



UvA-DARE (Digital Academic Repository)

Quantitative or qualitative development in decision making?

Huizenga, H.M.; Zadelaar, J.; Jansen, B.R.J.

DOI

[10.1016/j.jecp.2021.105198](https://doi.org/10.1016/j.jecp.2021.105198)

Publication date

2021

Document Version

Final published version

Published in

Journal of Experimental Child Psychology

License

CC BY

[Link to publication](#)

Citation for published version (APA):

Huizenga, H. M., Zadelaar, J., & Jansen, B. R. J. (2021). Quantitative or qualitative development in decision making? *Journal of Experimental Child Psychology*, 210, [105198]. <https://doi.org/10.1016/j.jecp.2021.105198>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.



ELSEVIER

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp



Quantitative or qualitative development in decision making?



Hilde M. Huizenga^{a,b,c,1,*}, Jacqueline Zadelaar^{a,1}, Brenda R.J. Jansen^{a,b,c}

^a Department of Developmental Psychology, University of Amsterdam, 1001 NK Amsterdam, the Netherlands

^b Amsterdam Brain and Cognition Center, University of Amsterdam, 1001 NK Amsterdam, the Netherlands

^c Research Priority Area Yield, University of Amsterdam, 1018 WS Amsterdam, the Netherlands

ARTICLE INFO

Article history:

Received 28 August 2019

Revised 15 March 2021

Available online 4 June 2021

Keywords:

Quantitative vs qualitative development

Judgement and choice

Strategies and heuristics

MIMIC analysis

Information integration theory

Cumulative prospect theory

Risk return theory

Adaptive strategy selection

ABSTRACT

A key question in the developmental sciences is whether developmental differences are quantitative or qualitative. For example, does age increase the speed in processing a task (quantitative differences) or does age affect the way a task is processed (qualitative differences)? Until now, findings in the domain of decision making have been based on the *assumption* that developmental differences are either quantitative or qualitative. In the current study, we took a different approach in which we *tested* whether development is best described as being quantitative or qualitative. We administered a judgment version and a choice version of a decision-making task to a developmental sample ($n_{\text{judgment}} = 109$ and $n_{\text{choice}} = 137$; $M_{\text{age}} = 12.5$ years, age range = 9–18). The task, the so-called Gambling Machine Task, required decisions between two options characterized by constant gains and probabilistic losses; these characteristics were known beforehand and thus did not need to be learned from experience. Data were analyzed by comparing the fit of quantitative and qualitative latent variable models, so-called multiple indicator multiple cause (MIMIC) models. Results indicated that individual differences in both judgment and choice tasks were quantitative and pertained to individual differences in “consideration of gains,” that is, to what extent decisions were guided by gains. These differences were affected by age in the judgment version, but not in the choice version, of the task. We discuss implications for theories of decision making and discuss potential limitations and extensions. We also argue that

* Corresponding author at: Department of Developmental Psychology, University of Amsterdam, 1001 NK Amsterdam, the Netherlands.

E-mail address: h.m.huizenga@uva.nl (H.M. Huizenga).

¹ These authors share first authorship.

the MIMIC approach is useful in other domains, for example, to test quantitative versus qualitative development of categorization, reasoning, math, and memory.

© 2021 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Introduction

A key question in the developmental sciences is whether development is best described as quantitative or qualitative (Jones & Dekker, 2018; Siegler, 2007). For example, the developmental increase in working memory may best be attributed to a quantitative increase in inhibition of distracting information (Plebanek & Sloutsky, 2019). On the other hand, the developmental increase in the speed with which children solve simple additions may originate in qualitative changes in strategy use; young children use a slow counting strategy, whereas older children use a fast memory retrieval strategy (Ashcraft & Fierman, 1982). In the domain of decision making, it is still largely unknown whether development in decision making (Defoe, Dubas, Figner, & van Aken, 2012; Rosenbaum, Venkatraman, Steinberg, & Chein, 2018) is quantitative or qualitative. Quantitative development would signify that all people use the same strategy to arrive at a decision but that there would be quantitative developmental changes in the parameters describing such a strategy. Qualitative development, on the other hand, would imply that there would be developmental changes in strategies being used. The purpose of the current study therefore was to test whether development of decision making is best characterized as quantitative or qualitative. We did so in a sample of 9- to 18-year-olds using a latent variable modeling approach specifically tailored to differentiate between these two types of development.

In the current study, decisions were made given described (i.e., known) gain amounts, loss amounts, and their associated probabilities, so-called attributes (Rosenbaum et al., 2018). Some theories are best suited to describe quantitative development in such decision making. First, according to information integration theory (Anderson, 1980; Schlottmann & Tring, 2005; Wilkening & Anderson, 1982), all people base decisions on a weighted sum of attributes. Age-related changes may then be due to quantitative developmental changes in the weights given to these attributes, for example, in the weights given to gains (in a different domain: Liu, Gonzalez, & Warneken, 2019; Wilkening, 1981). In addition there may also be quantitative developmental changes in how strictly people base their decisions on this strategy, which is captured in the so-called sensitivity parameter (Glöckner & Pachur, 2012; Nussenbaum & Hartley, 2019). Second, according to cumulative prospect theory, decisions are based on subjective utility, that is, a sum of subjectively evaluated gain and loss amounts multiplied by their subjectively evaluated probabilities (Tversky & Kahneman, 1992). Age-related changes may then be due to quantitative changes in the parameters governing subjective evaluation (e.g., of gains) or in the sensitivity parameter (Harbaugh, Krause, & Vesterlund, 2002; Steelandt, Broihanne, Romain, Thierry, & Dufour, 2013). Third, according to risk return theory (Weber, Shafir, & Blais, 2004), decisions are made given a weighted sum of expected value and risk. Quantitative developmental differences may then exist in the weighting of risk or in the sensitivity parameter (Paulsen, Platt, Huettel, & Brannon, 2011; van Duijvenvoorde et al., 2015; Wolf, Wright, Kilford, Dolan, & Blakemore, 2013).

However, other theories are better suited to describe qualitative developmental changes; that is, there may be age-related changes in strategies being used to arrive at a decision. First, according to dual process theory, people use either an intuitive or a deliberate decision-making strategy (Evans, 2008, 2011). Development may then originate in a shift from the intuitive strategy to the deliberate strategy (Kokis, Macpherson, Toplak, West, & Stanovich, 2002; Stanovich, West, & Toplak, 2011). Second, fuzzy trace theory suggests a reverse developmental shift—from a deliberate strategy to an intuitive strategy (Reyna & Ellis, 1994). Third, according to lexicographic theory (Tversky, Sattath, & Slovic, 1988), which resembles the theory on proportional reasoning (Jansen & van der Maas, 2002; Siegler &

Chen, 2002), development may originate in a shift from simple to complex strategy use (Betsch, Lehmann, Jekel, Lindow, & Glöckner, 2018; Huizenga, Crone, & Jansen, 2007; Lang & Betsch, 2018; but see Jansen, Van Duijvenvoorde, & Huizenga, 2012; Lindow, Lang, & Betsch, 2017). Fourth, it has been argued that there is a developmental shift in adaptive strategy selection (Betsch & Lang, 2013; Betsch, Lehmann, Lindow, Lang, & Schoemann, 2016; Lindow & Betsch, 2018, 2019), for example, that children, but not adults, use complex strategies even if simple strategies suffice (Mata, von Helversen, & Rieskamp, 2011).

In the literature reviewed above, developmental differences were explained by *assuming* either quantitative or qualitative developmental differences. That is, in the quantitative literature, it was assumed that everyone uses the same strategy but that there may exist quantitative differences in the parameters governing this strategy. For example, in risk return theory studies, it is assumed that every decision maker bases a decision on a weighted sum of expected value and risk; developmental differences are then assumed to originate in the weighting of risk (e.g., Paulsen et al., 2011; van Duijvenvoorde et al., 2015; Wolf et al., 2013). But it was not tested whether these developmental differences could be better described as qualitative. For example, it was not tested whether some participants based their decisions on a strategy in which they only considered gains whereas other did so given a strategy in which they considered both gains and losses. Vice versa, in the qualitative literature, differences in strategy use were reported, but it was never tested whether these differences could better be explained by quantitative differences. For example, it was reported that there were developmental differences in strategy use (Betsch & Lang, 2013; Betsch et al., 2016, 2018; Huizenga et al., 2007; Jansen et al., 2012; Lang & Betsch, 2018; Lindow & Betsch, 2018, 2019; Lindow et al., 2017; Mata et al., 2011; Reyna & Ellis, 1994), but it was not tested whether these could be better explained by quantitative differences in, for example, the weighting of gains.

In the current study, we adopted a new approach in which we directly tested which perspective, either quantitative or qualitative, better describes developmental differences in decision making. Such a testing of the nature of individual differences is important for theoretical and applied reasons (in a different domain, see also McGrath & Walters, 2012). Suppose that it is indeed found that developmental differences in decision making are quantitative in nature. This then provides evidence for quantitative theories such as information integration theory, cumulative prospect theory, and risk return theory. In that case, interventions to improve decision making are then also best rooted in these quantitative conceptualizations.

To this end, we administered a decision-making task in a large and developmentally diverse sample. Resulting data were analyzed given the following rationale. If individual differences in decision-making differences are quantitative, the latent variable driving these differences should be quantitative. However, if individual differences are qualitative, the latent variable should be qualitative. Therefore, we compared two multiple indicator multiple cause (MIMIC) models (Bollen & Bauldry, 2011; Joreskog, 1975; Kievit et al., 2012; Zadelaar et al., 2019) (see Fig. 1). In the quantitative model, age and potential other individual difference variables influence the score on a quantitative latent variable, which in turn determines responses to decision items. The quantitative latent variable thus can be conceived as a factor, just as in a factor analysis. In the qualitative model, age and potential other individual difference variables influence the likelihood of belonging to qualitatively different latent classes, each characterized by a different decision strategy that determines responses to decision items. The qualitative latent variable thus can be conceived as a categorical variable with multiple latent classes, just as in a mixture analysis. If the quantitative model provides the best fit, we conclude that individual differences are best described as quantitative. On the other hand, if the qualitative model proves to be better, we conclude that these differences are qualitative in nature.

Apart from age, we included the individual difference variables sex, neuroticism, and math performance. Effects of sex on decision making have been reported in every age range, with male participants taking more risks than female participants (Blais & Weber, 2001; Byrnes, Miller, & Schafer, 1999). Among personality variables, neuroticism has been shown to have the largest effect on decision making (Hilbig, 2008; Lauriola & Levin, 2001). Hilbig (2008) showed that high neuroticism was associated with simple over complex decision strategy use. Lauriola and Levin (2001) showed that high neuroticism was associated with decreased risk taking in gain domains but increased risk taking

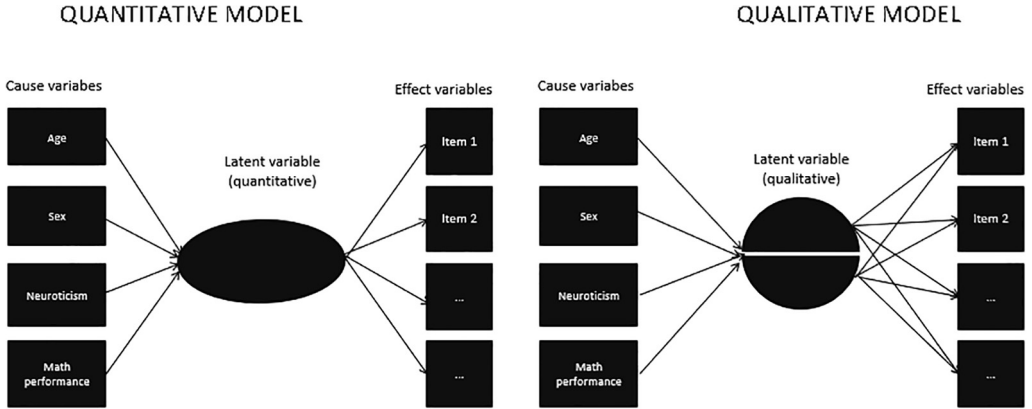


Fig. 1. Schematic illustration of multiple indicator multiple cause (MIMIC) models. In the quantitative model (left panel), individual difference variables (“cause variables”) influence the score on a quantitative latent variable, which in turn influences item responses (“effect variables”), which can be either judgment or choice responses. In the qualitative model (right panel), individual difference variables influence the logistic probability of belonging to different latent classes. Each latent class is characterized by distinct item responses, which again can be either judgment or choice responses. In this schematic illustration, the qualitative latent variable consists of two classes of equal size, but more classes and/or different class sizes are possible.

in loss domains. Finally, we included math performance because some complex strategies require integration of attributes, whereas other simple strategies do not.

The decision-making task was the so-called Gambling Machine Task, which was specifically designed to be sensitive to individual differences in strategy use (Bexkens, Jansen, van der Molen, & Huizenga, 2016; Dekkers et al., 2020; Jansen et al., 2012; Steingroever, Jepma, Lee, Jansen, & Huizenga, 2019; Van Duijvenvoorde et al., 2016). In this task, participants make a decision concerning two options, that is, two gambling machines. They may be asked either to choose the best of both options (i.e., choice), or to indicate to what extent they judge one option to be better than the other (i.e., judgment) (Figner & Schaub, n.d.; Montgomery, Selart, Gärling, & Lindberg, 1994; Payne, Bettman, Coupey, & Johnson, 1992; Selart, 1996; Tversky et al., 1988; Westenberg & Koele, 1992; Wilkening & Anderson, 1982). It has, to our knowledge, never been investigated whether developmental differences are robust over choice and judgment in decision-making tasks. However, in the developmental literature on the balance scale task, it has been suggested that there is a developmental increase in complexity of strategy use in choice versions, indicative of qualitative differences, whereas such differences are absent in judgment versions (Wilkening & Anderson, 1982). Therefore, to explore potential differences between choice and judgment, we administered both a choice version and a judgment version of the task. For both the choice and judgment data, we compared the fit of quantitative and qualitative MIMIC models.

Method

Participants

Participants were recruited through schools in the Netherlands. Primary caretakers were informed about the experiment and had the opportunity to exempt their children from participating. A total of 260 children ($M_{age} = 12.5$ years, age range = 9–18; 46% male)² filled in the paper-and-pencil Gambling Machine Task (GMT). All procedures were approved by the local ethics committee of the University of Amsterdam.

² Some children did not report their age ($n = 5$) or their sex ($n = 3$). Thus, M_{age} is computed over 255 children, and percentage male is computed over 257 children, instead of 260 children.

Experimental design and materials

Participants needed to respond either by choice or by judgment; response type was varied between participants. Age, math performance, and neuroticism were continuous between-participants variables. Sex, response (judgment vs. choice), and version (1 vs. 2; in Version 2, items were administered in reverse order) were discrete between-participants variables. Participants were randomly allocated to version and response conditions.

Decision making was assessed with the GMT. The task consisted of 47 items; an example is given in Fig. 2. Each item depicted two gambling machines characterized by their frequency of loss (FL), amount of loss (AL) and certain gain (CG) (attributes of all items are given in Table 1 of the supplementary online material [SOM Table 1]). Participants were asked to imagine that balls would be tossed and tumbled and that one ball would be drawn. They were instructed that if a white ball would be drawn, they would gain the amount printed on the machine (e.g., with the left-hand machine in Fig. 2, they would gain 2) and would lose nothing. If, however, a gray ball would be drawn, they would gain the amount printed on the machine but would also lose the amount printed on the gray ball (e.g., with the left-hand machine in Fig. 2, they would lose 10). Participants were then asked to decide. In the choice GMT, participants needed to decide by ticking one of three boxes: “Gambling Machine A,” “Equal,” or “Gambling Machine B.” In the judgment GMT, they needed to mark a 7-point Likert scale, with 1 indicating *I choose only Gambling Machine A*, 4 indicating *I choose Gambling Machine A and Gambling Machine B equally*, and 7 indicating *I choose only Gambling Machine B*.

Items depicted two machines that did or did not differ in expected value: $EV = AL * (FL/10) + CG$. For example, in Fig. 2, the two machines differed, with $EV = -10 * (1/10) + 2 = 1$ in the left-hand panel and $EV = -2 * (5/10) + 4 = 3$ in the right-hand panel. The low- and high-EV machines were coined the worst and best machines, respectively; thus, in Fig. 2, the right-hand machine was the best machine. Expected values of the two machines in each item are given in SOM Table 1.

The total of 47 items could be grouped into 16 item types based on which attribute(s) differed between machines and which attribute(s) signaled the best machine. The GMT item types were inspired by Siegler (1976) rule assessment methodology. That is, given a set of prespecified strategies (“rules”), we constructed item types that allowed for differentiation between these strategies. So, differences between items were deliberately chosen to differentiate between strategies. Therefore, this necessitated an identification system for item types; simple numbering (1–16) would not suffice. The identification system, also used in previous studies (Bexkens et al., 2016; Dekkers et al., 2020; Jansen et al., 2012; Steingroever et al., 2019; Van Duijvenvoorde et al., 2016), is described below. In SOM Table 2, we outline how this set of item types may differentiate between decision strategies.

The E (equal) type was an item in which no attributes differed between machines. In item types where attribute(s) differed between machines, attributes signaling the best machine (i.e., the machine with the highest expected value) were printed in uppercase, attributes signaling the worst machine were printed in lowercase, and attributes that did not differ were not printed. For example, an item type labeled CG differed only in certain gain, and the machine with the highest certain gain was best. As another example, the item in Fig. 2 belonged to the fl_AL,CG type; the machines differed on all attributes, and the best option was the machine that was optimal on amount of loss and on certain gain (as indicated by the uppercase AL and CG), whereas the worst option was the machine that was optimal on frequency of loss (as indicated by the lowercase fl). Finally, in the fl_al_cg type, all attributes differed but machines had equal expected values; therefore, none of the attributes was printed in uppercase. An underscore indicates that attributes conflicted (i.e., signaled different machines). For example, fl_AL,CG tells us that the frequency of loss signaled one machine, whereas both amount of loss and certain gain signaled the other machine.

Worst and best options were semirandomly presented on the left-hand or right-hand side of the item; therefore, responses were coded in the following way. In the choice task, each response was coded as a preference for the worst option (1), as indifference (2), or as a preference for the best option (3). In the judgment task, each response was coded such that a low score (1) indicated a preference for the worst option and a high score (7) indicated a preference for the best option. Responses to items in which options did not differ in expected value (i.e., the fl_al_cg and fl_al item types) were not recorded, meaning that a high choice or judgment score indicated a preference for the right-hand option. In the

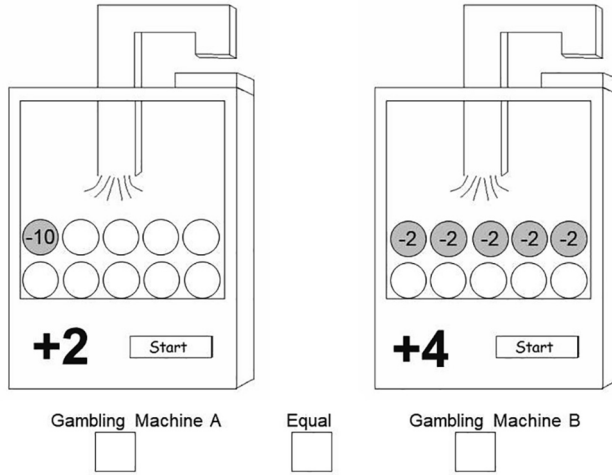


Fig. 2. Example of an item in the choice task. The left-hand and right-hand machines were respectively characterized by a frequency of loss (FL) of 1 and 5, an amount of loss (AL) of -10 and -2 , and a certain gain (CG) of $+2$ and $+4$.

fl_al_cg item type, the right-hand option was optimal on constant gain; in the fl_al item type, the right-hand option was optimal on amount of loss.

The GMT offers the opportunity to assess decision strategies, with different decision strategies giving rise to different response patterns on the task (Bexkens et al., 2016; Dekkers et al., 2020; Jansen et al., 2012; Steingroever et al., 2019; Van Duijvenvoorde et al., 2016; Zadelaar et al., 2019). For example, if people use a strategy in which they compare options on expected value, they will answer all items correctly. However, if they use a simple strategy in which they compare options only on their frequency of loss, they will answer only items with the uppercase FL correctly and will answer other items incorrectly. For example they will answer the fl_AL,CG item in Fig. 2 incorrectly because they will state that the machine with the lowest frequency of loss is optimal. In the supplementary online material, we have outlined how the GMT can differentiate among several strategies. The current version of the task was an adaptation of the GMT that was originally developed to assess qualitative individual differences, that is, differences in strategy use. To also be sensitive to quantitative differences, we adapted this task in two ways. First, we incorporated additional items generating more variations in gains, losses, and their probabilities. Increasing variation in attributes has been suggested to be important in detecting quantitative individual differences (Figner, Mackinlay, Wilkening, & Weber, 2009; Figner & Voelki, 2004; Wilkening & Anderson, 1982). Second, because it has been suggested that quantitative individual differences are more pronounced in judgment tasks than in choice tasks (Wilkening & Anderson, 1982), we included a judgment version of this task as well.

The GMT has been administered previously in developmental samples, for example, as young as 8 years (Jansen et al., 2012). In the latter study, consistency of responses over items was high, suggesting that children did understand the task and were not guessing.

Neuroticism was assessed with the neuroticism subscale of the Revised Junior Eysenck Personality Questionnaire (JEPQ-R) (Scholte & De Bruyn, 2001). The neuroticism subscale consisted of 20 statements on which participants could answer either *yes* or *no*. One item (“Do you sometimes feel life is just not worth living?”) was not administered because it was considered inappropriate for the current sample; therefore, the sum score was computed over 19 items. We were interested in effects of neuroticism over and above age and sex. Therefore, we regressed the neuroticism sum score on age and sex in the entire sample and saved the standardized residuals for further analysis.³

³ There was a significant effect of sex; female participants had higher neuroticism scores than male participants (cf. Scholte & De Bruyn, 2001). There were no effects of age.

Math performance was assessed with the Tempo Test Rekenen [Number Fact Retrieval Test] (De Vos, 1992). This test consisted of five subscales: addition, subtraction, multiplication, division, and a mix of these four problem types. Each subscale consisted of 40 items ranging from easy to difficult. Participants were allowed 1 min for each subscale. We used the mixed subscale for further analysis given that deciding may require all four mathematical capacities. Because we were interested in effects of math performance over and above age and sex, the number of correct items on the mixed subscale was regressed on age and sex⁴ in the entire sample; standardized residuals were saved for further analysis.

Procedure

Assessment was administered groupwise in a classroom and was led by two research assistants. Children were first handed the GMT booklet. Assistants explained the task and completed two example items together with the children. Instruction and examples were printed in the booklet as well. Participants were given 20 min to complete the GMT. After completion, GMT booklets were replaced by the neuroticism subscale and assistants instructed participants to choose the answer that suited them best and to not spend too much time on selecting an answer; participants were given 5 min to fill out this scale. After retrieving the completed questionnaires, assistants explained the math task. Because it was important that all participants started simultaneously, the sheet with the math problems was handed out upside down after the explanation. After a starting signal, participants flipped the sheet and started the first part of the task (addition). After 1 min, they started the second part, and so on. Assistants retrieved the sheets and handed out treats to thank the participants.

Analysis

A total of 16 item types were administered; however 4 item types were not included in the MIMIC analysis because these showed hardly any individual differences.⁵ That is, responses to the E type were consistently characterized by indifference, and responses to three other types (FL, AL, and CG) were characterized by a near unanimous preference for the best option (cf. SOM Table 1, columns with means and medians). Thus, this left $16 - 4 = 12$ item types for the MIMIC analyses.

We first fitted quantitative and qualitative MIMIC models to the judgment data and compared the fit of these models (Zadelaar et al., 2019), and we then did the same for the choice data. As outlined in Zadelaar et al. (2019), in a MIMIC model (cf. Fig. 1) a latent variable is influenced by a set of observable “cause variables” (age, sex, neuroticism, and math performance) while influencing another set of observable “effect variables” (continuous item judgments in the judgment task or ordinal item choices in the choice task). The latent variable, either quantitative or qualitative, represents an underlying construct of interest (Kievit et al., 2011, 2012). Quantitative individual differences (i.e., dimensional, spread across a continuum) in the construct of interest should be reflected by a quantitative latent variable, as represented in what is henceforth referred to as the *quantitative model* (see Fig. 1, left-hand panel). Similarly, qualitative individual differences (i.e., categorical, consisting of multiple homogeneous classes, each representing a strategy) in the construct of interest are reflected by a qualitative latent variable (see the *qualitative model* in Fig. 1, right-hand panel).

Thus, the quantitative model can be regarded as an extended factor analysis⁶: a *factor analysis* because the latent variable underlying item responses is quantitative (a factor) and an *extended factor analysis* because this factor is influenced by cause variables (e.g., age). As in a factor analysis, interpretation of the quantitative latent variable is based on the relationship between factor and item responses. To illustrate with an analogy, in the interpretation of factors underlying a personality questionnaire, a factor is interpreted as “conscientiousness” if this is the common aspect of items loading positively on

⁴ There was a significant effect of age, with math performance increasing with age. There were no effects of sex.

⁵ These omitted four simple item types comprised 13 items; hence, $47 - 13 = 34$ items were included in the analysis.

⁶ The term factor analysis is used when the latent factor underlies continuous, ordinal, or nominal variables.

this factor. The qualitative model can be considered as an extended mixture analysis⁷: a *mixture analysis* because the latent variable underlying item responses is qualitative (multiple latent classes) and an *extended mixture analysis* because the likelihood of belonging to each of these classes is influenced by cause variables (e.g., age). As in a mixture analysis, interpretation of the latent classes is based on item responses for each class. To illustrate again with an analogy, in the interpretation of latent classes underlying an internalizing problems questionnaire, a latent class is interpreted as “depression” if this is the common aspect of items endorsed by this class, whereas another class is interpreted as “anxiety” if this is the common aspect of items endorsed by this other class. In the analyses, the quantitative latent variable model provided the best fit for both the judgment and choice data. Therefore, we first describe the quantitative model and then proceed with its qualitative counterpart.

The quantitative model for the judgment data was defined by the following parameters. The cause parameters defined how age, sex, neuroticism, and math performance influenced the latent variable; a positive cause parameter indicated that an increase in a cause variable yielded an increase in the latent variable. The effect parameters consisted of slopes and intercepts for each item; a positive effect slope indicated that an increase in the latent variable yielded an increase in an item judgment. These effect slopes index the relationship between the latent variable and item responses and thus are crucial for interpretation of the latent variable. The effect intercepts denoted item judgments if the latent variable is zero. To obtain stability of the model, effect parameters of items belonging to the same item type were constrained to be equal. To identify the model (Bollen & Bauldry, 2011), the effect slope of one item type, fl_{al_cg} , was fixed to 1.⁸

In the quantitative model for the choice data, the cause parameters again defined how age, sex, neuroticism, and math performance influenced the latent variable; interpretation is equivalent to that for the judgment data. The effect parameters consisted of slopes and intercepts for each item. Because the choice data were ordinal (1, 2, 3), positive effect slopes indicated that an increase in the latent variable yielded an increase in the chances of a higher choice. Again, these effect slopes are crucial for interpretation of the latent variable. Effect intercepts denoted preferences if the latent variable was zero. There were two intercepts: one for the 1 choice as opposed to the 3 choice and one for the 2 choice as opposed to the 3 choice. To obtain stability of the model, effect parameters of items belonging to the same item type were constrained to be equal. To identify the model (Bollen & Bauldry, 2011), the effect slope of one item type, fl_{al_cg} , was fixed to 1.⁹

The qualitative model for the judgment data is defined by the following parameters. The cause parameters consist of an intercept and slopes. The cause slopes define how age, sex, neuroticism, and math performance influence the likelihood of belonging to a class as compared with a reference class. The cause intercept defines the likelihood of belonging to a certain class rather than the reference class when all cause indicators equal zero. Each class has its own effect parameters that define the mean judgment on an item for that particular class. Thus, these effect parameters are crucial for interpretation of these classes. To obtain stability of the model, effect parameters of items belonging to the same item type were constrained to be equal.

In the qualitative model for the choice data, the cause parameters are equivalent to those of the judgment data, as outlined above. With respect to the effect parameters, each class has two parameters per item that define the likelihood of individuals of that class giving a response (1 or 2) as opposed to the reference response (3). Again, these effect parameters are crucial for interpretation of classes. To obtain stability of the model, effect parameters of items belonging to the same item type were constrained to be equal.

For both the judgment and choice data, we adopted the following procedure. We fitted qualitative models with two to four latent strategy classes and a quantitative model with one latent variable. We

⁷ The term mixture analysis is used when the latent classes underlie continuous, ordinal, or nominal variables. Mixture analysis for continuous variables is also known as latent profile analysis; mixture analysis for ordinal or nominal data is also known as latent class analysis.

⁸ The model also contains a latent variable error variance because the latent variable is predicted by cause variables, and item error variances as item judgments are predicted by the latent variable (Bollen & Bauldry, 2011).

⁹ The model also contains a latent variable error variance because the latent variable is predicted by cause variables. Item variances are not estimated for ordinal data.

then determined, for the best fitting qualitative model and for the best fitting quantitative model, whether effects of age, sex, neuroticism, and math performance could be set to zero. We did so by comparing models by means of the Bayesian information criterion (BIC), where lower values indicate better model fit. Finally, we compared the best fitting qualitative model and the best fitting quantitative model by means of the BIC, which was warranted because this fit measure is also suited for non-nested models (Posada & Buckley, 2004; Vrieze, 2012). Analyses were performed in Mplus; code can be found in the supplementary online material.

Results

Judgment data

The MIMIC model required all cause variables (i.e., age, sex, neuroticism, and math performance) to be complete. Because 7 of the 116 participants had missing values on at least one cause variable, this left 109 participants for further analysis. Among the qualitative models, the two-class model converged, whereas the three- and four-class models did not do so.¹⁰ The quantitative model fitted best if the effect of age was included. The same was true for the qualitative two-class model. Fit was better for the quantitative model with age included (BIC = 13813.1) than for the qualitative model with age included (BIC = 13950.8) (see SOM Table 3 for all model selection results; see SOM Table 4 for additional relative fit measures, i.e., the Akaike information criterion [AIC] and AIC weights; and see Fig. 3 for absolute fit). Note that in both the quantitative and qualitative models, sex, neuroticism, and math performance were not required to describe the data adequately.

To provide more insight into the interpretation of the quantitative latent variable, we inspected the effect slope parameters¹¹ because they index how the latent variable is associated with the item responses. In the upper panel of Fig. 4, it can be seen that the latent variable was best described as a contrast between item types in which constant gain uniquely signaled the correct option (significant positive parameters) and item types in which constant gain did not do so (nonsignificant parameters or significant negative parameters).¹² Therefore, we interpreted the latent variable as “consideration of gains.” That is, this variable indexed to what extent decisions were influenced by gains. Note that this interpretation was consistent with the fixed positive parameter of the fl_al_cg item type.¹³ Cause slope estimates provided more insight into the effects of cause variables on the latent variable consideration of gains. The effect of age was significant ($b = -0.269, p = .034$). The negative age effect indicated that with increasing age, scores on the continuous latent variable consideration of gains decreased. Taken together, these results indicate that quantitative individual differences in this judgment task originated in a developmental decrease in consideration of gains.

Choice data

Of the 144 participants, 7 had missing values on at least one cause variable, leaving 137 participants for further analysis. The quantitative model converged, and no cause variables needed to be included. Among the qualitative models, the two- and three-class models converged, whereas the

¹⁰ The three- and four-class models did not converge because the number of parameters was too high. As a check, we also performed adapted analyses requiring fewer parameters. This indeed resulted in convergence for the two- and three-class models but not for the four-class model. The best fitting adapted qualitative model fitted the data less well than the best fitting quantitative model. Refer to SOM Table 6 for complete results.

¹¹ Although the quantitative model provided the best fit for judgment and choice data, in SOM Table 5 we also report results of the best fitting qualitative models.

¹² There is one exception to this pattern; the item type fl_al, in which both options had equal expected values, also was characterized by a significant positive effect parameter.

¹³ That is, people scoring high on the latent variable “consideration of gains” gave a high judgment score on this fl_al_cg item type. Because options did not differ in expected value on this item type, a high judgment score on this item incidentally indicated a preference for the option characterized by a high constant gain (cf. “Experimental design and materials” section). Hence, they considered gains in their judgment.

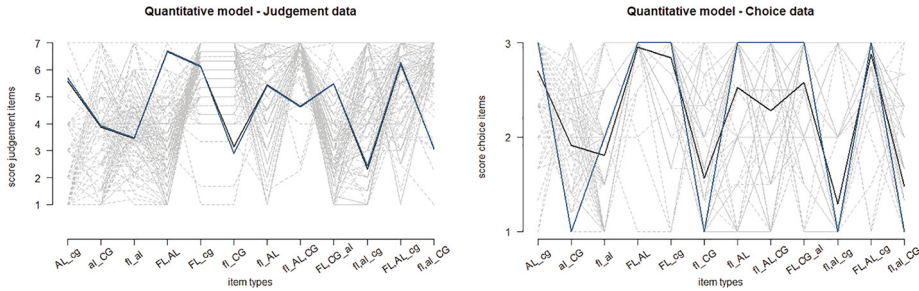


Fig. 3. Observed (black) and predicted (blue) mean response patterns across participants per item type (for an explanation of item types, see “Experimental design and materials” section) for the judgment data (left panel) and choice data (right panel) for the best fitting (i.e., quantitative) multiple indicator multiple cause (MIMIC) model. For the judgment data, this best fitting model included age; for the choice data, no cause variables were included. The x axis denotes item type; the y axis denotes the response on the corresponding item type. Dotted gray lines represent response patterns of individual participants. It can be seen that the quantitative model with age included fitted the mean response patterns well; however, absolute model fit for this model was mediocre to poor, $\chi^2(706) = 1698.244, p < .001$, root mean square error of approximation (RMSEA) = .114, comparative fit index (CFI) = .282, standardized root mean square residual (SRMS) = .149. Similar absolute fit statistics were not available for the choice data. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

four-class model did not do so.¹⁴ The qualitative three-class model outperformed the two-class model and fitted best if no cause variables were included. The quantitative model (BIC = 5747.6) outperformed the qualitative three-class model (BIC = 5872.2) (see SOM Table 3 for all model selection results; see SOM Table 4 for additional relative fit measures, i.e., the AIC and AIC weights; and see Fig. 3 for absolute fit). As can be seen in the lower panel of Fig. 4, item slopes again indicated that the quantitative latent variable was best described as a contrast between item types in which constant gain uniquely signaled the correct option and item types in which it did not do so. So, the latent variable could again be interpreted as consideration of gains. The conclusion therefore is that individual differences in this choice task were quantitative, could be interpreted as consideration of gains, and were not significantly related to age, sex, neuroticism, or math performance.

Discussion

Studies on the development of decision making assume that differences are either quantitative or qualitative. The current new MIMIC model comparison approach offers the possibility to test which of these two conceptualizations is most adequate. Current results, obtained by administering the GMT to a sample of 9- to 18-year-olds, indicate that individual differences are best described as quantitative individual differences in consideration of gains in making a judgment or a choice. In a judgment version of the task, these differences are affected by age; there is an age-related decrease in consideration of gains, whereas this decrease is not observed in a choice version of the task. Other individual difference variables (i.e., sex, neuroticism, and math performance) had no effects on considerations of gains.

These results offer five new insights. The first insight pertains to the main question of the current study: Are developmental differences in decision making quantitative or qualitative in nature? Current results provide support for quantitative conceptualizations, as implemented in information integration theory, cumulative prospect theory, and risk return theory (Anderson, 1980; Schlottmann & Tring, 2005; Tversky & Kahneman, 1992; Weber et al., 2004; Wilkening & Anderson, 1982).

¹⁴ As a check, for the choice data we also performed adapted analyses requiring fewer parameters for the two-, three-, and four-class models. This resulted in convergence for the two- and three-class models but not for the four-class model. The best fitting adapted qualitative model fitted the data less well than the best fitting quantitative model. Refer to SOM Table 6 for complete results.

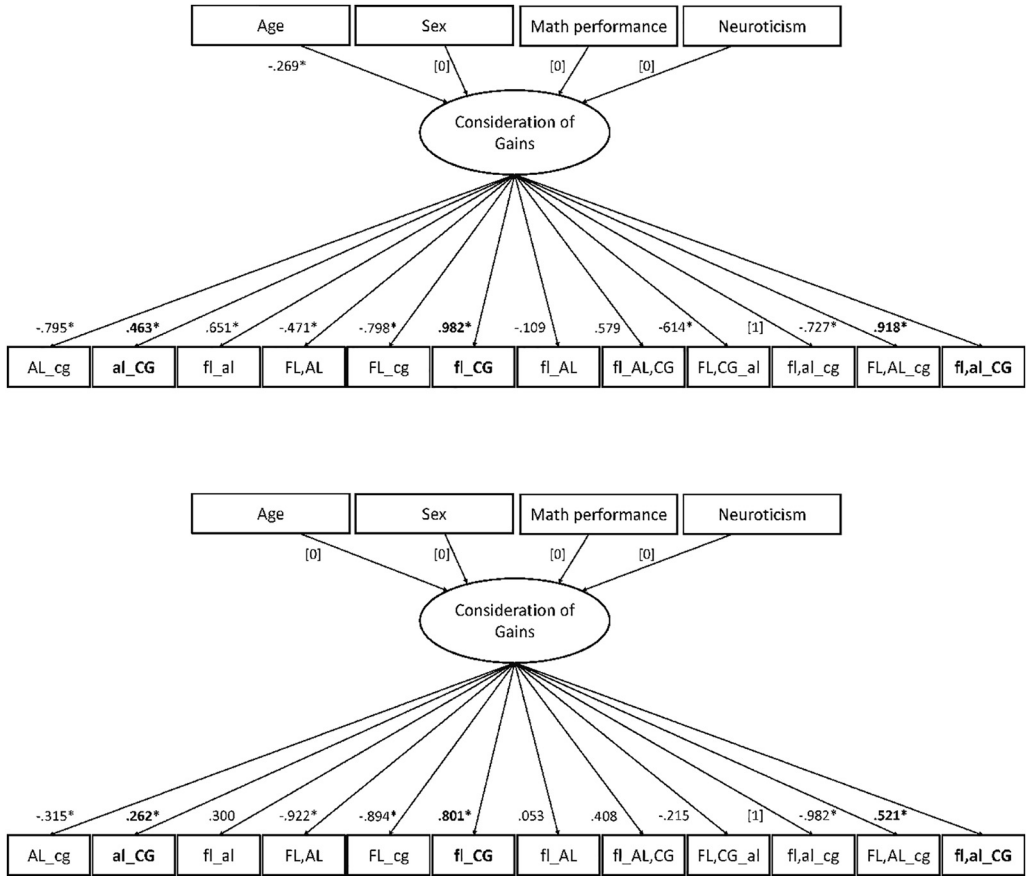


Fig. 4. Quantitative multiple indicator multiple cause (MIMIC) model parameters separately for the judgment data (top panel) and choice data (bottom panel). Parameters in brackets were fixed. An asterisk (*) indicates significance at $p < .05$. Items in which certain gain (CG) uniquely signaled the best option are bolded.

Conversely, they do not provide support for qualitative conceptualizations derived from dual process theory, fuzzy trace theory, or lexicographic theory (Evans, 2008, 2011; Jansen & van der Maas, 2002; Jansen et al., 2012; Kokis et al., 2002; Reyna & Ellis, 1994; Siegler & Chen, 2002; Stanovich et al., 2011; Steingroever et al., 2019; Tversky et al., 1988). The current study therefore motivates future studies in which developmental differences in decision making are studied from the three quantitative theoretical perspectives.

The second new insight is that quantitative developmental differences in decision making are related to a developmental decrease in consideration of gains. This finding generates the novel hypothesis that developmental differences in decision making will manifest themselves in parameters related to consideration of gains in each of the three quantitative theories. That is, in information integration theory they will appear in the weight parameter associated with gains, in cumulative prospect theory they will appear in the parameters governing subjective evaluation of gains, and in risk return theory they will appear in the weight associated with risk. To our knowledge, this hypothesis has hardly ever been tested. That is, no information integration studies tested the development of gain weights, only one cumulative prospect theory study investigated the development of subjective weighting of gains and found no such effect (Stelandt et al., 2013), and only a few risk return theory studies reported development in weights associated with risk (Paulsen et al., 2011; van Duijvenvoorde

et al., 2015; Wolf et al., 2013). The current findings therefore motivate future studies into the development of gain-related information in decision making (Horn, Mata, & Pachur, 2020).

The third new insight relates to an absence of an effect of sex, neuroticism, or math performance on quantitative individual differences in consideration of gains. Thus, male and female individuals do not differ in to what extent they use gain-related information in their judgment or choice. Because increased consideration of gains (i.e., overweighting of gains as compared with losses) will often result in risk taking, the absence of an effect of sex is inconsistent with previous findings that risk taking is enhanced in male individuals as compared with female individuals (Blais & Weber, 2001; Byrnes et al., 1999). Neuroticism also had no effect on consideration of gains. It has been shown that neuroticism has an opposite effect on risk taking in gain and loss domains (Lauriola & Levin, 2001), which may explain the absence of an effect in the current study where options were characterized by both gains and losses. Neuroticism has also been associated with simple over complex strategy use (Hilbig, 2008). Because the current results indicate no qualitative individual differences, and therefore no individual differences in strategy use, the current results are not inconsistent with Hilbig (2008) findings. Similarly, math performance was studied because it may be predictive of complex versus simple strategy use. Because we did not find individual differences in strategy use, the absence of an effect of math performance is reasonable.

The fourth new insight relates to that we investigated, for the first time, development of both judgment and choice in an otherwise equivalent decision-making task. In both the judgment and choice versions, a quantitative conceptualization, and not a qualitative conceptualization, provided the best fit. The absence of qualitative differences in judgment is in accordance with previous findings on the balance scale task, but the absence of qualitative differences in choice is not (Wilkening & Anderson, 1982). Several explanations may be put forward with respect to the latter discrepancy, for example, differences between decision making and balance scale tasks, differences in sample size and age range, and differences in methodology used to assess qualitative differences. This then would suggest that the notion that development in choice tasks consists of going through qualitatively different stages is not general but rather depends on task, sample, and method, all of which deserve further study. Another point worth mentioning is that developmental differences in consideration of gains were observed in the judgment version but not in the choice version of the task.¹⁵ We refrain from speculation on the origin of this developmental discrepancy between judgment and choice and only suggest that it requires further study.

The final new insight is that it is important to test, rather than assume, the quantitative versus qualitative nature of developmental differences. Although this testing of the nature of individual differences has become an active field of research in clinical developmental psychology (e.g., Beauchaine, 2003, 2007; Coghill & Sonuga-Barke, 2012; Haslam, Holland, & Kuppens, 2012), the current study is, to our knowledge, the first to investigate this in cognitive developmental domains. For example, in our previous GMT studies, we assumed qualitative individual and developmental differences (Bexkens et al., 2016; Dekkers et al., 2020; Jansen et al., 2012; Steingroever et al., 2019; Van Duijvenvoorde et al., 2016; Van Duijvenvoorde, Jansen, Visser, & Huizenga, 2010). Although a reanalysis of one of these studies using the current MIMIC approach confirmed that individual differences among adults were indeed better described by a qualitative model than by a quantitative model (Zadelaar et al., 2019), it remains to be tested whether this is also the case in these other studies.

These five novel insights should be considered in light of potential limitations. First, the current study was not preregistered, and thus the analyses were exploratory. Second, the current conclusions are, as always, specifically tied to the current task and population (Yarkoni, 2020). For example, if we would have administered more complex items, would have used time pressure, would have included more than two options in each item, and/or would have also included children of a younger age, qualitative age-related differences may have become apparent (Gonthier & Roulin, 2020; Jansen et al., 2012; Payne, Bettman, & Johnson, 1988; Van Duijvenvoorde et al., 2016). Relatedly, we used specific operationalizations of judgment and choice; other operationalizations may give rise to different results. That is, to be able to assess qualitative differences in strategy use, participants in the current

¹⁵ This may be a power issue because the third best model for the choice task did include an effect of age (cf. SOM Table 2).

study chose either one option or the other option or indicated that the options were equivalent. This possibility of equivalence is not always given in choice tasks. Moreover, to make the judgment task comparable to the choice task, participants in the current study rated one option in comparison with the other. In other judgment tasks, participants provide a rating of one option only. The current study therefore motivates future, preferably preregistered MIMIC studies into the generalizability of the current findings to different tasks and different populations.

Third, one may argue that the chosen models were too restrictive in that the qualitative MIMIC model did not allow for quantitative variation within classes. This potential limitation can be addressed in future studies by combining the MIMIC model with so-called factor mixture models that do allow for quantitative variation within classes (e.g., [Lubke & Muthén, 2005](#)) but that require a larger sample size. On the other hand, one may also argue that the models were not restrictive enough. That is, many free parameters needed to be estimated, and constraints on these parameters may have been beneficial. For example, in previous GMT research *assuming* qualitative differences, we identified several strategies ([Jansen et al., 2012](#); [Steingroever et al., 2019](#)). Given these results, we could have constrained parameters in the current analyses. We did not do so, however, because we wanted to create a level playing field for the qualitative and quantitative models; constraining parameters of the qualitative model, but not the quantitative model, may have favored the qualitative model.

Fourth, we defined quantitative differences as quantitative differences in parameters governing a single strategy. But what if people use the same strategy with qualitative differences, instead of quantitative differences, in parameters? For example, what if all people use an information integration strategy, but some have a gain weight parameter of zero, whereas for others this gain weight is markedly different from zero? Such differences may show up in the MIMIC analysis as qualitative changes even though people do use the same strategy ([Hofman, Visser, Jansen, & van der Maas, 2015](#)). Although this does not pose a problem in the current study because best fitting models were quantitative, this possibility should be kept in mind in future applications in which qualitative differences are observed.

The new MIMIC approach to testing quantitative versus qualitative developmental differences has provided four new insights into the nature of individual and developmental differences in decision making. Given these new insights obtained by using this methodology, we suggest that this novel MIMIC approach may also be informative in other cognitive domains. For example, to test the quantitative versus qualitative development of categorization, reasoning, math, or memory ([Bouwmeester, Vermunt, & Sijtsma, 2007](#); [Cottini, Basso, Saracini, & Palladino, 2019](#); [Hofman, Visser, Jansen, Marsman, & van der Maas, 2018](#); [Hofman et al., 2015](#); [Munakata & McClelland, 2003](#); [Reetzke, Maddox, & Chandrasekaran, 2015](#); [Schapiro & McClelland, 2009](#); [Stone, Blumberg, Blair, & Cancelli, 2016](#)).

Conclusion

The novel MIMIC approach proposed in this article provides a principled approach to test, rather than assume, quantitative versus qualitative developmental differences. The current study indicates that individual differences in decision making are quantitative, pertain to consideration of gains, and are related to age in judgment tasks but not in choice tasks. Because quantitative versus qualitative development is a key question in developmental research, the approach can also be informative in other domains.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by a Netherlands Organization of Scientific Research (NWO) Vici grant (453-12-005). We thank the children for their participation; Angelica Albers, Elisa Bindels, Aukje

den Blanken, Manouk van Egmond, and Mémé Veels for their assistance in data collection; Edmee Griffioen for supervising data collection; and Bernd Figner for helpful discussions on the task.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jecp.2021.105198>.

References

- Anderson, N. H. (1980). Information integration theory in developmental psychology. In F. Wilkening, J. Becker, & T. Trabasso (Eds.), *Information integration by children* (pp. 1–45). Hillsdale, NJ: Lawrence Erlbaum.
- Ashcraft, M. H., & Fierman, B. A. (1982). Mental addition in third, fourth, and sixth graders. *Journal of Experimental Child Psychology*, *33*, 216–234.
- Beauchaine, T. P. (2003). Taxometrics and developmental psychopathology. *Development and Psychopathology*, *15*, 501–527.
- Beauchaine, T. P. (2007). A brief taxometrics primer. *Journal of Clinical Child & Adolescent Psychology*, *36*, 654–676.
- 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.
- Betsch, T., Lehmann, A., Jekel, M., Lindow, S., & Glöckner, A. (2018). Children's application of decision strategies in a compensatory environment. *Judgment and Decision Making*, *13*, 514–528.
- Betsch, T., Lehmann, A., Lindow, S., Lang, A., & Schoemann, M. (2016). Lost in search: (Mal-)adaptation to probabilistic decision environments in children and adults. *Developmental Psychology*, *52*, 311–325.
- 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*, 357–367.
- Blais, A. R., & Weber, E. U. (2001). Domain-specificity and gender differences in decision making. *Risk, Decision and Policy*, *6*, 47–69.
- Bollen, K. A., & Bauldry, S. (2011). Three Cs in measurement models: Causal indicators, composite indicators, and covariates. *Psychological Methods*, *16*, 265–284.
- Bouwmeester, S., Vermunt, J. K., & Sijtsma, K. (2007). Development and individual differences in transitive reasoning: A fuzzy trace theory approach. *Developmental Review*, *27*, 41–74.
- Byrnes, J. P., Miller, D. C., & Schafer, W. D. (1999). Gender differences in risk taking: A meta-analysis. *Psychological Bulletin*, *125*, 367–383.
- Coghill, D., & Sonuga-Barke, E. J. S. (2012). Annual research review: Categories versus dimensions in the classification and conceptualisation of child and adolescent mental disorders—Implications of recent empirical study. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *53*, 469–489.
- Cottini, M., Basso, D., Saracini, C., & Palladino, P. (2019). Performance predictions and postdictions in prospective memory of school-aged children. *Journal of Experimental Child Psychology*, *179*, 38–55.
- De Vos, T. (1992). *Tempo Test Rekenen [Number Fact Retrieval Test]*. Nijmegen, Netherlands: Berkhout.
- Defoe, I. N., Dubas, J. S., Figner, B., & van Aken, M. A. G. (2012). A meta-analysis on age differences in risky decision making: Adolescents versus children and adults. *Psychological Bulletin*, *141*, 48–84.
- Dekkers, T. J., Huizenga, H. M., Popma, A., Bexkens, A., Zadelaar, J. N., & Jansen, B. R. J. (2020). Decision-making deficits in adolescent boys with and without attention-deficit/hyperactivity disorder (ADHD): An experimental assessment of associated mechanisms. *Journal of Abnormal Child Psychology*, *48*, 495–510.
- Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*, 255–278.
- Evans, J. S. B. T. (2011). Dual-process theories of reasoning: Contemporary issues and developmental applications. *Developmental Review*, *31*, 86–102.
- Figner, B., Mackinlay, R. J., Wilkening, F., & Weber, E. U. (2009). Affective and deliberative processes in risky choice: Age differences in risk taking in the Columbia Card Task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 709–730.
- Figner, B., & Schaub, S. (n.d.). Development of adaptive risky decision making: Risk sensitivity in judgment and choice. Unpublished working paper.
- Figner, B., & Voelki, N. (2004). Risky decision making in a computer card game: An information integration experiment. *Polish Psychological Bulletin*, *35*, 135–139.
- Glöckner, A., & Pachur, T. (2012). Cognitive models of risky choice: Parameter stability and predictive accuracy of prospect theory. *Cognition*, *123*, 21–32.
- Gonthier, C., & Roulin, J. L. (2020). Intraindividual strategy shifts in Raven's matrices, and their dependence on working memory capacity and need for cognition. *Journal of Experimental Psychology: General*, *149*, 564–579.
- Harbaugh, W. T., Krause, K., & Vesterlund, L. (2002). Risk attitudes of children and adults: Choices over small and large probability gains and losses. *Experimental Economics*, *5*, 53–84.
- Haslam, N., Holland, E., & Kuppens, P. (2012). Categories versus dimensions in personality and psychopathology: A quantitative review of taxometric research. *Psychological Medicine*, *42*, 903–920.
- Hilbig, B. E. (2008). Individual differences in fast-and-frugal decision making: Neuroticism and the recognition heuristic. *Journal of Research in Personality*, *42*, 1641–1645.
- Hofman, A. D., Visser, I., Jansen, B. R. J., Marsman, M., & van der Maas, H. L. J. (2018). Fast and slow strategies in multiplication. *Learning and Individual Differences*, *68*, 30–40.

- Hofman, A. D., Visser, I., Jansen, B. R. J., & van der Maas, H. L. J. (2015). The balance-scale task revisited: A comparison of statistical models for rule-based and information-integration theories of proportional reasoning. *PLoS One*, *10*(10) e136449.
- Horn, S. S., Mata, R., & Pachur, T. (2020). Good + bad = ? Developmental differences in balancing gains and losses in value-based decisions from memory. *Child Development*, *91*, 417–438.
- 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*, 814–825.
- 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.
- 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.
- Jones, P. R., & Dekker, T. M. (2018). The development of perceptual averaging: Learning what to do, not just how to do it. *Developmental Science*, *21*, 1–11.
- Joreskog, K. G. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, *70*, 631–639.
- Kievit, R. A., Romeijn, J. W., Waldorp, L. J., Wicherts, J. M., Scholte, H. S., & Borsboom, D. (2011). Mind the gap: A psychometric approach to the reduction problem. *Psychological Inquiry*, *22*, 67–87.
- Kievit, R. A., van Rooijen, H., Wicherts, J. M., Waldorp, L. J., Kan, K. J., Scholte, H. S., & Borsboom, D. (2012). Intelligence and the brain: A model-based approach. *Cognitive Neuroscience*, *3*, 89–97.
- Kokis, J. V., Macpherson, R., Toplak, M. E., West, R. F., & Stanovich, K. E. (2002). Heuristic and analytic processing: Age trends and associations with cognitive ability and cognitive styles. *Journal of Experimental Child Psychology*, *83*, 26–52.
- Lang, A., & Betsch, T. (2018). Children's neglect of probabilities in decision making with and without feedback. *Frontiers in Psychology*, *9*. <https://doi.org/10.3389/fpsyg.2018.00191>.
- Lauriola, M., & Levin, I. P. (2001). Personality traits and risky decision-making in a controlled experimental task: An exploratory study. *Personality and Individual Differences*, *31*, 215–226.
- Lindow, S., & Betsch, T. (2018). Child decision-making: On the burden of predecisional information search. *Journal of Cognition and Development*, *19*, 137–164.
- Lindow, S., & Betsch, T. (2019). Children's adaptive decision making and the costs of information search. *Journal of Applied Developmental Psychology*, *60*, 24–34.
- Lindow, S., Lang, A., & Betsch, T. (2017). Holistic information integration in child decision making. *Journal of Behavioral Decision Making*, *30*, 1131–1146.
- Liu, S., Gonzalez, G., & Warneken, F. (2019). Worth the wait: Children trade off delay and reward in self- and other-benefiting decisions. *Developmental Science*, *22* e12702.
- Lubke, G. H., & Muthén, B. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods*, *10*, 21–39.
- Mata, R., von Helversen, B., & Rieskamp, J. (2011). When easy comes hard: The development of adaptive strategy selection. *Child Development*, *82*, 687–700.
- McGrath, R. E., & Walters, G. D. (2012). Taxometric analysis as a general strategy for distinguishing categorical from dimensional latent structure. *Psychological Methods*, *17*, 284–293.
- Montgomery, H., Selart, M., Gärling, T., & Lindberg, E. (1994). The judgment-choice discrepancy: Noncompatibility or restructuring?. *Journal of Behavioral Decision Making*, *7*, 145–155.
- Munakata, Y., & McClelland, J. L. (2003). Connectionist models of development. *Developmental Science*, *6*, 413–429.
- Nussenbaum, K., & Hartley, C. A. (2019). Reinforcement learning across development: What insights can we draw from a decade of research?. *Developmental Cognitive Neuroscience*, *40* 100733.
- Paulsen, D. J., Platt, M. L., Huettel, S. A., & Brannon, E. M. (2011). Decision-making under risk in children, adolescents, and young adults. *Frontiers in Psychology*, *2*. <https://doi.org/10.3389/fpsyg.2011.00072>.
- Payne, J. W., Bettman, J. R., Coupey, E., & Johnson, E. J. (1992). A constructive process view of decision making: Multiple strategies in judgment and choice. *Acta Psychologica*, *80*, 107–141.
- Payne, J. W., Bettman, J. R., & Johnson, E. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 534–552.
- Plebanek, D. J., & Sloutsky, V. M. (2019). Selective attention, filtering, and the development of working memory. *Developmental Science*, *22* e12727.
- Posada, D., & Buckley, T. R. (2004). Model selection and model averaging in phylogenetics: Advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology*, *53*, 793–808.
- Reetzke, R., Maddox, W. T., & Chandrasekaran, B. (2015). The role of age and executive function in auditory category learning. *Journal of Experimental Child Psychology*, *142*, 48–65.
- Reyna, V. F., & Ellis, S. C. (1994). Fuzzy-trace theory and framing effects in children's risky decision making. *Psychological Science*, *5*, 275–279.
- Rosenbaum, G. M., Venkatraman, V., Steinberg, L., & Chein, J. M. (2018). The influences of described and experienced information on adolescent risky decision making. *Developmental Review*, *47*, 23–43.
- Schapiro, A. C., & McClelland, J. L. (2009). A connectionist model of a continuous developmental transition in the balance scale task. *Cognition*, *110*, 395–411.
- Schlottmann, A., & Tring, J. (2005). How children reason about gains and losses: Framing effects in judgement and choice. *Swiss Journal of Psychology*, *64*, 153–171.
- Scholte, R. H. J., & De Bruyn, E. E. J. (2001). The Revised Junior Eysenck Personality Questionnaire (JEPQ-R): Dutch replications of the full-length, short, and abbreviated forms. *Personality and Individual Differences*, *31*, 615–625.
- Selart, M. (1996). Structure compatibility and restructuring in judgment and choice. *Organizational Behavior and Human Decision Processes*, *65*, 106–116.
- Siegler, R. S. (1976). Three aspects of cognitive development. *Cognitive Psychology*, *8*, 481–520.
- Siegler, R. S. (2007). Cognitive variability. *Developmental Science*, *10*, 104–109.

- 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.
- Stanovich, K. E., West, R. F., & Toplak, M. E. (2011). The complexity of developmental predictions from dual process models. *Developmental Review*, *31*, 103–118.
- Stelandt, S., Broihanne, M. H., Romain, A., Thierry, B., & Dufour, V. (2013). Decision-making under risk of loss in children. *PLoS One*, *8*(1) e52316.
- Steingroever, H., Jepma, M., Lee, M. D., Jansen, B. R. J., & Huizenga, H. M. (2019). Detecting strategies in developmental psychology. *Computational Brain & Behavior*, *2*, 128–140.
- Stone, M. M., Blumberg, F. C., Blair, C., & Cancelli, A. A. (2016). The “EF” in deficiency: Examining the linkages between executive function and the utilization deficiency observed in preschoolers. *Journal of Experimental Child Psychology*, *152*, 367–375.
- Tversky, A., & Kahneman, D. (1992). Advances in prospect-theory—Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*, 297–323.
- Tversky, A., Sattath, S., & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological Review*, *95*, 371–384.
- Van Duijvenvoorde, A. C. K., Figner, B., Weeda, W. D., van der Molen, M. W., Jansen, B. R. J., & Huizenga, H. M. (2016). Neural mechanisms underlying compensatory and noncompensatory strategies in risky choice. *Journal of Cognitive Neuroscience*, *28*, 1358–1373.
- Van Duijvenvoorde, A. C. K., Jansen, B. R. J., Visser, I., & Huizenga, H. M. (2010). Affective and cognitive decision-making in adolescents. *Developmental Neuropsychology*, *35*, 539–554.
- van Duijvenvoorde, A. C. K., Huizenga, H. M., Somerville, L. H., Delgado, M. R., Powers, A., Casey, B. J., ... Figner, B. (2015). Neural correlates of expected risks and returns in risky choice across development. *Journal of Neuroscience*, *35*, 1549–1560.
- Vrieze, S. I. (2012). Model selection and psychological theory: A discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods*, *17*, 228–243.
- Weber, E. U., Shafir, S., & Blais, A.-R. (2004). Predicting risk sensitivity in humans and lower animals: Risk as variance or coefficient of variation. *Psychological Review*, *111*, 430–445.
- Westenberg, M. R. M., & Koele, P. (1992). Response modes, decision processes and decision outcomes. *Acta Psychologica*, *80*, 169–184.
- Wilkening, F. (1981). Integrating velocity, time, and distance information: A developmental study. *Cognitive Psychology*, *13*, 231–247.
- Wilkening, F., & Anderson, N. H. (1982). Comparison of two rule-assessment methodologies for studying cognitive development and knowledge structure. *Psychological Bulletin*, *92*, 215–237.
- Wolf, L. K., Wright, N. D., Kilford, E. J., Dolan, R. J., & Blakemore, S. J. (2013). Developmental changes in effects of risk and valence on adolescent decision-making. *Cognitive Development*, *28*, 290–299.
- Yarkoni, T. (2020). The generalizability crisis. *Behavioral and Brain Sciences*, *21*, 1–37.
- Zadelaar, J. N., Weeda, W. D., Waldorp, L. J., Van Duijvenvoorde, A. C. K., Blankenstein, N. E., & Huizenga, H. M. (2019). Are individual differences quantitative or qualitative? An integrated behavioral and fMRI MIMIC approach. *NeuroImage*, *202*, 116058.