonance imaging study of Piaget's conservation-of-number task in preschool and school-age children: A neo-Piagetian approach. *Journal of Experimental Child Psychology*, 110(3), 332–346. <https://doi.org/10.1016/j.jecp.2011.04.008>.
- Huizenga, H. M., Crone, E. A., & Jansen, B. R. J. (2007). Decision-making in healthy children, adolescents and adults explained by the use of increasingly complex proportional reasoning rules. *Developmental Science*, 10(6), 814–825. <https://doi.org/10.1111/j.1467-7687.2007.00621.x>.
- Jacobs, J., & Potenza, M. (1991). The use of judgement heuristics to make social and object decisions: A developmental perspective. *Child Development*, 62(1), 166–178.
- Jaki, T., Kim, M., Lamont, A., George, M., Chang, C., Feaster, D., Van Horn, M. L. (in press). The effects of sample size on the estimation of regression mixture models. *Educational and Psychological Measurement*.
- 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. <https://doi.org/10.1006/drev.2001.0530>.
- 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(4), 383–416. <https://doi.org/10.1006/jecp.2002.2664>.
- Jansen, B. R. J., van Duijvenvoorde, A. C. K., & Huizenga, H. M. (2012). Development of decision making: Sequential versus integrative rules. *Journal of Experimental Child Psychology*, 111, 87–100. <https://doi.org/10.1016/j.jecp.2011.07.006>.
- Kwak, Y., Payne, J. W., Cohen, A. L., & Huettel, S. A. (2015). The rational adolescent: Strategic information processing during decision making revealed by eye tracking. *Cognitive Development*, 36, 20–30. <https://doi.org/10.1016/j.cogdev.2015.08.001>.
- Lang, A., & Betsch, T. (2018). Children's neglect of probabilities in decision making with and without feedback. *Frontiers in Psychology*, 9(FEB), 1–14. <https://doi.org/10.3389/fpsyg.2018.00191>.
- Lee, M. D. (2016). Bayesian outcome-based strategy classification. *Behavior Research Methods*, 48, 29–41. <https://doi.org/10.3758/s13428-014-0557-9>.
- Lee, M. D. (2018a). Bayesian methods for analyzing true-and-error models. *Judgment and Decision making*, 13, 622–635.
- Lee, M. D. (2018b). Bayesian methods in cognitive modeling. In J. Wixted & E.-J. Wagenmakers (Eds.) *The Stevens' handbook of experimental psychology and cognitive neuroscience*, volume 5: Methodology (Fourth Edition). John Wiley & Sons.
- Lee, M. D., & Sarnecka, B. W. (2011). Number knower-levels in young children: Insights from a Bayesian model. *Cognition*, 120, 391–402.
- Lee, M. D., & Wagenmakers, E.-J. (2005). Bayesian statistical inference in psychology: Comment on Trafimow (2003). *Psychological Review*, 112, 662–668. <https://doi.org/10.1037/0033-295X.112.3.662>.
- Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian modeling for cognitive science: A practical course*. MA: Cambridge University Press Cambridge.
- Lunn, D., Jackson, C., Best, N., Thomas, A., & Spiegelhalter, D. (2012). *The BUGS book: A practical introduction to Bayesian analysis*. Boca Raton, FL: Chapman & Hall/CRC Press.
- McCutcheon, A. L. (1987). *Latent class analysis*. Beverly Hills, CA: Sage.
- Mistry, P. K., Lee, M. D., & Newell, B. R. (2016). An empirical evaluation of models for how people learn cue search orders. In *Proceedings of the 38th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.
- Nylund, K. L., Asparouhov, T., & Muthén, B. (2007a). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*, 14, 535–569. <https://doi.org/10.1080/10705510701575396>.
- Nylund, K., Bellmore, A., Nishina, A., & Graham, S. (2007b). Subtypes, severity, and structural stability of peer victimization: What does latent class analysis say? *Child Development*, 78(6), 1706–1722.
- Palminteri, S., Kilford, E. J., Coricelli, G., & Blakemore, S.-J. (2016). The computational development of reinforcement learning during adolescence. *PLoS Computational Biology*, 12, 1–25. <https://doi.org/10.1371/journal.pcbi.1004953>.
- Perry, L. K., Samuelson, L. K., Malloy, L. M., & Schiffer, R. N. (2010). Learn locally, think globally: Variability supports high generalization and word learning. *Psychological Science*, 21(12), 1894–1902.
- Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing* (Vol. 124, p. 125).
- Poirel, N., Borst, G., Simon, G., Rossi, S., Cassotti, M., Pineau, A., & Houdé, O. (2012). Number conservation is related to children's prefrontal inhibitory control: An fMRI study of a piagetian task. *PLoS One*, 7(7), e40802. <https://doi.org/10.1371/journal.pone.0040802>.
- Potter, T., Bryce, N., & Hartley, C. (2017). Cognitive components underpinning the development of model-based learning. *Developmental Cognitive Neuroscience*, 25, 272–280. <https://doi.org/10.1016/j.dcn.2016.10.005>.
- R Development Core Team. (2008). *R: A language and environment for statistical computing* [computer software manual]. Vienna, Austria. Retrieved January 18, 2019, from <http://www.R-project.org>.
- Rabi, R., Miles, S. J., & Minda, J. P. (2015). Learning categories via rules and similarity: Comparing adults and children. *Journal of Experimental Child Psychology*, 131, 149–169. <https://doi.org/10.1016/j.jecp.2014.10.007>.
- Raijmakers, M. E. J., Jansen, B. R. J., & van der Maas, H. L. J. (2004). Rules and development in triad classification task performance. *Developmental Review*, 24(3), 289–321. <https://doi.org/10.1016/j.dr.2004.06.002>.
- Regenwetter, M., Dana, J., & Davis-Stober, C. P. (2011). Transitivity of preferences. *Psychological Review*, 118, 42–56. <https://doi.org/10.1037/a0021150>.
- Reyna, V. (2008). Development and decision making in adolescence. *Psychological Science*, 7, 1–44.
- Reyna, V. F., & Ellis, S. C. (1994). Fuzzy-trace theory and framing effects in children's risky decision-making. *Psychological Science*, 5, 275–279.
- Rieskamp, J. (2008). The probabilistic nature of preferential choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 1446–1465. <https://doi.org/10.1037/a0013646>.
- Rieskamp, J., & Otto, P. E. (2006). SSL: A theory of how people learn to select strategies. *Journal of Experimental Psychology: General*, 135, 207–236.
- Rindskopf, D. (1983). A general framework for using latent class analysis to test hierarchical and nonhierarchical learning models. *Psychometrika*, 48, 85–97. <https://doi.org/10.1007/BF02314678>.
- Rindskopf, D. (1987). Using latent class analysis to test developmental models. *Developmental Review*, 7, 66–85. [https://doi.org/10.1016/0273-2297\(87\)90005-0](https://doi.org/10.1016/0273-2297(87)90005-0).
- Rouder, J. N., & Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin & Review*, 12, 573–604.
- Schlottmann, A. (2000). Children's judgements of gambles: A disordinal violation of utility. *Journal of Behavioral Decision Making*, 13(1), 77–89. [https://doi.org/10.1002/\(SICI\)1099-0771\(200001/03\)13:1<77::AID-BDM344>3.0.CO;2-Y](https://doi.org/10.1002/(SICI)1099-0771(200001/03)13:1<77::AID-BDM344>3.0.CO;2-Y).
- Schlottmann, A. (2001). Children's probability intuitions: Understanding the expected value of complex gambles. *Child Development*, 72(1), 103–122.

- Schlottmann, A., & Anderson, N. H. (1994). Children's judgments of expected value. *Developmental Psychology*, 30(1), 56–66. <https://doi.org/10.1037//0012-1649.30.1.56>.
- Schmittmann, V. D., Visser, I., & Raijmakers, M. E. J. (2006). Multiple learning modes in the development of performance on a rule-based category-learning task. *Neuropsychologia*, 44(11), 2079–2091. <https://doi.org/10.1016/j.neuropsychologia.2005.12.011>.
- Schmittmann, V. D., van der Maas, H. L. J., & Raijmakers, M. E. J. (2012). Distinct discrimination learning strategies and their relation with spatial memory and attentional control in 4- to 14-year-olds. *Journal of Experimental Child Psychology*, 111(4), 644–662. <https://doi.org/10.1016/j.jecp.2011.10.010>.
- Schwartz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461–464.
- Shiffrin, R. M., Lee, M. D., Kim, W., & Wagenmakers, E. J. (2008). A survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods. *Cognitive Science*, 32(8), 1248–1284. <https://doi.org/10.1080/03640210802414826>.
- Siegler, R. S. (1976). Three aspects of cognitive development. *Cognitive Psychology*, 8, 481–520. [https://doi.org/10.1016/0010-0285\(76\)90016-5](https://doi.org/10.1016/0010-0285(76)90016-5).
- Siegler, R. S. (1987). The perils of averaging data over strategies: An example from Children's addition. *Journal of Experimental Psychology General*, 116(3), 250–264.
- Siegler, R. S. (1988). Individual differences in strategy choices: Good students, not-so-good students, and perfectionists. *Child Development*, 59(4), 833–851.
- Siegler, R. S. (2007). Cognitive variability. *Developmental Science*, 10(1), 104–109. <https://doi.org/10.1111/j.1467-7687.2007.00571.x>.
- Siegler, R. S. & Chen, Z. (2002). Development of rules and strategies: Balancing the old and the new. *Journal of Experimental Child Psychology*, 81(4), 446–457.
- Siegler, R. S., Strauss, S., & Levin, I. (1981). Developmental sequences within and between concepts. *Monographs of the Society for Research in Child Development*, 46(2), 1–84. <https://doi.org/10.2307/1165995>.
- Stevenson, C. E., Heiser, W. J., & Resing, W. C. M. (2016). Dynamic testing of analogical reasoning in 5- to 6-year-olds: Multiple-choice versus constructed-response training items. *Journal of Psychoeducational Assessment*, 34(6), 550–565. <https://doi.org/10.1177/0734282915622912>.
- Thomas, H., & Horton, J. J. (1997). Competency criteria and the class inclusion task: Modeling judgments and justifications. *Developmental Psychology*, 33(6), 1060–1073. <https://doi.org/10.1037/0012-1649.33.6.1060>.
- Torbeyns, J., De Smedt, B., Ghesquière, P., & Verschaffel, L. (2009). Acquisition and use of shortcut strategies by traditionally schooled children. *Educational Studies in Mathematics*, 71(1), 1–17. <https://doi.org/10.1007/s10649-008-9155-z>.
- Vandekerckhove, J., Rouder, J. N., & Kruschke, J. (2018). Editorial: Bayesian methods for advancing psychological science. *Psychonomic Bulletin & Review*, 25, 1–4.
- Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, 14, 779–804.
- Wagenmakers, E.-J., Morey, R. D., & Lee, M. D. (2016). Bayesian benefits for the pragmatic researcher. *Current Directions in Psychological Science*, 25, 169–176.
- Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E.-J. (2011). Statistical evidence in experimental psychology: an empirical comparison using 855 t tests. *Perspectives on Psychological Science*, 6, 291–298.
- Wilkening, F. (1981). Integrating velocity, time, and distance information: A developmental study. *Cognitive Psychology*, 13(2), 231–247. [https://doi.org/10.1016/0010-0285\(81\)90009-8](https://doi.org/10.1016/0010-0285(81)90009-8).
- Wilkening, F. (1988). A misrepresentation of knowledge representation. *Developmental Review*, 8, 361–367.